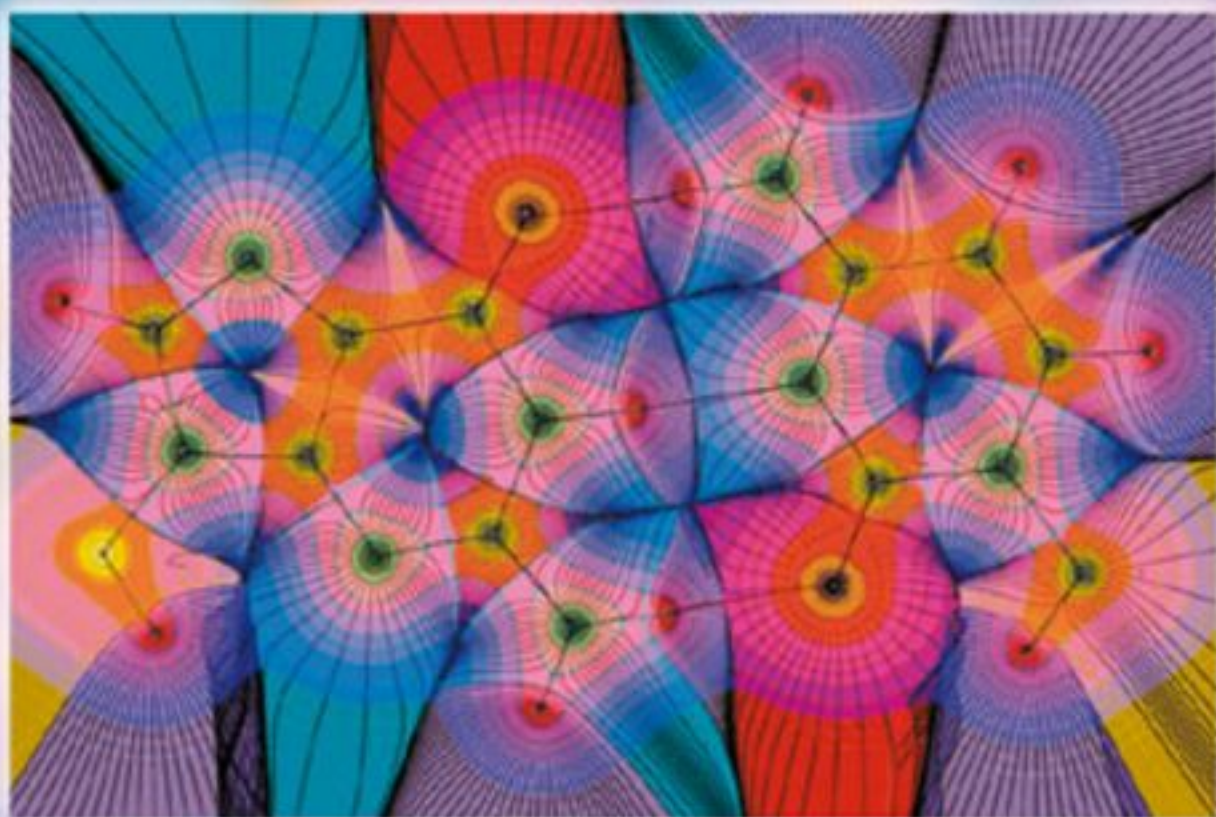


Edited by Chérif F. Matta

 WILEY-VCH

Quantum Biochemistry

Electronic Structure and Biological Activity



Edited by
Chérif F. Matta

Quantum Biochemistry



**WILEY-
VCH**

WILEY-VCH Verlag GmbH & Co. KGaA

Edited by
Chérif F. Matta

Quantum Biochemistry

Related Titles

Feig, M. (Ed.)

Modeling Solvent Environments

Applications to Simulations of Biomolecules

2010

Hardcover

ISBN: 978-3-527-32421-7

Morokuma, K., Musaev, D. (eds.)

Computational Modeling for Homogeneous and Enzymatic Catalysis

A Knowledge-Base for Designing Efficient Catalysts

2008

Hardcover

ISBN: 978-3-527-31843-8

Reiher, M., Wolf, A.

Relativistic Quantum Chemistry

The Fundamental Theory of Molecular Science

2009

Hardcover

ISBN: 978-3-527-31292-4

Matta, C. F., Boyd, R. J. (eds.)

The Quantum Theory of Atoms in Molecules

From Solid State to DNA and Drug Design

2007

Hardcover

ISBN: 978-3-527-30748-7

Meyer, H.-D., Gatti, F., Worth, G. A. (eds.)

Multidimensional Quantum Dynamics

MCTDH Theory and Applications

2009

Hardcover

ISBN: 978-3-527-32018-9

Rode, B.M., Hofer, T., Kugler, M.

The Basics of Theoretical and Computational Chemistry

2007

Hardcover

ISBN: 978-3-527-31773-8

Comba, P., Hambley, T. W., Martin, B.

Molecular Modeling of Inorganic Compounds

2009

Hardcover

ISBN: 978-3-527-31799-8

Edited by
Chérif F. Matta

Quantum Biochemistry



**WILEY-
VCH**

WILEY-VCH Verlag GmbH & Co. KGaA

The Editor

Prof. Chérif F. Matta

Dept. of Chemistry & Physics
Mount Saint Vincent Univ.
Halifax, Nova Scotia
Canada B3M 2J6

and

Dept. of Chemistry
Dalhousie University
Halifax, Nova Scotia,
Canada B3H 4J3

Cover:

About the cover graphic (from Chapter 14):

A superimposition of (1) the electron density ρ contour map of a Guanine-Cytosine Watson-Crick base pair in the molecular plane (the outermost contour is the $0.001 \text{ e}^-/\text{bohr}^3$ iso-contour followed by 2×10^{-4} , 4×10^{-4} , and $8 \times 10^{-4} \text{ e}^-/\text{bohr}^3$ with n starting at -3 and increasing in steps of unity); and (2) representative lines of the gradient of the density $\nabla \rho$. The density is partitioned into non-spherical color-coded “atoms-in-molecules (AIM)”, each containing a single nucleus. (Adapted from: C. F. Matta, *PhD Thesis*, McMaster University, Hamilton, Canada, 2002). (Courtesy of Chérif F. Matta).

Credit:

The phrase “*Quantum Biochemistry*” used in the title of this book has been coined by Bernard Pullman and Alberte Pullman (B. Pullman and A. Pullman, *Quantum Biochemistry*; Interscience Publishers: New York, 1963).

All books published by Wiley-VCH are carefully produced. Nevertheless, authors, editors, and publisher do not warrant the information contained in these books, including this book, to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

Library of Congress Card No.: applied for

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <http://dnb.d-nb.de>.

© 2010 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

All rights reserved (including those of translation into other languages). No part of this book may be reproduced in any form – by photoprinting, microfilm, or any other means – nor transmitted or translated into a machine language without written permission from the publishers. Registered names, trademarks, etc. used in this book, even when not specifically marked as such, are not to be considered unprotected by law.

Printed in the Federal Republic of Germany
Printed on acid-free paper

ISBN: 978-3-527-32322-7

*To every experimentalist and theoretician who has contributed to Quantum Biochemistry,
and to every scientist, practitioner, and philosopher in whom its advancement, use, and
interpretation finds fruition.*

Acknowledgment

This book is the result of the contributions of Ms. Alya A. Arabi, Dr. J. Samuel Arey, Prof. Paul W. Ayers, Prof. Richard F.W. Bader, Dr. José Enrique Barquera-Lozada, Dr. Joan Bertran, Dr. Michel Bitbol, Mr. Hugo J. Bohrquez, Prof. Russell J. Boyd, Dr. Denis Bucher, Dr. Steven K. Burger, Prof. Roberto Cammi, Prof. Chiara Cappelli, Dr. Constanza Cárdenas, Prof. Paolo Carloni, Dr. Lung Wa Chung, Dr. Fernando Clemente, Prof. Fernando Cortés-Guzmán, Prof. Gabriel Cuevas, Prof. Matteo Dal Peraro, Prof. Katherine V. Darvesh, Prof. Sultan Darvesh, Prof. Bijoy K. Dey, Prof. Leif A. Eriksson, Dr. Laura Estévez, Dr. Michael J. Frisch, Prof. James W. Gauld, Dr. Konstantinos Gkionis, Dr. María J. González Moa, Dr. Ana M. Graña, Dr. Anna V. Gubskaya, Ms. Mireia Güell, Dr. Mark Hicks, Dr. J. Grant Hill, Dr. Lulu Huang, Dr. Marek R. Janicki, Dr. Jerome Karle, Dr. Nouredin El-Bakali Kassimi, Prof. Eugene S. Kryachko, Dr. Xin Li, Ms. Yuli Liu, Dr. Jorge Llano, Mr. Jean-Pierre Llored, Dr. Marcos Mandado, Prof. Earl Martin, Prof. Lou Massa, Dr. Fanny Masson, Prof. Robert S. McDonald, Prof. Benedetta Mennucci, Prof. Keiji Morokuma, Prof. Ricardo A. Mosquera, Dr. Klefah A.K. Musa, Dr. Marc Noguera, Prof. Manuel E. Patarroyo, Prof. Jason K. Pearson, Dr. James A. Platts, Prof. Paul L.A. Popelier, Prof. Ian R. Pottie, Prof. Arvi Rauk, Dr. Arturo Robertazzi, Prof. Jorge H. Rodriguez, Dr. Luis Rodríguez-Santiago, Prof. Ursula Röthlisberger, Ms. Debjani Roy, Ms. Lesley R. Rutledge, Dr. Utpal Sarkar, Prof. Paul von Ragué Schleyer, Prof. Mariona Sodupe, Prof. Miquel Solà, Dr. David N. Stamos, Dr. Marcel Swart, Prof. Ajit J. Thakkar, Prof. Jacopo Tomasi, Prof. Alejandro J. Vila, Dr. Thom Vreven, Prof. Donald F. Weaver, Prof. Stacey D. Wetmore, and Prof. Ada Yonath. I cannot thank each contributor enough for accepting my invitation. I feel honored to have had the chance of working with such an exceptional group of scientists.

The staff of Wiley-VCH has been instrumental in all phases of the development of this project from its conception by copy-editing, proof reading, preparing galley proofs, contacting authors, and for the timely production of this book. I have been very lucky to work with them and extend my deepest thanks to Dr. Heike Noethe, Dr. Eva-Stina Riihimäki, Dr. Ursula Schling-Brodersen, Dr. Martin Ottmar, Ms. Claudia Nussbeck, and Ms. Hiba-tul-Habib Nayyer for their considerable effort, professionalism, experience, and expertise on which I have constantly relied in the past two years.

I am very grateful to Prof. Lou Massa for his invaluable help in the form of opinion and advice about the concept and design of this book.

I thank my colleagues and the administration at *Mount Saint Vincent University*, past and present, for their moral and administrative support and continual encouragement. I am also indebted to *Dalhousie University* and the *Université Henri Poincaré (Nancy Université – 1)* for access to their resources, including their libraries, by virtue of, first, an “honorary Adjunct Professorship”, and second, a “Visiting Professorship”.

Extremely fortunate would be an understatement as to how I personally feel about knowing, working with, and benefiting from the exceptional professional mentorship of Professors Richard F. W. Bader, Russell J. Boyd, Claude Lecomte, Lou Massa, and John C. Polanyi. I cannot see how I could have edited this book without having considerably benefited in numerous ways from my association with each.

The funding received by my research group was indispensable for the completion of this project. I am much obliged to the *Natural Sciences and Engineering Research Council of Canada* (NSERC), *Canada Foundation for Innovation* (CFI), and *Mount Saint Vincent University* for financial support.

In closing, and on a more personal note, I wish to express my deepest and most affectionate gratitude to the memory of those who gave me life: Farid A. Matta, and Nabila Matta (née Nassif Abdel-Nour) for bringing me up in a rich and vibrant intellectual atmosphere with a well-stocked library and art collection at our home in Alexandria, and to the other members of our family who have always supported me unconditionally, in particular during the unfolding of this demanding project: Maged, Heba, Sara, and Nadine Matta.

Chérif F. Matta

Congratulations to Professor Ada Yonath for Winning the 2009 Nobel Prize in Chemistry

The editor this book and the staff of Wiley-VCH extend their warmest congratulations to Professor Ada Yonath for winning the 2009 Nobel Prize in Chemistry. They undertake this opportunity to thank her again for her contribution to this book (Chapter 16) that she has co-authored with Prof. Lou Massa, Prof. Chérif F. Matta, and Dr. Jerome Karle.

Introductory Reflections on Quantum Biochemistry: From Context to Contents

Chérif F. Matta

“I will at least report novel properties of gases, the effects of which are regular, by proving that these substances combine among each other in very simple ratios, and that the volume contraction that they experience by the combination follows also a regular law. I hope to provide through that a proof of what has been put forward by very distinguished chemists, that we are perhaps not far from the epoch in which we will be able to submit to calculation the majority of chemical phenomena.”¹⁾

Louis-Joseph Gay-Lussac, 31 December 1808 [1].

Two hundred and one years ago, almost to the day, Gay-Lussac (1778–1850) made the far-reaching prediction that, one day, the “majority of chemical phenomena” will be amenable to calculations. The boldness of this prediction is as extraordinary as the accuracy with which it has been (and is being) realized. The history of science since the early nineteenth century to the present is extremely rich and complex and studded with important milestones that fall well beyond the scope of these short introductory remarks and outside of the knowledge comfort zone of the writer, so only a few relevant highlights will be offered to set the stage for this book. One of these milestones was the award of the 1998 Nobel Prize in Chemistry, two centuries short of a decade after Gay-Lussac’s prediction, “to Walter Kohn for his development of the density-functional theory and to John Pople for his development of computational methods in quantum chemistry”. This visionary opening quotation, with wording such as “soumettre au calcul” or “submit to calculation”, cannot have a more contemporary ring!

1) Translated by the present writer from the original text in French: «Je vais du moins faire connaître des propriétés nouvelles dans les gaz, dont les effets sont réguliers, en prouvant que ces substances se combinent entre elles dans des rapports très-simples, et que la contraction de volume qu’elles éprouvent par

la combinaison suit aussi une loi régulière. J’espère donner par là une preuve de ce qu’ont avancé des chimistes très – distingués, qu’on n’est peut-être pas éloigné de l’époque à laquelle on pourra soumettre au calcul la plupart des phénomènes chimiques » [1]. (See Figure 1).

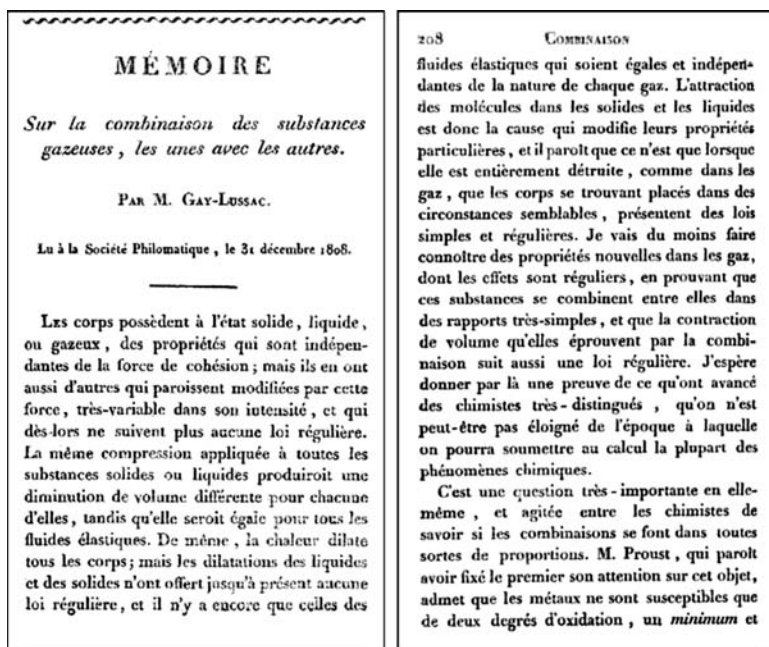


Figure 1 The first two pages of L. J. Gay-Lussac's 1809 paper (Ref. [1]). The paper was read in the last day of 1808 but was published in 1809. (The "M." before the name of the author is the title "Monsieur", or Mr.)

The quotation is extracted from the second page of Gay-Lussac's 1809 paper [1] "*On the combination of gaseous substances, one another*" (Figure 1). In this paper Gay-Lussac applies the concepts of the modern atomic theory formulated by his contemporary, John Dalton [2], to explain why gases combine in simple volumetric proportions. An immediate progeny of Gay-Lussac's paper was one by Amedeo Avogadro (1776–1856), who, *in a single paper*, introduced the concepts of "mole", the number later to be named in his honor N_A , a method to calculate atomic and molecular weights, and the distinction between "elementary molecules [atoms]" and "molecules" [3]. Avogadro's work led Stanislao Cannizzaro (1826–1910) to the determination of atomic weights for the first time in 1858 [4]. Two years later, in September 1860, Kékulé, Wurtz, and Weltzien organized the *Karlsruhe Congress* [5, 6], an international meeting that was attended by prominent chemists at the time, later to evolve into the *International Union of Pure and Applied Chemistry* (IUPAC) [5]. Among the participants in the 1860 meeting were the likes of Cannizzaro but also less established young scientists including 26-year-old Dmitri Ivanovich Mendeleev (and also 30-year-old Julius L. Meyer). Reprints of Cannizzaro's paper [4] were distributed to the participants [5], including Mendeleev and Meyer, the principal characters in the following act in the historical drama of chemistry culminating with the periodic classification of the elements, initially on the basis of Cannizzaro's atomic weights.

In 1916, a century and eight years after Gay-Lussac read his Memoire before the *Société de physique et de chimie de la Société d'Arcueil*, Gilbert Newton Lewis (1875–1946) proposed his model of the chemical bond [7, 8]. Lewis recognized, for the first time, the tendency of free atoms to complete the noble gas electronic shell configuration and the central role played by the electron pair. Recognizing the importance of electron pairing in 1916 [7, 8] *before the advent of modern quantum mechanics and the discovery of spin*, is an extraordinary achievement. Without the benefit of the knowledge of electronic spin, Lewis was compelled to go as far as questioning the applicability of Coulomb's law itself at very small distances: "*Coulomb's law of inverse squares must fail at small distances*" [7]. Lewis's paper has marked, in the humble opinion of the writer, the conception of the modern electronic theory of chemical bonding.

In 1929, at the dawn of the era of quantum mechanics, Paul A. M. Dirac (1902–1984) opens his paper entitled "*Quantum Mechanics of Many-Electron System*" [9] by the, now well-known, statement:

"The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble."

What Dirac meant is that the solution of Schrödinger equation, the wavefunction Ψ , provides a complete description and thus contains all the information that can be known about the system in a given quantum state. But since the Schrödinger equation can be solved exactly only for a very small number of very simple systems (composed of one or two particles at the most), Dirac goes on to close the opening paragraph to his paper wishing that [9]:

"It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation."

Eighty years later, today in 2009, much of Dirac's wish to develop approximate methods to extend the application of quantum mechanics to "complex atomic systems" has been realized, but the search for better and faster approximations to solve the Schrödinger equation remains a subject of prime importance and current interest in theoretical and quantum chemical research. The need for these "*approximate practical methods*" is particularly pertinent to quantum biochemistry where quantum mechanics is applied to biological systems of staggering complexity, unimaginable just a few decades ago.

The Born-Oppenheimer (BO) approximation, that electrons being much lighter than nuclei are capable of readjusting their distribution "instantaneously" on the time-scale of nuclear motion, is one of the most accurate and seminal approximations in quantum chemistry. This approximation decouples the nuclear and electronic Hamiltonians, a considerable simplification by virtue of which the nuclei move on a

potential energy surface (PES) generated by solving the electronic Schrödinger equation for all possible nuclear geometries [10–12].²⁾

The concept of potential energy surface was advanced for the first time in 1931 by Henry Eyring and Michael Polanyi in their treatment of the $\text{H} + \text{H}_2$ reaction [13]. The concept has been further developed by Polanyi and Eyring but also by F. W. London, S. Sato, Philip M. Morse, and others [14–16]. Laidler's book [17] presents an excellent exposition of the role of PES in chemical kinetics and dynamics as well as biographies of 41 of the early pioneers in this field. The book edited by Back and Laidler [16] is a compilation of commented reprints of a selection of key papers on PES, dynamics, and kinetics including a reproduction of Savante Arrhenius' 1889 paper on $k = Ae^{-E_a/RT}$. An extraordinary collection of scholarly essays dedicated to Michael Polanyi by leading scientists (including his son, John C. Polanyi, who went on to win the 1986 Nobel Prize in Chemistry), economists, historians, and philosophers – a mix of disciplines that reflect the grandeur and the breadth of the intellect of Michael Polanyi – was published in 1961 on the occasion of his 70th birthday [18].

Thus the BO approximation allows for a separate solution of the electronic and nuclear problems. The solution of the *electronic, time-independent, non-relativistic, Born-Oppenheimer* molecular Schrödinger equation represents much of modern quantum chemistry (and quantum biochemistry), while the prediction of IR and Raman spectra require the solution of the nuclear Schrödinger equation.

Further approximations have given rise to the evolution of two equivalent branches of electronic structure theory: Valence Bond (VB) theory and Molecular Orbital (MO) theory. Valence bond theory was founded by W. H. Heitler and F. W. London, and further developed by J. C. Slater, L. C. Pauling, E. A. Hylleraas and several others. The theory is reviewed qualitatively in Pauling's monograph "*The Nature of the Chemical Bond*" [19] and in C. A. Coulson's "*Valence*" [20] and its updated version by R. McWeeney's "*Coulson's Valence*" [21]. VB theory has been reviewed in the recent books by S. Shaik and P. Hiberty [22] and by G. A. Gallup [23].

F. Hund and R. S. Mulliken developed the Molecular Orbitals approach, to which several others have also made substantial contributions, including J. Lennard-Jones, J. C. Slater, E. Hückel, C. Coulson, and John Pople. A set of coupled differential equations, one for each spin orbital, is obtained by the application of the variational principle. The solution is obtained in the form of a single Slater determinant in an iterative manner, the self-consistent field (SCF) approach, constituting what is now known as the Hartree-Fock (H-F) method [12, 24–29].

The spherical symmetry of atoms enables a separation of variables that facilitates the solution of the SCF problem. This advantage is lost in molecules, a problem that was solved by the introduction of the *linear combination of atomic orbitals* (LCAO) credited to Roothaan [30] and Hall [31]. The Roothaan equations can be solved from first principle (*ab initio* SCF theory) or through empirical parametrization and further simplifying approximations (*semi-empirical* methods). Depending on how Coulom-

2) There are cases where the BO approximation breaks down. See for example Ref. [168]. These cases are of considerable interest but of no implications in quantum biochemistry at the present stage of knowledge, to the best of the writer's knowledge.

bic correlation is accounted for in post-Hartree-Fock methods, a hierarchy of methods of different degrees of approximation is obtained.

An excellent commented exposition of reprints of early historical papers on MO and VB theories is available in a recent book edited by H. Hettema [32]. The “rivalry” between MO and VB theories has been the subject of a recent mind-stimulating tripartite conversation between Roald Hoffman, Sason Shaik, and Philippe Hiberty, a highly recommended reading [33].

A radically distinct approach to solve the electronic problem with the incorporation of Coulombic correlation, comparable in accuracy to post-HF methods but with a considerable computational economy, is modern Density Functional Theory (DFT) [34–36]. Perdew *et al.* [37] have recently published a very clear non-mathematical conceptual review of DFT’s basic principles and ideas, an excellent read.

While originally proposed by L. Tomas and E. Fermi, the *modern formulation* of DFT was born in 1964 when P. Hohenberg and W. Kohn announced their celebrated (HK) theorems [38]. The first HK theorem was reached through an elegant proof *ad absurdum* that there exists a unique functional relationship between the external potential and the electron density, and as a consequence, between the density and the total energy of the system. The second theorem states that the exact electron density of the ground state is one that minimizes the total energy. In other words, the second theorem states that the variational principle can be invoked to calculate the energy of the ground state. These powerful theorems in themselves offer no procedure to compute the energy given the density. W. Kohn and L. Sham devised a workable practical solution to this problem a year later, in 1965, when they cast the theory into a formalism that resembles the Hartree-Fock SCF method in structure but with a completely new meaning and interpretation of the (KS) orbitals [39]. The problem of finding the exact functional remains unsolved to the time of writing.

DFT has evolved to become a formidable computational tool in the arsenal of the solid state physicists, quantum and computational chemists, and computational biochemists thanks to the subsequent pioneering work of Walter Kohn, Axel D. Becke, Robert Parr, Weitao Yang, John Purdue, Donald Truhlar, Tom Ziegler and others [34–36]. DFT has branched into a utilitarian/computational flavor used extensively to generate the results similar to the ones reviewed in this book, but also into a branch often called “conceptual DFT” aiming at deepening our understanding of the physical bases of chemistry and pioneered by the Belgium school including P. Geerlings, F. De Proft, P. Bultinck, and the McMaster research group of Paul Ayers, among others (see for example Refs. [40, 41]).

The application of electronic structure calculations (wavefunction and density functional methods) to real problems has been pushed to the forefront by scientists such as John Pople, Paul von Ragué Schleyer, Henry F. Schaeffer III, Leo Radom, Warren J. Hehre, Keiji Morokuma, Jacopo Tomasi, Kendall Houk, and a number of other pioneers [42–45]. A crowning achievement of the computational implementation of electronic structure methods is the development over several decades of very sophisticated software such as GAUSSIAN [43, 46] and GAMESS [47] in molecular quantum mechanics, and CRYSTAL [48, 49] in solid state physics.

Electronic structure calculations, the primary focus of this book, represent a principal branch of a wider field that can be called theoretical and computational chemistry [44] and which includes, for example, molecular mechanics and force field methods, Monte Carlo simulations, molecular dynamics simulations, molecular modeling and docking, informatics, etc. [50–54].

The early uses of digital computers in chemistry marked the birth of computational chemistry in the 1950s. This period coincided with spectacular advances in structural biology that culminated in the discovery of the alpha-helical structure of DNA by James Watson and Francis Crick [55–58] on the basis of a well-resolved X-ray diffraction pattern obtained by Rosalind Franklin [59].

Interestingly, a book appeared in 1944 based on a series of lectures delivered a year earlier at Trinity College, Dublin, in the midst of World War II (in 1943), by Erwin Schrödinger. The book was not about wave mechanics but about biology viewed through a physicist's lens with the daring question *What is life?* as its title [60]. In this book, the word “code” was used for the first time in the context of genetics when Schrödinger described the chromosome as a “code-script”. In an incredibly unique leap of insight, and in a section entitled “The Variety of Contents Compressed in the Miniature Code”, Schrödinger writes [60]:

“It has often been asked how this tiny speck of material, nucleus of the fertilized egg, could contain an elaborate code-script involving all the future development of the organism. A well-ordered association of atoms, endowed with sufficient resistivity to keep its order permanently, appears to be the only conceivable material structure that offers a variety of possible (‘isomeric’) arrangements, sufficiently large to embody a complicated system of ‘determinations’ within a small spatial boundary. Indeed, the number of atoms in such a structure need not be very large to produce an almost unlimited number of possible arrangements. For illustration, think of the Morse code. The two different signs of dot and dash in well-ordered groups of not more than four allow thirty different specifications. Now, if you allowed yourself the use of a third sign, in addition to dot and dash, and used groups of not more than ten, you could form 88,572 different ‘letters’; with five signs and groups up to 25, the number is 372,529,029,846,191,405.”

“That the gene is to be thought of as an *information carrier*”, Watson says [58], was “the most important point made by Schrödinger”. Schrödinger's book was instrumental in its influence on a young generation of structural biologists that included James Watson and Francis Crick. In fact, apparently it is *What is Life?* that ignited the interest of Francis Crick to switch from physics to biology, as recounted by Watson [58].

It is a particularly remarkable piece of history that Schrödinger, the discoverer and inventor of much of quantum mechanics, was also the one who planted many of the seeds of modern structural and molecular biology, whether directly by underscoring the importance of investigating the nature of information coding in the gene (unknown at the time) or through his considerable influence on the careers, enthusiasm, and thoughts of major players such as Watson and Crick. *Thus the*

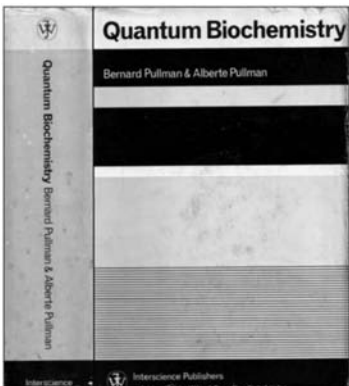
	<table> <tr> <td>Part I :</td> <td>MOLECULAR ORBITALS FOR BIOCHEMISTS</td> </tr> <tr> <td>I</td> <td>Why molecular orbitals?</td> </tr> <tr> <td>II</td> <td>Fundamental concepts</td> </tr> <tr> <td>III</td> <td>Conjugated molecules</td> </tr> <tr> <td>IV</td> <td>The principal applications of the indices of the electronic structure</td> </tr> <tr> <td>Part II :</td> <td>ELECTRONIC STRUCTURE OF FUNDAMENTAL BIOCHEMICAL SUBSTANCES</td> </tr> <tr> <td>V</td> <td>Purines, pyrimidines and the submolecular structure of nucleic acids</td> </tr> <tr> <td>VI</td> <td>Proteins as conjugated systems</td> </tr> <tr> <td>VII</td> <td>Energy-rich compounds</td> </tr> <tr> <td>VIII</td> <td>Pteridines</td> </tr> <tr> <td>IX</td> <td>Porphyrins and bile pigments</td> </tr> <tr> <td>X</td> <td>Conjugated biochemical polymers</td> </tr> <tr> <td>XI</td> <td>Quinones</td> </tr> <tr> <td>Part III :</td> <td>ELECTRONIC ASPECTS OF ENZYMIC REACTIONS</td> </tr> <tr> <td>XII</td> <td>General aspects of enzymic reactions</td> </tr> <tr> <td>XIII</td> <td>Oxidation-reduction enzymes</td> </tr> <tr> <td>XIV</td> <td>Folic acid coenzymes</td> </tr> <tr> <td>XV</td> <td>Pyridoxal phosphate enzymes</td> </tr> <tr> <td>XVI</td> <td>Thiamine-pyrophosphate catalyzed reactions</td> </tr> <tr> <td>XVII</td> <td>Enzymic hydrolysis</td> </tr> <tr> <td>XVIII</td> <td>Conclusion: electronic delocalization and the processes of life</td> </tr> </table>	Part I :	MOLECULAR ORBITALS FOR BIOCHEMISTS	I	Why molecular orbitals?	II	Fundamental concepts	III	Conjugated molecules	IV	The principal applications of the indices of the electronic structure	Part II :	ELECTRONIC STRUCTURE OF FUNDAMENTAL BIOCHEMICAL SUBSTANCES	V	Purines, pyrimidines and the submolecular structure of nucleic acids	VI	Proteins as conjugated systems	VII	Energy-rich compounds	VIII	Pteridines	IX	Porphyrins and bile pigments	X	Conjugated biochemical polymers	XI	Quinones	Part III :	ELECTRONIC ASPECTS OF ENZYMIC REACTIONS	XII	General aspects of enzymic reactions	XIII	Oxidation-reduction enzymes	XIV	Folic acid coenzymes	XV	Pyridoxal phosphate enzymes	XVI	Thiamine-pyrophosphate catalyzed reactions	XVII	Enzymic hydrolysis	XVIII	Conclusion: electronic delocalization and the processes of life
Part I :	MOLECULAR ORBITALS FOR BIOCHEMISTS																																										
I	Why molecular orbitals?																																										
II	Fundamental concepts																																										
III	Conjugated molecules																																										
IV	The principal applications of the indices of the electronic structure																																										
Part II :	ELECTRONIC STRUCTURE OF FUNDAMENTAL BIOCHEMICAL SUBSTANCES																																										
V	Purines, pyrimidines and the submolecular structure of nucleic acids																																										
VI	Proteins as conjugated systems																																										
VII	Energy-rich compounds																																										
VIII	Pteridines																																										
IX	Porphyrins and bile pigments																																										
X	Conjugated biochemical polymers																																										
XI	Quinones																																										
Part III :	ELECTRONIC ASPECTS OF ENZYMIC REACTIONS																																										
XII	General aspects of enzymic reactions																																										
XIII	Oxidation-reduction enzymes																																										
XIV	Folic acid coenzymes																																										
XV	Pyridoxal phosphate enzymes																																										
XVI	Thiamine-pyrophosphate catalyzed reactions																																										
XVII	Enzymic hydrolysis																																										
XVIII	Conclusion: electronic delocalization and the processes of life																																										

Figure 2 (a) Dust cover and (b) Abbreviated Table of Content of “Quantum Biochemistry” by Bernard Pullman and Alberte Pullman published in 1963 [61]. Note how current the topics listed in the table of content by today’s standards, more than four decades after its publication.

phrase “*Quantum Biochemistry*”, coined in 1963 by Bernard and Alberte Pullman [61] (Figure 2), while describing impeccably a definitive modern field of research whereby quantum mechanics is applied to biological molecules and reactions, the subject of this book, also epitomizes an era during which the synergy between physics and biology has benefited humankind in a manner that is rarely encountered in human intellectual history.

The discovery of the chemical nature and structure of the genetic material has, thus, brought biology within reach of the tools of a branch of applied quantum mechanics, namely, quantum chemistry, which when applied to biological systems is termed quantum biochemistry (QB). Among the earliest work in QB was the now well-known mechanism of spontaneous and induced mutation, proposed by Per-Olov Löwdin in 1963, in which a mutation is the result of tautomeric transitions of the two bases accompanied with double proton transfer by tunneling through the two barriers of the pair of double potential wells, each corresponding to a hydrogen bond linking the Watson-Crick partners [62, 63]. (See Chapter 31 of this book for a very interesting review of this mechanism and its evolutionary consequences). If this change in the hydrogen-bonding signature happens prior to transcription it results in the incorporation of an erroneous base in mRNA and may lead to a non-silent mutation if the altered codon is not a synonym of the original one. (An important three-volume collective work dedicated to the memory of Per-Olov Löwdin has recently been edited by E. J. Brändas and E. S. Kryachko [64] and includes chapters that review recent research done on this mechanism of mutation).

Another notable example of early insightful uses of computational quantum chemistry in biology was the elucidation of the nature of the high energy phosphate bond and the nature of its chelate with magnesium by Fukui *et al.* [65, 66]. Other early efforts in QB were spearheaded by the Pullmans. They relied on early semiempirical methods such as Hückel Theory or the PPP (Pariser-Parr-Pople) method to elucidate the electronic structure of polycyclic aromatic hydrocarbons (PAHs) and correlate it to carcinogenicity [67, 68], the electronic structure of nucleic acids [69], and to explore

stacking interactions between PAHs and nucleic acid bases [70]. Further examples are reviewed in Pullman and Pullman's remarkable monograph *Quantum Biochemistry* [61].

What is particularly commendable and admirable in the contribution of the Pullmans is their boldness in attacking problems of biology by performing calculations on molecules of sizes reaching a few dozens of atoms at a time when the results of *ab initio* calculations on diatomics were publishable in the best journals. To the Pullmans' credit also is their total mastery of both the biology and the physics and their ability to look beyond the calculation to the larger picture, evolutionary biology being a noted example [71]. A glance at the table of content of their book cannot convey a more timely impression even today in 2009 (Figure 2). This present book aims at contributing to review the state-of-the art of quantum biochemistry supplementing several excellent other books that have a similar goal (see for example Refs. [72–76]).

Naturally, the transformation of theoretical chemistry into computational chemistry has been greatly facilitated not only by the very fast increase in the power and availability of computers but also by the development of methods tailored for large molecules as they occur in quantum biochemistry. In the 1960s, performing an *ab initio* calculation on a small molecule composed of a handful of atoms represented the limit of what could be achieved. Nowadays, computational strategies have allowed for the calculation of ever increasingly large and complex systems.

In recent years, the need to study enzyme active sites under the influence of the surrounding (whether the surroundings are the remainder of the protein, of the immediate surrounding amino acid residues near the active site) has provided the impetus for the development of methods that treat the active site of interest at the highest achievable computational level of theory and treating the surrounding as the source of a perturbing field at lower (more economical) level(s) of theory, hence optimizing the balance of accuracy and speed. If the active site is treated quantum mechanically (QM) and the remainder of the protein by molecular mechanics (MM) the method is known as QM/MM [77, 78]. Hybrid methods have found numerous applications in biochemistry and are now a standard and very powerful tool in the hands of quantum and computational biochemists. (See Chapters 2, 3, 4, and 17 of this book for excellent reviews on hybrid quantum mechanical methods).

Another important breakthrough concerned with very large systems such as proteins and nucleic acids is the reconstruction of the density matrix of the target macromolecule from density matrices of its composing pieces termed “kernels”. This method, developed in its present form by Lulu Huang, Lou Massa, and Jerome Karle, the subject of the opening chapter of this book, is termed *Quantum Crystallography (QCr)* and is also sometimes referred to as the *Kernel Energy Method (KEM)*.

The QCr/KEM method has been rigorously and repeatedly tested by comparing *ab initio* wavefunctions obtained directly on full molecules to the corresponding wavefunctions reconstructed from kernels. This repeated benchmarking has established the accuracy and validity of this approximation. The crowning achievement of this

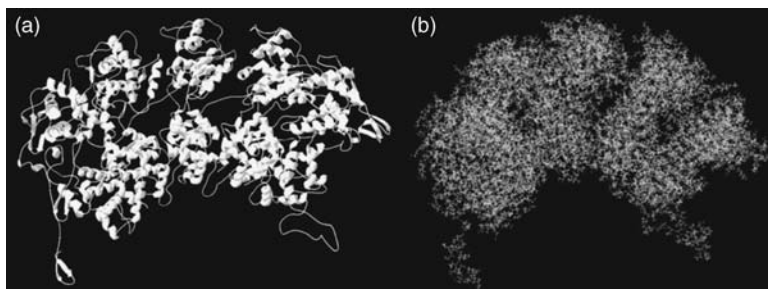


Figure 3 The crystal structure of vesicular stomatitis virus nucleocapsid protein Ser290Trp mutant (2QVJ) [87] (a) ribbon model (b) atomic model (without hydrogen atoms). The *ab initio* energy of this gigantic molecule composed of some 33,175 has been calculated using the Kernel Energy Method [79]. This is the largest *ab initio* calculation known to the writer at the time of writing.

approach has been the calculation of the Hartree-Fock [HF/6-31G(d,p)] energy as well as the MP2/6-31G(d,p) interaction energies within the vesicular stomatitis virus nucleoprotein, a protein composed of a staggering 33,175 atoms (Figure 3) [79]. This result has been the fruit of decades of development going back to the late 1960s [80–82] and more recently with applications to very large molecules such as DNA [83], tRNA [84], the ribosome [85], and insulin [86].

Solvation is another area of prime importance to the quantum chemistry of biological molecules. While solvation is still not considered as a “solved” problem in quantum chemistry, considerable advances have been achieved already. Solvent effects are commonly accounted for by either (a) the explicit incorporation of solvent molecules into the quantum mechanical calculation, sometimes referred to as the super-molecule approach, or (b) implicit solvation known as the self-consistent reaction field (SCRF) approach in which the solute is placed in a cavity inside the solvent (the shape of this cavity depends on the particular model chosen). The solvent is then modeled as a continuum characterized by its uniform dielectric constant [88–91]. Scientists such as Jacopo Tomasi, Donald Truhlar, and Cristopher Cramer are among the pioneers in this field. (See Chapter 4 for an authoritative review).

The discovery of solutions to the phase problem of X-ray crystallography, e.g., the discovery of direct methods by Jerome Karle and Herbert A. Hauptman (the Nobel Laureates in Chemistry for 1985), the dramatic engineering advances in the design of diffractometers and of data collection devices, most notably, the invention of the CCD (charge-coupled device) camera, and the advent of bright synchrotron X-ray sources, all contributed to an unprecedented shortening of the data collection and structure solution times. As a result, the solution of X-ray crystallographic structures has become standardized and faster than ever. Incidentally, the invention of the CCD is a theme of the 2009 Nobel Prize in Physics awarded to Willard S. Boyle and George E. Smith.

As a result of these exciting developments, and because of the widespread availability of the internet, we are now witnessing an exponential proliferation of

massive databases of structural information. Besides the deposition of crystallographic information files (cif) as electronic supplementary material to published articles, there are now several repositories of structural information, and to name a few important examples we list The Cambridge Structural Database (CSD), the Crystallography Open Database (COD), the Nucleic Acid Database, and the Protein Data Bank (PDB).

The largest object that has been crystallized to this day is the ribosome, a task generally believed impossible just a few years ago. The crystallization of the ribosome and the solution of its structure are achievements of epical proportions because they provide the atomic details necessary to understand how it reads the genetic information encoded in the mRNA and how it translates this information into a polypeptide. This is tantamount to uncovering one of life's most jealously guarded secrets. The implications of this fundamental knowledge are considerable for example in the design of selective protein synthesis inhibitors, *i.e.*, antibiotics that selectively target the ribosomes of harmful bacteria leaving human ribosomes intact. Venkatraman Ramakrishnan, Thomas A. Steitz, and Ada E. Yonath were awarded the 2009 Nobel Prize in Chemistry for solving the difficult jigsaw puzzle leading to the full atomic structure of the ribosome. Besides her contributions in working out key aspects of the structure and function of the ribosome, Ada Yonath is also credited for the development of an entirely new technique termed cryo-bio-crystallography, indispensable for the crystallization and subsequent solution of the ribosomal architecture [92]. Ada Yonath is the fourth woman to win the Prize in Chemistry, joining the league of Marie Curie (1911), Irène Joliot-Curie (1935), and Dorothy Crowfoot Hodgkin (1964).

Besides its primary role in yielding structural information about molecules of widely varying sizes and chemical composition, X-ray crystallography has also evolved into another direction concerned with the "nature of the chemical bond" in Pauling's words. In a "routine" crystallographic data treatment, the experimental structure factors are refined by iterative comparison with those obtained by a reverse Fourier transform of a model density. The model density of the unit cell is obtained from a guessed structure where spherical atomic densities are placed at the positions of the nuclei assumed in the model [93]. Only the atomic positions are allowed to change during the refinement cycles but not their spherical shape. This approach is suitable for molecular geometries but is not capable of capturing the subtle deformations of the electron density in regions relatively removed from the nuclei, as in the regions of chemical bonding. For that purpose, an aspherical multipolar refinement strategy is necessary [94]; a widely used multipolar model is that of Hansen and Coppens [95–97].

When the quality of a crystal is good and if the experiment is carefully conducted (preferably at very low temperatures) followed by the appropriate corrections and multipolar refinement, it can yield very accurate electron density maps of the bonding regions. *The question now is how to analyze these electron density maps? How to extract the chemistry folded and encoded within the density?* These questions are equally valid with reference to the output of the electronic structure calculations described above.

The answers to these important questions are rooted in the early 1960s, when Richard F. W. Bader *et al.* calculated and analyzed *ab initio* molecular electron density distributions well before the electron density was an object of intense interest [98–100]. In 1963 Richard F. W. Bader and Glenys A. Jones write [99]:

“The manner in which the electron density is disposed in a molecule has not received the attention its importance would seem to merit. Unlike the energy of a molecular system which requires a knowledge of the second-order density matrix for its evaluation [101] many of the observable properties of a molecule are determined in whole or in part by the simple three-dimensional electron-density distribution. In fact, these properties provide a direct measure of a wide spectrum of different moments averaged directly over the density distribution. Thus the diamagnetic susceptibility, the dipole moment, the diamagnetic contribution to the nuclear screening constant, the electric field, and the electric field gradient (as obtained from nuclear quadrupole coupling constants) provide a measure of (aside from any angular dependencies) $\langle r_i^2 \rangle$, $\langle r_i \rangle$, $\langle r_i^{-1} \rangle$, $\langle r_i^{-2} \rangle$, and $\langle r_i^{-3} \rangle$, respectively. The electric field at a nucleus due to the electron density distribution is of particular interest due to the theorem derived by Hellmann [102] and Feynman [103]. They have demonstrated that the force acting on a nucleus in a molecule is determined by the electric field at that nucleus due to the other nuclei and to the electron-density distribution.”

Over the past three decades, Bader and his students have constructed a theory of great elegance, beauty, generality, and power. This theory is referred to in the older literature as the “*Theory of Atoms-in-Molecules (AIM)*”, and in the more recent literature as the “*Quantum Theory of Atoms in Molecules (QTAIM)*” [104–109]. The theory in one stroke provides a framework to discuss, classify, and understand chemical structure and its (in)stability and transformations, chemical bonding interactions (note the usage as a verb [110]), and a coherent and physically and mathematically sound partitioning of the molecular space into individual atoms, hence the designation “Atoms-in-Molecules”. The partitioning of the molecular space into *non-overlapping non-spherical* atoms (see the cover graphic of this book) allows the partitioning of any molecular property that can be expressed as a local density into additive atomic and group contributions. In doing so, the theory has been shown on numerous occasions to recover experimental transferability and additivity schemes [104].

The theory has deep roots in quantum mechanics [111] and is founded on the analysis of Dirac observables (see Chapter 14 of this book for a brief introduction). The theory presents an interpretative and predictive scheme for chemistry that parallels experiment (see Refs. [112, 113]).

It has recently been proposed to re-name QTAIM as “Quantum Chemical Topology” and detailed and very compelling arguments to do so have been presented [114]. However, in the present writer’s view, changing the designation that everyone uses “*The Quantum Theory of Atoms in Molecules*” to another designation is not recommended because it can cause confusion in the vast

literature on the subject and will complicate literature searches. As a result, this is likely to diminish the impact of the theory. More important, perhaps, is that changing the designation of the theory may lead to the dilution of the credit that its principal developer, Richard F. W. Bader, deserves. Finally, in the opinion of this writer, it is incumbent on the principal developer of the theory to choose how to name it. Ref. [114] is a highly recommended reading.

QTAIM is becoming the standard theory used to interpret and analyze experimental charge densities [96, 97, 115–122] and has gained a broad acceptance in the computational chemistry community (as several of the chapters of this book show). QTAIM has been extensively applied to calculated and experimental electron densities [96] to predict and interpret molecular properties at an atomic resolution, including for example, heats of formation [123], magnetic susceptibilities [124, 125], atomic electrostatic moments and polarizabilities [126, 127] Raman intensities [126–129], IR intensities [130, 131], electron localization and delocalization [132, 133], pK_a [134], biological and physicochemical properties of the amino acids [135], protein retention times [136], HPLC column capacity factors [137], and NMR spin-spin coupling constants [138, 139]. The theory was also applied in the design of protein force fields by atom typing [140], to automate the search for pharmacophores and/or (re)active sites in a series of related molecules [141–145] and to reconstruct large molecules not amenable to direct computation [146–148] or easy crystallization [120] from transferable fragments.

In most of these studies, the analysis is applied to stationary points on the PES and, generally, in the absence of external perturbations such as external fields (with the exception of studies of polarizabilities). The advent of time-resolved crystallography, pioneered by scientists such as Philip Coppens, has brought the fourth dimension into the world of the experimental electron density [149–151]. A pump-probe approach is used to first excite the crystal with ultra-short laser or X-ray pulses followed by the “interrogating” pulse(s), the latter often polychromatic (Laue technique) to improve the time resolution. The work has generated “images” of the electron density and its deformation upon electronic excitation and allowed a “real-time” observation of the change in the geometry of molecules upon charge transfer induced by the external perturbation. Experimental activation energies have been measured by analyzing the temperature-dependence of the rate constant of photoisomerization [152]. Paralleling these exciting experimental advances on the theoretical side, studies that analyze the topology of the electron density as it evolved over the full PES landscape, or along the steepest path of descent from TS to the reactants and products valleys, the so-called “reaction path (RP)” [153–156], started to appear in the literature [157–160].

Further, there exists a bijective mapping between the points of a PES and the corresponding points belonging to each property surface such as “dipole moment or polarizability surfaces” [161, 162]. Examples of such surfaces for the reaction $F^\bullet + CH_4 \rightarrow HF + \bullet CH_3$, are displayed in Figure 4.

In the presence of an external laser field, and at the low frequency limit, the effective potential along the reaction path ($X + CH_4$, C_{3v} symmetry) can be approximated by [161]: $V = V(s) - \mu(s)\epsilon_0 \cos(\phi) - \frac{1}{2}\alpha_{zz}(s)\epsilon_0^2 \cos^2(\phi)$, where $V(s)$ is the laser-

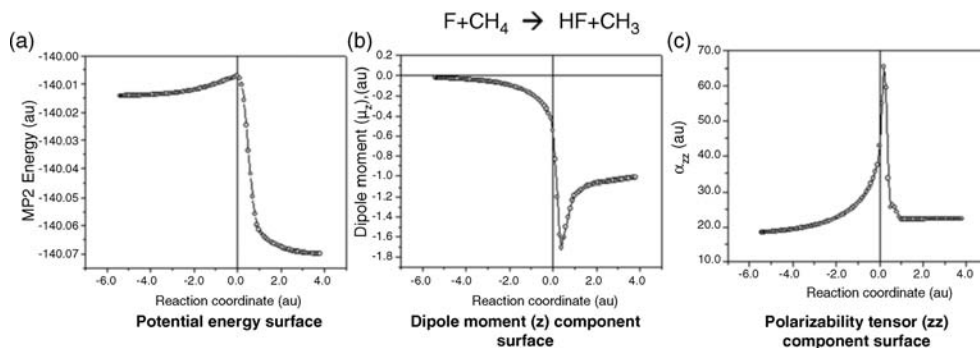


Figure 4 (a) Potential energy surface, (b) z-component of the dipole moment surface, and (c) zz-component of the polarizability tensor surface, for the reaction between a fluorine atom and methane (Adapted from Ref. [161] with permission from the American Institute of Physics).

free *ab initio* potential, $\mu(s)$ and $\alpha(s)$ are the dipole moment and polarizability components along the C_3 axis, and ϕ the phase.

With a proper choice of phase, the coupling between the field and the peaks in the dipole moment and polarizability surfaces can result in the *inversion* of the transition state into a bound state when $X = Cl$, and significantly reduce the height of the energy barrier in the case of $X = F$.

These results suggest that the evolution of properties that accompany the excursions of the system on the PES landscape are important not only for insight into chemical reactivity, kinetics, and thermodynamics of reactions, but also because of the potential use in the coherent control of reaction kinetics and dynamics through interferences with external fields. The writer's former postdoctoral supervisor, Professor John C. Polanyi, summed it up in his Nobel Lecture [163]:

"In closing I mention two further approaches which could assist materially in the quest for understanding of the choreography of chemical reaction. In the first, attempts are being made to observe the molecular partners while they are, so to speak, on the stage, rather than immediately prior to and following the reactive dance . . . In the second novel approach the intention, stated a little grandiosely, is to have a hand in writing the script according to which the dynamics occurs. . ."

The time appears to be ripe to extend the analysis of the topology of the electron density in the fourth dimension "on the stage" and influence the script of the molecular dance.³⁾

3) The writer has been analyzing the *atomic contributions to energies of reactions* and the *atomic contributions to activation energy barriers* since 2005. The latter interest constitutes an extension of his former studies of the atomic

partitioning of the BDE and of the energies of reactions [169–171], of the barrier for rotation in biphenyl [172], and of $X + CH_4$ reactions [161, 162].

An Apology to the Reader

This writer is neither a historian of science nor an expert in every field that was touched upon in these introductory remarks. The historical approach was chosen to set the tone for this collective work and to put Quantum Biochemistry in historical and scientific contexts. The highlights in this contextual introduction are necessary biased, incomplete, and, likely, at times imprecise. Because of that and because of space limitations, there is no doubt that important milestones, references, names of key scientists, and other contributions of those scientists who are named, have been omitted. The writer seeks the forgiveness of the reader for these unavoidable biases, errors, and omissions. Those who are interested in the history of chemistry can find better and comprehensive accounts elsewhere [164–167].

Book Contents

The book is organized in five logical parts. **Part I** is devoted to novel theoretical, computational, and experimental developments. In **Chapter 1**, Huang, Massa, and Karle review the biological applications of their Kernel Energy Method (Quantum Crystallography), whereby experiment and theory are combined to obtain the wavefunctions of biological macromolecules. Clemente, Vreven, and Frisch of GAUSSIAN, Inc., contributed **Chapter 2** in which they provide an excellent tutorial on the ONIOM method paying particular attention to practical guidelines and common pitfalls. Modeling enzymatic reactions in metalloenzymes and in photobiology is the subject of **Chapter 3** in which Chung, Li, and Morokuma show how to use a combination of quantum mechanical and QM/MM methods to obtain physically and biologically meaningful answers. **Chapter 4**, contributed by Tomasi, Cappelli, Mennucci, and Cammi, builds from the molecular electrostatic potentials to solvation models and closing with photophysical processes of biological significance. Finally, Liu, Burger, Dey, Sarkar, Janicki, and Ayers review their new method for the fast determination of reaction paths to elucidate complex reaction mechanisms in **Chapter 5**.

Part II focuses on key biological molecules and building blocks such as nucleic acids, amino acids, and peptides, as well as their interactions. In **Chapter 6**, Roy and Schleyer present complete reaction pathways explaining the mode of combinations of hydrogen cyanide molecules to form the nucleic acid base adenine under prebiotic and interstellar conditions. The effect of ionization on hydrogen bonding and proton transfer in DNA base pairs, amino acids and peptides is the topic of **Chapter 7** by Rodríguez-Santiago, Noguera, Bertran, and Sodupe. Kryachko's **Chapter 8** is about nano-biochemistry, exploring the interactions of gold atoms and clusters with DNA. **Chapter 9** by Rutledge and Wetmore reviews non-covalent DNA–protein interactions and their significance. Bader and Cortés-Guzmán examine the role of the virial field, in the context of QTAIM, in accounting for the transferability upon DNA base-pairing in **Chapter 10**. The next chapter, **Chapter 11**, by Mosquera, Moa, Estévez, Mandado, and Graña, investigates the origin of the ubiquitous stacking interactions in terms of

the topology of the electron density. The following three chapters deal with the properties of the amino acids. In **Chapter 12** Kassimi and Thakkar contrast, and compare the performance of, additive models and the *ab initio* calculations of the polarizabilities of the amino acids. This is followed by a contribution from Bohórquez, Cárdenas, Matta, Boyd, and Patarroyo, **Chapter 13**, in which the results of quantum chemical calculations are used as descriptors to yield a “physicochemical” classification of the amino acids into related classes and sub-classes. **Chapter 14** by Matta, the last one dealing with the amino acids, shows how the electron density of the atoms composing the genetically-encoded amino acids is related to the genetic code, protein stability, and several other (physicochemical) properties. This section ends with **Chapter 15** by Matta and Arabi in which the authors review a study where the energy storage in ATP’s “high energy phosphate bonds” is investigated at atomic resolution through the tracking of the changes in atomic energies upon hydrolysis.³

Part III includes studies on reactivity, catalysis, reaction paths and reaction mechanisms. The opening chapter of this section, **Chapter 16** written by Massa, Matta, Yonath, and Karle, explores the transition state, reaction path, and reaction mechanism of the peptide bond formation in the ribosome during the elongation step of protein synthesis. In **Chapter 17**, Bucher, Masson, Arey, and Rothlisberger use hybrid QM/MM to simulate enzyme-catalyzed DNA repair reactions. Rodriguez reviews the electronic structure of spin-coupled di-iron-oxoproteins in **Chapter 18**. Accurate description of spin states and its implications in catalysis is the topic of **Chapter 19** authored by Swart, Güell, and Solà. This is followed by **Chapter 20** on selenium biochemistry by Pearson and Boyd. In **Chapter 21**, Dal Peraro, Vila, and Carloni review computational and experimental studies of the mechanism of catalysis by metallo β -lactamase enzymes. 8-Epiconfertifin is then used as a case study in the exploration of the terminal biogenesis of sesquiterpenes in **Chapter 22** written by Barquera-Lozada and Cuevas. The final chapter in this section, **Chapter 23** by Llano and Gauld, investigates the effect of the size of the computational model of the active site on the emerging mechanistic picture of enzyme catalysis.

Part IV has a more “applied” flavor as it focuses on the uses of quantum biochemistry as a tool in the pharmacological, medical, and pharmaceutical sciences, especially in the domain of the conceptualization and design of new drugs and therapeutic agents. The first chapter in this section, **Chapter 24** by Popelier, reviews his method termed “Quantum Topological Molecular Similarity” (or QTMS). In **Chapter 25**, Gubskaya presents a critical review of the quantum chemical descriptors commonly used in studies of quantitative structure-to-activity/property relationship (QSAR/QSPR). **Chapter 26** by Gkionis, Hicks, Robertazzi, Hill, and Platts is a review on the role, structure, and activation of complexes of platinum as anti-cancer drugs. The next three chapters in this section are about the protein folding disease par excellence, namely, Alzheimer’s Disease (AD). **Chapter 27** written by Weaver reviews his group’s quantum biochemical searches for a cure to this disease. Darvesh, Pottie, McDonald, Martin, and Darvesh explore therapies to this disease by targeting Butyrylcholinesterase in **Chapter 28**. Finally, Rauk, in **Chapter 29**, discusses the relevance of reduction potentials of peptide-bound Cu^{2+} to AD and also to Prion Diseases, another example of a protein folding disease. In the final chapter of this

section, **Chapter 30**, Musa and Eriksson investigate the mechanisms of photo-degradation of non-steroidal anti-inflammatory drugs (NSAID).

Part V is written by three philosophers of science who have strong interest in quantum biochemistry. In **Chapter 31**, Stamos presents powerful arguments for and against the “trickling up” of the quantum indeterminism of individual acts of spontaneous mutation, brought about through Löwdin’s mechanism, to the macroscopic evolutionary level. In the closing chapter of the book, **Chapter 32**, Llored and Bitbol present a condensed and scholarly reflective essay on the meaning of molecular orbitals in a wider epistemological context with particular reference to Quantum Biochemistry.

Acknowledgment

The writer thanks Professors Lou Massa, Paul Ayers, for discussions and suggestions and Professor Anna Small for her corrections to the manuscript. Professor Massa has brought the historical events at the 1860 Karlsruhe Congress to the writer’s attention.

References

- 1 L. J. Gay-Lussac; Sur la combinaison des substances gazeuses, les unes avec les autres. *Mémoires de la Société de physique et de chimie de la Société d’Arcueil*, tome 2 1809, 2, 207–234 (with two tables, pp. 252–253).
- 2 R. A. Smith *Memoir of John Dalton and History of the Atomic Theory Up to His Time*; H. Bailliere: London, 1856.
- 3 A. Avogadro; Essai d’une manière de déterminer les masses relatives des molécules élémentaires des corps, et les proportions selon lesquelles elles entrent dans ces combinaisons. *Journal de physique, de chimie et d’histoire naturelle* 1811, 58–76.
- 4 S. Cannizzaro *Sketch of a Course of Chemical Philosophy (English Translation from the 1858 Italian Edition: Sunto di un corso di Filosofia chimica)*; The Alembic Club and University of Chicago Press: Edinburgh, Chicago, 1911.
- 5 Wikipedia. Karlsruhe Congress. Web Page, http://en.wikipedia.org/wiki/Karlsruhe_Congress, accessed 2009.
- 6 M. G. Fayershtein: The Evolution of the Theory of Valency. (V. I. Kuznetsov, Ed.) *Theory of Valency in Progress (English Translation)*; Mir Publishers: Moscow.
- 7 G. N. Lewis; The atom and the molecule. *J. Am. Chem. Soc.* 1916, 38, 762–785.
- 8 G. N. Lewis *Valence and the Structure of Atoms and Molecules*; Dover Publications, Inc.: New York, 1966.
- 9 P. A. M. Dirac; Quantum mechanics of many-electron systems. *Proc. Roy. Soc., Ser. A* 1929, 123, 714–733.
- 10 M. Born, R. Oppenheimer; Zur quantentheorie der moleküle (On the quantum theory of molecules). *Ann. Phys.* 1927, 84, 457–484.
- 11 I. N. Levine *Quantum Chemistry, (Sixth Edition)*; Pearson Prentice Hall: Upper Saddle River, New Jersey, 2009.
- 12 A. Szabo, N. S. Ostlund *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications, Inc.: New York, 1989.
- 13 H. Eyring, M. Polanyi; On simple gas reaction. *Z. physik. Chem. B* 1931, 12, 279–311.
- 14 S. Glasstone, K. J. Laidler, H. Eyring *The Theory of Rate Processes (First Edition)*; McGraw-Hill Book Company, Inc.: New York, 1941.
- 15 H. Eyring, E. M. Eyring *Modern Chemical Kinetics*; Reinhold Publishing Corporation: New York, 1963.

- 16 M. H. Back, K. J. E. Laidler (Eds.) *Selected Readings in Chemical Kinetics*; Pergamon Press, Ltd.: Oxford, 1967.
- 17 K. J. Laidler *Chemical Kinetics*; Harper and Row, Publishers: Cambridge, 1987.
- 18 P. Ignotus, J. Polanyi, E. Schmid, H. Eyring, E. D. Bergmann, A. Koestler, C. V. Wedgwood, J. R. Ravetz, J. R. Baker, R. Aron, E. Shils, I. Kristol, E. Devons, D. M. Mackinnon, E. Sewell, M. Grene, M. Calvin, E. P. Wigner *The Logic of Personal Knowledge: Essays Presented to Michael Polanyi on his Seventieth Birthday, 11th March 1961*; Routledge and Kegan Paul: London, 1961.
- 19 L. Pauling *The Nature of the Chemical Bond, (Third Ed.)*; Cornell University Press: Ithaca, N.Y., 1960.
- 20 C. A. Coulson *Valence, (Second Edition)*; Oxford University Press: New York, 1961.
- 21 R. McWeeny *Coulson's Valence*; The English Language Book Society and Oxford University Press: Oxford, 1979.
- 22 S. Shaik, P. C. Hiberty *A Chemist's Guide to Valence Bond Theory*; John Wiley and Sons, Inc.: New Jersey, 2008.
- 23 G. A. Gallup *Valence Bond Methods*; Cambridge University Press: Cambridge, 2002.
- 24 D. R. Hartree; The wave mechanics of an atom with a non-Coulomb central field. Part I. Theory and methods. *Proc. Cambridge Phil. Soc.* 1928, **24**, 89–110.
- 25 D. R. Hartree; The wave mechanics of an atom with a non-Coulomb central field. Part II. Some results and discussion. *Proc. Cambridge Phil. Soc.* 1928, **24**, 111–132.
- 26 D. R. Hartree *The Calculation of Atomic Structures*; John Wiley and Sons, Inc.: New York, 1957.
- 27 J. C. Slater; Note on Hartree's method. *Phys. Rev.* 1930, **35**, 210–211.
- 28 V. Fock; Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems. *Z. Physik* 1930, **61**, 126–148.
- 29 S. M. Blinder; Basic concepts of self-consistent-field theory. *Am. J. Phys.* 1965, **33**, 431–443.
- 30 C. C. J. Roothaan; New developments in molecular orbital theory. *Rev. Mod. Phys.* 1951, **23**, 69–89.
- 31 G. G. Hall; The molecular orbital theory of chemical valency. VIII. A method of calculating ionization potentials. *Proc. Roy. Soc., Ser. A* 1951, **A205**, 541–552.
- 32 H. Hettema *Quantum Chemistry: Classic Scientific Papers*; World Scientific: Singapore, 2000.
- 33 R. Hoffmann, S. Shaik, P. C. Hiberty; A conversation on VB vs MO theory: A never-ending rivalry? *Acc. Chem. Res.* 2003, **36**, 750–756.
- 34 R. G. Parr, W. Yang *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: Oxford, 1989.
- 35 T. Ziegler; Approximate density functional theory as a practical tool in molecular energetics and dynamics. *Chem. Rev.* 1991, **91**, 651–667.
- 36 W. Koch, M. C. Holthausen *A Chemist's Guide to Density Functional Theory, (Second Edition)*; Wiley-VCH: New York, 2001.
- 37 J. P. Perdew, A. Ruzsinszky, L. A. Constantin, J. Sun, G. I. Csonka; Some fundamental issues in ground-state density functional theory: A guide for the perplexed. *J. Chem. Theory Comput.* 2009, **5**, 902–908.
- 38 P. Hohenberg, W. Kohn; Inhomogeneous electron gas. *Phys. Rev. B* 1964, **136**, 864–871.
- 39 W. Kohn, L. J. Sham; Self consistent equations including exchange and correlation effects. *Phys. Rev. A* 1965, **140** (4A), 1133–1138.
- 40 F. Geerlings, F. De Proft, W. Langenaeker; Conceptual Density Functional Theory. *Chem. Rev.* 2003, **103**, 1793–1874.
- 41 F. Geerlings, F. De Proft; Conceptual DFT: the chemical relevance of higher response functions. *Phys. Chem. Chem. Phys. (PCCP)* 2008, **10**, 3028–3042.
- 42 W. J. Hehre, L. Radom, J. A. Pople, P. v. R. Schleyer *Ab Initio Molecular Orbital Theory*; Wiley-Interscience: New York, 1986.
- 43 J. B. Foresman, A. Frisch *Exploring Chemistry with Electronic Structure Methods, (Second Edition)*; Gaussian, Inc.: Pittsburgh, 1996.
- 44 P. v.-R. Schleyer (Ed.) *Encyclopedia of Computational Chemistry*; John Wiley and Sons: Chichester, UK, 1998.

- 45 S. M. Bachrach *Computational Organic Chemistry*; John Wiley and Sons, Inc.: Hoboken, New Jersey, 2007.
- 46 Frisch, M. J., Trucks, G. W., Schlegel, H. B., *et al.*; Gaussian Inc.: Pittsburgh PA, 2003.
- 47 M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis, J. A. Montgomery; General atomic and molecular electronic-structure system (GAMESS). *J. Comput. Chem.* 1993, **14**, 1347–1363.
- 48 Saunders, V. R., Dovesi, R., Roetti, C., Orlando, R., Zicovich-Wilson, C. M., Harrison, N. M., Doll, K., Civalieri, B., Bush, I. J., D'Arco, Ph., and Llunell, M.; 2003.
- 49 Pisani C. (Ed.) *Quantum-Mechanical Ab-initio Calculations of the Properties of Crystalline Materials*; Springer-Verlag: Berlin, 1996.
- 50 J. W. Ponder, D. A. Case; Force fields for protein simulations. *Adv. Protein Chem.* 2003, **66**, 27–85.
- 51 J.-P. Doucet, J. Weber *Computer-Aided Molecular Design: Theory and Applications*; Academic Press, Ltd.: London, 1996.
- 52 T. Schlick *Molecular Modeling and Simulation: An Interdisciplinary Guide*; Springer: New York, 2002.
- 53 D. Frenkel, B. Smit *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press: New York, 2002.
- 54 C. J. Cramer *Essentials of Computational Chemistry: Theories and Models*; John Wiley & Sons, Ltd.: New York, 2002.
- 55 J. D. Watson, F. H. C. Crick; Genetical implications of the structure of deoxyribose nucleic acid. *Nature* 1953, **171**, 964–967.
- 56 J. D. Watson *Molecular Biology of the Gene (Second Edition)*; W. A. Benjamin, Inc.: New York, 1970.
- 57 J. D. Watson, F. H. C. Crick; A structure for deoxyribose nucleic acid. *Nature* 1953, **171**, 737–738.
- 58 J. D. Watson *The Double Helix: A Personal Account of the Discovery of the Structure of DNA (Edited by G. S. Stent)*; W. W. Norton & Co.: New York, 1980.
- 59 R. E. Franklin, R. G. Gosling; Molecular configuration in sodium thymonucleate. *Nature* 1953, **171**, 740–741.
- 60 E. Schrödinger *What is Life?*; Cambridge University Press: Cambridge, 1944.
- 61 B. Pullman, A. Pullman *Quantum Biochemistry*; Interscience Publishers: New York, 1963.
- 62 P.-O. Löwdin; Proton tunneling in DNA and its biological implications. *Rev. Mod. Phys.* 1963, **35**, 721–733.
- 63 P.-O. Löwdin; Quantum genetics and the aperiodic solid: some aspects on the biological problems of heredity, mutation, aging, and tumors in view of the quantum theory of the DNA molecule. *Adv. Quantum Chem.* 1965, **2**, 213–360.
- 64 E. J. Brändas, E. S. Kryachko (Eds.) *Fundamental World of Quantum Chemistry: A Tribute to the Memory of Per-Olov Löwdin*; Kluwer Academic Publishers: Dordrecht, 2003.
- 65 K. Fukui, K. Morokuma, C. Nagata; A molecular orbital treatment of phosphate bonds of biochemical interest. I. Simple LCAO MO treatment. *Bull. Chem. Soc. Jpn.* 1960, **33**, 1214–1219.
- 66 K. Fukui, A. Imamura, C. Nagata; A molecular orbital treatment of phosphate bonds of biochemical interest. II. Metal chelates of adenosine triphosphate. *Bull. Chem. Soc. Jpn.* 1963, **36**, 1450–1453.
- 67 A. Pullman, B. Pullman Electronic structure and carcinogenic activity of aromatic molecules: New developments. *Advances in Cancer Research (Volume 3)*; Academic Press: New York, 1955, p 117–169.
- 68 B. Pullman, A. Pullman; Electron-donor or electron-acceptor properties and carcinogenic activity of organic molecules. *Nature (London)* 1963, **199**, 467–469.
- 69 B. Pullman, A. Pullman; Submolecular structure of the nucleic acids. *Nature (London)* 1961, **189**, 725–727.
- 70 B. Pullman, P. Claverie, J. Caillet; Intermolecular forces in association of purines with polybenzenoid hydrocarbons. *Science* 1965, **147**, 1305–1307.
- 71 B. Pullman, A. Pullman; Electronic delocalization and biochemical

- evolution. *Nature (London)* 1962, **196**, 1137–1142.
- 72 D. L. Beveridge, R. E. Lavery *Theoretical Biochemistry and Molecular Biophysics (Vol. 1: DNA; Vol. 2: Proteins)*; Adenine Press: Schenectady, NY, 1991.
 - 73 L. A. Eriksson (Ed.) *Theoretical Biochemistry - Processes and Properties of Biological Systems*; Elsevier Science B. V.: Amsterdam, 2001.
 - 74 O. M. Becker, A. D. MacKerell Jr., B. Roux, M. Watanabe *Computational Biochemistry and Biophysics*; Marcel Dekker, Inc.: New York, 2001.
 - 75 A. Warshel, G. Naray-Szabo *Computational Approaches to Biochemical Reactivity*; Kluwer Academic Publishers: 2002.
 - 76 P. Carloni, F. E. Alber (Eds.) *Quantum Medicinal Chemistry*; Wiley-VCH: Weinheim, 2003.
 - 77 H. M. Senn, W. Thiel; QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed.* 2009, **48**, 1198–1229.
 - 78 A. Warshel *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; John Wiley and Sons, Inc.: New York, 1991.
 - 79 L. Huang, L. Massa, J. Karle; Kernel energy method applied to vesicular stomatitis virus nucleoprotein. *Proc. Natl. Acad. Sci. USA* 2009, **106**, 1731–1736.
 - 80 W. L. Clinton, A. J. Galli, L. J. Massa; Direct determination of pure-state density matrices. II. Construction of constrained idempotent one-body densities. *Phys. Rev.* 1969, **177**, 7–12.
 - 81 W. L. Clinton, L. J. Massa; Determination of the electron density matrix from x-ray diffraction data. *Phys. Rev. Lett.* 1972, **29**, 1363–1366.
 - 82 L. Massa, L. Huang, J. Karle; Quantum crystallography and the use of kernel projector matrices. *Int. J. Quantum. Chem: Quantum Chem. Symp.* 1995, **29**, 371–384.
 - 83 L. Huang, L. Massa, J. Karle; Kernel energy method: Application to DNA. *Biochemistry* 2005, **44**, 16747–16752.
 - 84 L. Huang, L. Massa, J. Karle; The Kernel Energy Method: Application to a tRNA. *Proc. Natl. Acad. Sci. USA* 2006, **103**, 1233–1237.
 - 85 A. Gindulyte, A. Bashan, I. Agmon, L. Massa, A. Yonath, J. Karle; The transition state for the formation of the peptide bond in the ribosome. *Proc. Natl. Acad. Sci. USA* 2006, **103**, 13327–13332.
 - 86 L. Huang, L. Massa, J. Karle; Kernel energy method: Application to insulin. *Proc. Natl. Acad. Sci. USA* 2005, **102**, 12690–12693.
 - 87 X. Zhang, T. J. Green, J. Tsao, S. Qiu, M. Luo; Role of intermolecular interactions of vesicular stomatitis virus nucleoprotein in RNA encapsidation. *J. Virol.* 2008, **82**, 674–682.
 - 88 C. J. Cramer, D. G. Truhlar Continuum solvation models: Classical and quantum mechanical implementations. *Reviews in Computational Chemistry (Vol. 6)*; VCH Publishers: New York, 1995, pp 1–72.
 - 89 C. J. Cramer, D. G. Truhlar; Implicit solvation models: Equilibria, structure, spectra, and dynamics. *Chem. Rev.* 1999, **99**, 2161–2200.
 - 90 J. Tomasi; Thirty years of continuum solvation chemistry: A review, and prospects for the near future. *Theor. Chem. Acc.* 2004, **112**, 184–203.
 - 91 J. Tomasi, B. Mennucci, R. Cammi; Quantum mechanical continuum solvation models. *Chem. Rev.* 2005, **105**, 2999–3093.
 - 92 A. Yonath The quest for high resolution phasing for large macromolecular assemblies exhibiting severe non-isomorphism, extreme beam sensitivity and no internal symmetry. In: *Structure and Dynamics of Biomolecules: Neutron and Synchrotron Radiation for Condensed Matter Studies*; (E. Fanchon, et al. Eds), Oxford University Press: Oxford, 2000.
 - 93 G. H. Stout, L. H. Jensen *X-Ray Structure Determination: A Practical Guide, (Second Edition)*; John-Wiley and Sons: New York, 1989.
 - 94 R. F. Stewart; Electron population analysis with rigid pseudoatoms. *Acta Cryst.* 1976, **A32**, 565–574.
 - 95 N. K. Hansen, P. Coppens; Testing aspherical atom refinement on small molecules data sets. *Acta Cryst.* 1978, **A34**, 909–921.

- 96 P. Coppens *X-ray Charge Densities and Chemical Bonding*; Oxford University Press, Inc.: New York, 1997.
- 97 T. S. Koritsanszky, P. Coppens; Chemical applications of X-ray charge-density analysis. *Chem. Rev.* 2001, **101**, 1583–1628.
- 98 R. F. W. Bader, G. A. Jones; The electron density distributions in hydride molecules, III, The hydrogen fluoride molecule. *Can. J. Chem.* 1963, **41**, 2251–2264.
- 99 R. F. W. Bader, G. A. Jones; The electron density distribution in hydride molecules. The ammonia molecule. *J. Chem. Phys.* 1963, **38**, 2791–2802.
- 100 R. F. W. Bader, G. A. Jones; The electron density distributions in hydride molecules, I, The water molecule. *Can. J. Chem.* 1963, **41**, 586–606.
- 101 P.-O. Löwdin; Correlation problem in many-electron quantum mechanics I. Review of different approaches and discussion of some current ideas. *Adv. Chem. Phys.* 1959, **2**, 207–322.
- 102 H. Hellmann *Einführung in die Quantenchemie*; Deuticke: Leipzig and Vienna, 1937.
- 103 R. P. Feynman; Forces in molecules. *Phys. Rev.* 1939, **56**, 340–343.
- 104 R. F. W. Bader *Atoms in Molecules: A Quantum Theory*; Oxford University Press: Oxford, U.K., 1990.
- 105 P. L. A. Popelier *Atoms in Molecules: An Introduction*; Prentice Hall: London, 2000.
- 106 Matta, C. F., Boyd, R. J. (Eds.) *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*; Wiley-VCH: Weinheim, 2007.
- 107 R. F. W. Bader; The quantum mechanical basis of conceptual chemistry. *Monatsh Chem* 2005, **136**, 819–854.
- 108 R. F. W. Bader, T. T. Nguyen-Dang; Quantum theory of atoms in molecules - Dalton revisited. *Adv. Quantum Chem.* 1981, **14**, 63–124.
- 109 R. F. W. Bader, T. T. Nguyen-Dang, Y. Tal; A topological theory of molecular structure. *Rep. Prog. Phys.* 1981, **44**, 893–948.
- 110 R. F. W. Bader; Bond paths are not chemical bond. *J. Phys. Chem. A* 2009, **113**, 10391–10396.
- 111 R. F. W. Bader; Principle of stationary action and the definition of a proper open system. *Phys. Rev. B* 1994, **49**, 13348–13356.
- 112 R. G. Parr, P. W. Ayers, R. F. Nalewajski; What is an atom in a molecule. *J. Phys. Chem. A* 2005, **109**, 3957–3959.
- 113 C. F. Matta, R. F. W. Bader; An experimentalist's reply to "What is an atom in a molecule?". *J. Phys. Chem. A* 2006, **110**, 6365–6371.
- 114 P. L. A. Popelier Quantum chemical topology: On bonds and potentials. In: *Intermolecular Forces and Clusters, Structure and Bonding*, Vol. 115; (D. J. Wales, Ed.), Springer: 2005, pp 1–56.
- 115 E. Espinosa, E. Molins, C. Lecomte; Hydrogen bond strengths related by topological analyses of experimentally observed electron densities. *Chem. Phys. Lett.* 1998, **285**, 170–173.
- 116 D. Housset, F. Benabicha, V. Pichon-Pesme, C. Jelsch, A. Maierhofer, S. David, J. C. Fontecilla-Camps, C. Lecomte; Towards the charge-density study of proteins: A room-temperature scorpion-toxin structure at 0.96 Å resolution as a first test case. *Acta Cryst.* 2000, **D56**, 151–160.
- 117 F. Benabicha, V. Pichon-Pesme, C. Jelsch, C. Lecomte, A. Khmou; Experimental charge density and electrostatic potential of glycyl-L-threonine dihydrate. *Acta Cryst.* 2000, **B56**, 155–165.
- 118 L. Leherter, B. Guillot, D. P. Vercauteren, V. Pichon-Pesme, C. Jelsch, A. Lagoutte, C. Lecomte Topological analysis of proteins as derived from medium and high-resolution electron density: Applications to electrostatic properties. In: *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*; (C. F. Matta and R. J. Boyd, Eds.), Wiley-VCH: Weinheim, 2007, pp 285–315.
- 119 B. Dittrich, T. Koritsanszky, M. Grosche, W. Scherer, R. Flaig, A. Wagner, H. G. Krane, H. Kessler, C. Riemer, A. M. M. Schreurs, P. Luger; Reproducibility and transferability of topological properties; experimental charge density of the hexapeptide cyclo-(D, L-Pro)₂-(L-Ala)₄ monohydrate. *Acta Cryst. B* 2002, **58**, 721–727.

- 120 S. Scheins, M. Messerschmidt, P. Luger; Submolecular partitioning of morphine hydrate based on its experimental charge density at 25 K. *Acta Cryst. B* 2005, **61**, 443–448.
- 121 P. Luger, B. Dittrich Fragment transferability studied theoretically and experimentally with QTAIM - Implications for electron density and invarium modeling. *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*; (C. F. Matta and R. J. Boyd, Eds.), Wiley-VCH: Weinheim, 2007, pp 317–341.
- 122 P. Luger; Fast electron density methods in the life sciences - a routine application in the future? *Org. Biomolec. Chem.* 2007, **5**, 2529–2540.
- 123 K. B. Wiberg, R. F. W. Bader, C. D. H. Lau; Theoretical analysis of hydrocarbons properties. 2. Additivity of group properties and the origin of strain energy. *J. Am. Chem. Soc.* 1987, **109**, 1001–1012.
- 124 T. A. Keith, R. F. W. Bader; Calculation of magnetic response properties using atoms in molecules. *Chem. Phys. Lett.* 1992, **194**, 1–8.
- 125 T. A. Keith, R. F. W. Bader; Use of electron charge and current distributions in the determination of atomic contributions to magnetic properties. *Int. J. Quantum Chem.* 1996, **60**, 373–379.
- 126 R. F. W. Bader, T. A. Keith, K. M. Gough, K. E. Laidig; Properties of atoms in molecules: additivity and transferability of group polarizabilities. *Mol. Phys.* 1992, **75**, 1167–1189.
- 127 K. M. Gough, M. M. Yacowar, R. H. Cleve, J. R. Dwyer; Analysis of molecular polarizabilities and polarizability derivatives in H₂, N₂, F₂, CO, and HF, with the theory of atoms in molecules. *Can. J. Chem.* 1996, **74**, 1139–1144.
- 128 K. M. Gough, H. K. Srivastava, K. Belohorcová; Molecular polarizability and polarizability derivatives in cyclohexane analyzed with the theory of atoms in molecules. *J. Phys. Chem.* 1994, **98**, 771–776.
- 129 K. M. Gough, H. K. Srivastava; Electronic charge flow and Raman trace scattering intensities for CH stretching vibrations in *n*-pentane. *J. Phys. Chem.* 1996, **100**, 5210–5216.
- 130 R. L. A. Haiduke, R. E. Bruns; An atomic charge-charge flux-dipole flux atom-in-molecule decomposition for molecular dipole-moment derivatives and infrared fundamental intensities. *J. Phys. Chem. A* 2005, **109**, 2680–2688.
- 131 J. V. da Silva, R. L. A. Haiduke, R. E. Bruns; QTAIM Charge-charge flux-dipole flux models for the infrared fundamental intensities of the fluorochloromethanes. *J. Phys. Chem. A* 2006, **110**, 4839–4845.
- 132 X. Fradera, M. A. Austen, R. F. W. Bader; The Lewis model and beyond. *J. Phys. Chem. A* 1999, **103**, 304–314.
- 133 Y.-G. Wang, C. F. Matta, N. H. Werstiuk; Comparison of localization and delocalization indices obtained with Hartree-Fock and conventional correlated methods: Effect of Coulomb correlation. *J. Comput. Chem.* 2003, **24**, 1720–1729.
- 134 K. R. Adam; New density functional and atoms in molecules method of computing relative p*K*_a values in solution. *J. Phys. Chem. A* 2002, **106**, 11963–11972.
- 135 C. F. Matta, R. F. W. Bader; Atoms-in-molecules study of the genetically-encoded amino acids. III. Bond and atomic properties and their correlations with experiment including mutation-induced changes in protein stability and genetic coding. *Proteins: Struct. Funct. Genet.* 2003, **52**, 360–399.
- 136 M. Song, C. M. Breneman, J. Bi, N. Sukumar, K. P. Bennett, S. Cramer, N. Tugcu; Prediction of protein retention times in anion-exchange chromatography systems using support vector regression. *J. Chem. Inf. Comput. Sci.* 2002, **42**, 1347–1357.
- 137 C. M. Breneman, M. Rhem; QSPR Analysis of HPLC column capacity factors for a set of high-energy materials using electronic van der Waals surface property descriptors computed by transferable atom equivalent method. *J. Comput. Chem.* 1997, **18**, 182–197.
- 138 C. F. Matta, J. Hernández-Trujillo, R. F. W. Bader; Proton spin-spin coupling and electron delocalisation. *J. Phys. Chem. A* 2002, **106**, 7369–7375.

- 139 N. Castillo, C. F. Matta, R. J. Boyd; Fluorine-Fluorine spin-spin coupling constants: Correlations with the delocalization index and with the internuclear separation. *J. Chem. Inf. Mod.* 2005, **45**, 354–359.
- 140 P. L. A. Popelier, F. M. Aicken; Atomic properties of amino acids: Computed atom types as a guide for future force-field design. *CHEMPYSCHEM* 2003, **4**, 824–829.
- 141 P. L. A. Popelier; Quantum Molecular Similarity. 1. BCP space. *J. Phys. Chem. A* 1999, **103**, 2883–2890.
- 142 S. E. O'Brien, P. L. A. Popelier; Quantum molecular similarity. Part 2: the relation between properties in BCP space and bond length. *Can. J. Chem.* 1999, **77**, 28–36.
- 143 S. E. O'Brien, P. L. A. Popelier; Quantum molecular similarity. 3. QTMS descriptors. *J. Chem. Inf. Comput. Sci.* 2001, **41**, 764–775.
- 144 U. A. Chaudry, P. L. A. Popelier; Estimation of pKa using quantum topological molecular similarity descriptors: Application to carboxylic acids, anilines and phenols. *J. Org. Chem.* 2004, **69**, 233–241.
- 145 P. L. A. Popelier, F. M. Aicken; Atomic properties of selected biomolecules: Quantum topological atom types of carbon occurring in natural amino acids and derived molecules. *J. Am. Chem. Soc.* 2003, **125**, 1284–1292.
- 146 C. M. Breneman, T. R. Thompson, M. Rhem, M. Dung; Electron density modeling of large systems using the transferable atom equivalent method. *Comput. Chem.* 1995, **19**, 161–179.
- 147 R. F. W. Bader, C. F. Matta, F. J. Martín; Atoms in medicinal chemistry. *Medicinal Quantum Chemistry*, (P. Carloni and F. Alber, Eds.), Wiley-VCH: Weinheim, 2003, pp 201–231.
- 148 C. F. Matta; Theoretical reconstruction of the electron density of large molecules from fragments determined as proper open quantum systems: the properties of the oripavine PEO, enkephalins, and morphine. *J. Phys. Chem. A* 2001, **105**, 11088–11101.
- 149 P. Coppens, M. Pitak, M. Gembicky, M. Messerschmidt, S. Scheins, J. B. Benedict, S.-I. Adachi, T. Sato, S. Nozawa, K. Ichiiyanagi, M. Chollet, S.-Y. Koshihara; The RATIO method for time-resolved Laue crystallography. *J. Synchrotron Rad.* 2009, **16**, 226–230.
- 150 I. Vorontsov, T. Graber, A. Kovalevsky, I. Novozhilova, M. Gembicky, Y.-S. Chen, P. Coppens; Capturing and analyzing the excited-state structure of a Cu(I) phenanthroline complex by time-resolved diffraction and theoretical calculations. *J. Am. Chem. Soc.* 2009, **131**, 6566–6573.
- 151 P. Coppens; The new photocrystallography. *Angew. Chem. Int. Ed.* 2009, **48**, 4280–4281.
- 152 S.-L. Zheng, C. M. L. Vande Velde, M. Messerschmidt, A. Volkov, M. Gembicky, P. Coppens; Supramolecular solids as a medium for single-crystal-to-single-crystal *E/Z*-photoisomerization: Kinetic study of the photoreactions of two Zn coordinated tiglic acid molecules. *Chem. Eur. J.* 2008, **14**, 706–713.
- 153 K. Fukui; A formulation of the reaction coordinate. *J. Phys. Chem.* 1970, **74**, 4161–4163.
- 154 K. Fukui; The path of chemical reactions - The IRC approach. *Acc. Chem. Res.* 1981, **14**, 363–368.
- 155 C. Gonzalez, H. B. Schlegel; An improved algorithm for reaction path following. *J. Chem. Phys.* 1989, **90**, 2154.
- 156 C. Gonzalez, H. B. Schlegel; Reaction path following in mass-weighted internal coordinates. *J. Phys. Chem.* 1990, **94**, 5523–5527.
- 157 M. García-Revilla, J. Hernández-Trujillo; Energetic and electron density analysis of hydrogen dissociation of protonated benzene. *Phys. Chem. Chem. Phys.* 2009, **11**, 8425–8432.
- 158 J. P. Salinas-Olvera, R. M. Gomez, F. Cortes-Guzman; Structural evolution: Mechanism of olefin insertion in hydroformylation reaction. *J. Phys. Chem. A* 2008, **112**, 2906–2912.
- 159 Y. Zeng, L. Meng, X. Li, S. Zheng; Topological characteristics of electron density distribution in $SSXY \Rightarrow XSSY$ (X or $Y = F, Cl, Br, I$) isomerization

- reactions. *J. Phys. Chem. A* 2007, **111**, 9093–9101.
- 160 Farrugia L. J., C. Evans, M. Tegel; Chemical bonds without “chemical bonding”? A combined experimental and theoretical charge density study on an iron trimethylenemethane complex. *J. Phys. Chem. A* 2006, **110**, 7952–7961.
- 161 A. D. Bandrauk, E. S. Sedik, C. F. Matta; Effect of absolute laser phase on reaction paths in laser-induced chemical reactions. *J. Chem. Phys.* 2004, **121**, 7764–7775.
- 162 A. D. Bandrauk, E. S. Sedik, C. F. Matta; Laser control of reaction paths in ion-molecule reactions. *Mol. Phys.* 2006, **104**, 95–102.
- 163 J. C. Polanyi; Some concepts in reaction dynamics (Nobel Lecture, 8 December, 1986). *Chem. Script.* 1987, **27**, 229–247.
- 164 B. Pullman *The Atom in the History of Human Thought*; Oxford University Press: Oxford, 2004.
- 165 E. R. Scerri *The Periodic Table: Its Story and Its Significance*; Oxford University Press: Oxford, 2006.
- 166 W. H. Brock *The Fontana History of Chemistry*, Fontana Press: London, 1993.
- 167 V. I. Kuznetsov *Theory of Valence in Progress (translated from the original 1977 Russian edition by A. Rosinkin)*; Mir Publishers: Moscow, 1980.
- 168 S. Pisana, M. Lazzeri, C. Casiraghi, K. S. Novoselov, A. K. Geim, A. C. Ferrari, F. Mauri; Breakdown of the adiabatic Born–Oppenheimer approximation in graphene. *Nature Materials* 2007, **6**, 198–201.
- 169 C. F. Matta, N. Castillo, R. J. Boyd; Atomic contributions to bond dissociation energies in aliphatic hydrocarbons. *J. Chem. Phys.* 2006, **125**, 204103_1–204103_13.
- 170 C. F. Matta, A. A. Arabi, T. A. Keith; Atomic Partitioning of the Dissociation Energy of the PO(H) Bond in Hydrogen Phosphate Anion (HPO_4^{2-}): Disentangling the Effect of Mg^{2+} . *J. Phys. Chem. A* 2007, **111**, 8864–8872.
- 171 A. A. Arabi, C. F. Matta; Where is energy stored in adenosine triphosphate? *J. Phys. Chem. A* 2009, **113**, 3360–3368.
- 172 J. Hernández-Trujillo, C. F. Matta; Hydrogen-hydrogen bonding in biphenyl revisited. *Struct. Chem.* 2007, **18**, 849–857.

Contents

Acknowledgment	<i>VII</i>
Congratulations to Professor Ada Yonath for Winning the 2009 Nobel Prize in Chemistry	<i>IX</i>
Introductory Reflections on Quantum Biochemistry: From Context to Contents	<i>XI</i>
<i>Chérif F. Matta</i>	
List of Contributors	<i>LI</i>

Vol I

Part One Novel Theoretical, Computational, and Experimental Methods and Techniques 1

1	Quantum Kernels and Quantum Crystallography: Applications in Biochemistry	3
	<i>Lulu Huang, Lou Massa, and Jerome Karle</i>	
1.1	Introduction	3
1.2	Origins of Quantum Crystallography (QCr)	4
1.2.1	General Problem of <i>N</i> -Representability	4
1.2.2	Single Determinant <i>N</i> -Representability	5
1.2.3	Example Applications of Clinton's Equations	7
1.2.3.1	Beryllium	7
1.2.3.2	Maleic Anhydride	9
1.3	Beginnings of Quantum Kernels	10
1.3.1	Computational Difficulty of Large Molecules	10
1.3.2	Quantum Kernel Formalism	11
1.3.3	Kernel Matrices: Example and Results	14
1.3.4	Applications of the Idea of Kernels	17
1.3.4.1	Hydrated Hexapeptide Molecule	17
1.3.4.2	Hydrated Leu ¹ -Zervamicin	18
1.4	Kernel Density Matrices Led to Kernel Energies	22
1.4.1	KEM Applied to Peptides	24

1.4.2	Quantum Models within KEM	29
1.4.2.1	Calculations and Results Using Different Basis Functions for the ADPGV7b Molecule	32
1.4.2.2	Calculations and Results Using Different Quantum Methods for the Zaib4 Molecule	34
1.4.2.3	Comments Regarding KEM	36
1.4.3	KEM Applied to Insulin	36
1.4.3.1	KEM Calculation Results	36
1.4.3.2	Comments Regarding the Insulin Calculations	38
1.4.4	KEM Applied to DNA	39
1.4.4.1	KEM Calculation Results	39
1.4.4.2	Comments Regarding the DNA Calculations	41
1.4.5	KEM Applied to tRNA	41
1.4.6	KEM Applied to Rational Design of Drugs	43
1.4.6.1	Importance of the Interaction Energy for Rational Drug Design	43
1.4.6.2	Sample Calculation: Antibiotic Drug in Complex (1O9M) with a Model Aminoacyl Site of the 30s Ribosomal Subunit	44
1.4.6.3	Comments Regarding the Drug–Target Interaction Calculations	46
1.4.7	KEM Applied to Collagen	47
1.4.7.1	Interaction Energies	47
1.4.7.2	Collagen 1A89	47
1.4.7.3	Comments Regarding the Collagen Calculations	50
1.4.8	KEM Fourth-Order Calculation of Accuracy	50
1.4.8.1	Molecular Energy as a Sum over Kernel Energies	50
1.4.8.2	Application to Leu ¹ -zervamicin of the Fourth-Order Approximation of KEM	51
1.4.9	KEM Applied to Vesicular Stomatitis Virus Nucleoprotein, 33 000 Atom Molecule	53
1.4.9.1	Vesicular Stomatitis Virus Nucleoprotein (2QVJ) Molecule	53
1.4.9.2	Hydrogen Bond Calculations	54
1.4.9.3	Comments regarding the 2QVJ Calculations	54
1.5	Summary and Conclusions	55
	References	57

2 **Getting the Most out of ONIOM: Guidelines and Pitfalls** 61

Fernando R. Clemente, Thom Vreven, and Michael J. Frisch

2.1	Introduction	61
2.2	QM/MM	62
2.3	ONIOM	63
2.4	Guidelines for the Application of ONIOM	65
2.4.1	Summary	72
2.5	The Cancellation Problem	72
2.6	Use of Point Charges	77
2.7	Conclusions	81
	References	82

3	Modeling Enzymatic Reactions in Metalloenzymes and Photobiology by Quantum Mechanics (QM) and Quantum Mechanics/Molecular Mechanics (QM/MM) Calculations	85
	<i>Lung Wa Chung, Xin Li, and Keiji Morokuma</i>	
3.1	Introduction	85
3.2	Computational Strategies (Methods and Models)	86
3.2.1	Quantum Mechanical (QM) Methods	86
3.2.2	Active-Site Model	88
3.2.3	QM/MM Methods	88
3.2.4	QM/MM Model and Setup	90
3.3	Metalloenzymes	91
3.3.1	Heme-Containing Enzymes	91
3.3.1.1	Binding and Photodissociation of Diatomic Molecules	91
3.3.1.2	Heme Oxygenase (HO)	95
3.3.1.3	Indoleamines Dioxygenase (IDO) and Tryptophan Dioxygenase (TDO)	97
3.3.1.4	Nitric Oxide Synthase (NOS)	101
3.3.2	Cobalamin-Dependent Enzymes	105
3.3.2.1	Methylmalonyl-CoA Mutase	105
3.3.2.2	Glutamine Mutase	108
3.4	Photobiology	109
3.4.1	Fluorescent Proteins (FPs)	109
3.4.1.1	Green Fluorescent Proteins (GFP)	110
3.4.1.2	Reversible Photoswitching Fluorescent Proteins (RPFs)	111
3.4.1.3	Photoconversion of Fluorescent Proteins	115
3.4.2	Luciferases	117
3.5	Conclusion	120
	References	120
4	From Molecular Electrostatic Potentials to Solvation Models and Ending with Biomolecular Photophysical Processes	131
	<i>Jacopo Tomasi, Chiara Cappelli, Benedetta Mennucci, and Roberto Cammi</i>	131
4.1	Introduction	131
4.2	The Molecular Electrostatic Potential and Noncovalent Interactions among Molecules	132
4.2.1	Molecular Electrostatic Potential	132
4.2.1.1	Use of MEP	133
4.2.1.2	Semiclassical Approximation	133
4.2.1.3	MEP as a Component of the Intermolecular Interaction	134
4.2.1.4	Definition of the Coulomb Interaction Term	135
4.2.1.5	Simplifications in the Expression of E_{es} : Point Charge Descriptions	135
4.2.1.6	Simplifications in the Expression of E_{es} : Atomic Charges	136
4.2.1.7	Simplifications in the Expression of E_{es} : Multipolar Expansions	136

4.2.2	Interaction Energy between Two Molecules	137
4.2.3	Examples of Energy Decomposition Analyses	139
4.2.3.1	Interactions with a Proton	139
4.2.3.2	Interactions with Other Cations	139
4.2.3.3	Hydrogen Bonding	140
4.2.4	Interaction Potentials (Force Fields) for Computer Simulations of Liquid Systems	140
4.3	Solvation: the “Continuum Model”	142
4.3.1	Basic Formulation of PCM	142
4.3.2	Beyond the Basic Formulation	146
4.3.2.1	Dielectric Function	146
4.3.2.2	Cavity Surface	147
4.3.2.3	Definition of the Apparent Charges	147
4.3.2.4	Description of the Solute	147
4.3.3	Other Continuum Solvation Methods	148
4.3.3.1	Apparent Surface Charge (ASC) Methods	148
4.3.3.2	Multipole Expansion Methods (MPE)	149
4.3.3.3	Generalized Born Model	149
4.3.3.4	Finite Element Method (FEM) and Finite Difference Method (FDM)	150
4.4	Applications of the PCM Method	150
4.4.1	Solvation Energies	150
4.4.2	About the PES	152
4.4.3	Chemical Equilibria	152
4.4.3.1	Tautomeric Equilibria	153
4.4.3.2	Equilibria in Molecular Aggregation	153
4.4.3.3	pK_a of Acids	153
4.4.4	Reaction Mechanisms	154
4.4.5	Solvent Effects on Molecular Properties and Spectroscopy	156
4.4.5.1	<i>N</i> -Acetylproline Amide (NAP)	157
4.4.5.2	Glucose	158
4.4.5.3	Local Field Effects	159
4.4.5.4	Dynamic Effects	160
4.4.6	Effect of the Environment on Formation and Relaxation of Excited States	161
4.4.7	Electronic Transitions and Related Spectroscopies	162
4.4.8	Photoinduced Electron and Energy Transfers	164
	References	166
5	The Fast Marching Method for Determining Chemical Reaction Mechanisms in Complex Systems	171
	<i>Yuli Liu, Steven K. Burger, Bijoy K. Dey, Utpal Sarkar, Marek R. Janicki, and Paul W. Ayers</i>	
5.1	Motivation	171

5.2	Background	172
5.2.1	Minimum Energy Path	172
5.2.2	Two End Methods	172
5.2.3	Surface Walking Algorithms	173
5.2.4	Metadynamics Methods	174
5.2.5	Fast Marching Method	174
5.3	Fast Marching Method	175
5.3.1	Introduction to FMM	175
5.3.2	Upwind Difference Approximation	176
5.3.3	Heapsort Technique	176
5.3.4	Shepard Interpolation	177
5.3.5	Interpolating Moving Least-Squares Method	179
5.3.6	FMM Program	180
5.3.6.1	Setup, Definitions and Notation	180
5.3.6.2	Initialize the Calculation	181
5.3.6.3	Updating the Heap	181
5.3.6.4	Backtracing from the Ending Point to the Starting Point on the Energy Cost Surface	181
5.3.7	Application	182
5.3.7.1	Four-Well Analytical PES	182
5.3.7.2	S_N2 Reaction	184
5.3.7.3	Dissociation of Ionized O-Methylhydroxylamine	185
5.4	Quantum Mechanics/Molecular Mechanics (QM/MM) Methods Applied to Enzyme-Catalyzed Reactions	187
5.4.1	QM/MM Methods	187
5.4.2	Incorporating the QM/MM-MFEP Methods with FMM	189
5.4.3	Application of the Incorporated FMM and QM/MM-MFEP Method to Enzyme-Catalyzed Reactions	190
5.4.3.1	S_N2 Reaction in Solvent	190
5.4.3.2	Isomerization Reaction Catalyzed by 4-Oxalocrotonate Tautomerase (4-OT)	190
5.4.3.3	Dechlorination Reaction Catalyzed by <i>trans</i> -3-Chloroacrylic Acid Dehalogenase (CAAD)	191
5.5	Summary	191
	References	192

Part Two Nucleic Acids, Amino Acids, Peptides and Their Interactions 197

6	Chemical Origin of Life: How do Five HCN Molecules Combine to form Adenine under Prebiotic and Interstellar Conditions	199
	<i>Debjani Roy and Paul von Ragué Schleyer</i>	
6.1	Introduction	199
6.1.1	Prebiotic Chemistry: Experimental Endeavor to Synthesize the Building Blocks of Biopolymers	199

6.1.2	Key Role of HCN as a Precursor for Prebiotic Compounds	201
6.1.3	Prebiotic Experiments and Proposed Pathways for the Formation of Adenine	202
6.2	Computational Investigation	202
6.2.1	Method	204
6.2.2	Thermochemistry of Pentamerization	204
6.2.3	Detailed Step by Step Mechanism	205
6.2.3.1	DAMN vs AICN as Adenine Precursors	205
6.2.3.2	Is an Anionic Mechanism Feasible in Isolation?	205
6.2.3.3	Two Tautomeric forms of AICN: Which one is the Favorable Precursor for Adenine Formation under Prebiotic Conditions?	207
6.2.3.4	Validating the Methods Used for Computing Barrier Heights	213
6.3	Conclusion	213
	References	216
7	Hydrogen Bonding and Proton Transfer in ionized DNA Base Pairs, Amino Acids and Peptides	219
	<i>Luis Rodríguez-Santiago, Marc Noguera, Joan Bertran, and Mariona Sodupe</i>	
7.1	Introduction	219
7.2	Methodological Aspects	220
7.3	Ionization of DNA Base Pairs	221
7.3.1	Equilibrium Geometries and Dimerization Energies	222
7.3.2	Single and Double Proton Transfer Reactions	223
7.4	Ionization of Amino Acids	227
7.4.1	Structural Features of Neutral and Radical Cation Amino Acids	227
7.4.2	Intramolecular Proton-Transfer Processes	231
7.5	Ionization of Peptides	234
7.5.1	Ionization of <i>N</i> -Glycylglycine	234
7.5.2	Influence of Ionization on the Ramachandran Maps of Model Peptides	236
7.6	Conclusions	239
	References	241
8	To Nano-Biochemistry: Picture of the Interactions of DNA with Gold	245
	<i>Eugene S. Kryachko</i>	
8.1	Introductory Nanoscience Background	245
8.1.1	Gold in Nanodimensions	246
8.1.2	Gold and DNA: Meeting Points in Nanodimensions	248
8.2	DNA–Gold Bonding Patterns: Some Experimental Facts	253
8.3	Adenine–Gold Interaction	254
8.3.1	Adenine–Au and Adenine–Au ₃ Bonding Patterns	254
8.3.2	Propensity of Gold to Act as Nonconventional Proton Acceptor	257
8.3.2.1	Pause: A Short Excursion to Hydrogen Bonding Theory	259

8.3.2.2	Proof that $N-H \cup Au \equiv N-H \cdots Au$ in $A \cdot Au_3(N_{i=1,3,7})$	260
8.3.2.3	Nonconventional Hydrogen Bonds $N-H \cdots Au$ in $A \cdot Au_3(N_{i=1,3,7})$	261
8.3.3	Complex $A \cdot Au_3(N_6)$	262
8.3.4	Interaction between Adenine and Chain Au_3 Cluster	262
8.4	Guanine–Gold Interaction	263
8.5	Thymine–Gold Interactions	268
8.6	Cytosine–Gold Interactions	272
8.7	Basic Trends of DNA Base–Gold Interaction	273
8.7.1	Anchoring Bond in DNA Base–Gold Complexes	276
8.7.2	Energetics in $Z = 0$ Charge State	278
8.7.3	$Z = -1$ Charge State	282
8.8	Interaction of Watson–Crick DNA Base Pairs with Gold Clusters	286
8.8.1	General Background	286
8.8.2	$[A \cdot T] \cdot Au_3$ Complexes	289
8.8.3	$[G \cdot C] \cdot Au_3$ Complexes	293
8.8.4	Au_6 Cluster Bridges the WC G·C Pair	296
8.9	Summary and Perspectives	297
	References	298
9	Quantum Mechanical Studies of Noncovalent DNA–Protein Interactions	307
	<i>Lesley R. Rutledge and Stacey D. Wetmore</i>	
9.1	Introduction	307
9.2	Computational Approaches for Studying Noncovalent Interactions	308
9.3	Hydrogen-Bonding Interactions	315
9.3.1	Interactions between the Protein Backbone and DNA Nucleobases	315
9.3.2	Interactions between Protein Side Chains and DNA Backbone	316
9.3.3	Interactions between Protein Side Chains and DNA Nucleobases	317
9.4	Interactions between Aromatic DNA–Protein Components	318
9.4.1	Stacking Interactions	319
9.4.2	T-Shaped Interactions	323
9.5	Cation– π Interactions between DNA–Protein Components	326
9.5.1	Cation– π Interactions between Charged Nucleobases and Aromatic Amino Acids	326
9.5.2	Cation– π Interactions Involving Charged Aromatic Amino Acids	330
9.5.3	Cation– π Interactions Involving Charged Non-aromatic Amino Acids	330
9.5.4	Simultaneous Cation– π and Hydrogen-Bonding Interactions (DNA–Protein Stair Motifs)	332
9.6	Conclusions	333
	References	333

10	The Virial Field and Transferability in DNA Base-Pairing	337
	<i>Richard F.W. Bader and Fernando Cortés-Guzmán</i>	
10.1	A New Theorem Relating the Density of an Atom in a Molecule to the Energy	337
10.2	Computations	339
10.3	Chemical Transferability and the One-Electron Density Matrix	339
10.3.1	The Virial Field	340
10.3.2	Short-Range Nature of the Virial Field and Transferability	342
10.4	Changes in Atomic Energies Encountered in DNA Base Pairing	343
10.4.1	Dimerization of the Four Bases A, C, G and T	346
10.4.2	Energy Changes in CC	349
10.4.3	Energy Changes in AA1	349
10.4.4	Energy Changes in GG4	350
10.4.5	Energy Changes in TT2	350
10.5	Energy Changes in the WC Pairs GC and AT	350
10.6	Discussion	355
10.6.1	Attractive and Repulsive Contributions to the Atomic Virial and its Short-Range Nature	356
10.6.2	Can One Go Directly to the Virial Field?	360
	References	363
11	An Electron Density-Based Approach to the Origin of Stacking Interactions	365
	<i>Ricardo A. Mosquera, María J. González Moa, Laura Estévez, Marcos Mandado, and Ana M. Graña</i>	
11.1	Introduction	365
11.2	Computational Method	366
11.3	Charge-Transfer Complexes: Quinhydrone	367
11.4	π - π Interactions in Hetero-Molecular Complexes: Methyl Gallate-Caffeine Adduct	371
11.5	π - π Interactions between DNA Base Pair Steps	374
11.6	π - π Interactions in Homo-Molecular Complexes: Catechol	378
11.7	C-H/ π Complexes	381
11.8	Provisional Conclusions and Future Research	385
	References	385
12	Polarizabilities of Amino Acids: Additive Models and <i>Ab Initio</i> Calculations	389
	<i>Noureddin El-Bakali Kassimi and Ajit J. Thakkar</i>	
12.1	Introduction	389
12.2	Models of Polarizability	389
12.3	Polarizabilities of the Amino Acids	393

12.4	Concluding Remarks	398
	References	400
13	Methods in Biocomputational Chemistry: A Lesson from the Amino Acids	403
	<i>Hugo J. Bohórquez, Constanza Cárdenas, Chérif F. Matta, Russell J. Boyd, and Manuel E. Patarroyo</i>	
13.1	Introduction	403
13.2	Conformers, Rotamers and Physicochemical Variables	404
13.3	QTAIM Side Chain Polarizations and the Theoretical Classification of Amino Acids	408
13.4	Quantum Mechanical Studies of Peptide–Host Interactions	414
13.5	Conclusions	419
	References	420
14	From Atoms in Amino Acids to the Genetic Code and Protein Stability, and Backwards	423
	<i>Chérif F. Matta</i>	
14.1	Context of the Work	423
14.2	The Electron Density $\rho(\mathbf{r})$ as an Indirectly Measurable Dirac Observable	426
14.3	Brief Review of Some Basic Concepts of the Quantum Theory of Atoms in Molecules	430
14.4	Computational Approach and Level of Theory	438
14.5	Empirical Correlations of QTAIM Atomic Properties of Amino Acid Side Chains with Experiment	439
14.5.1	Partial Molar Volumes	439
14.5.2	Free Energy of Transfer from the Gas to the Aqueous Phase	448
14.5.3	Simulation of Genetic Mutations with Amino Acids Partition Coefficients	448
14.5.4	Effect of Genetic Mutation on Protein Stability	451
14.5.5	From the Genetic Code to the Density and Back	454
14.6	Molecular Complementarity	456
14.7	Closing Remarks	462
14.8	Appendix A X-Ray and Neutron Diffraction Geometries of the Amino Acids in the Literature	462
	References	467
15	Energy Richness of ATP in Terms of Atomic Energies: A First Step	473
	<i>Chérif F. Matta and Alya A. Arabi</i>	
15.1	Introduction	473
15.2	How “(De)Localized” is the Enthalpy of Bond Dissociation?	474
15.3	The Choice of a Theoretical Level	477
15.3.1	The Problem	477

15.3.2	Empirical Correlation of Trends in the Atomic Contributions to BDE: Comparison of MP2 and DFT(B3LYP) Results	478
15.3.3	Theory	478
15.3.3.1	QTAIM Atomic Energies from the <i>ab initio</i> Methods	478
15.3.3.2	Atomic Energies from Kohn–Sham Density Functional Theory Methods	482
15.3.3.3	Atomic Contributions to the Energy of Reaction	484
15.4	Computational Details	484
15.5	(Global) Energies of the Hydrolysis of ATP in the Absence and Presence of Mg^{2+}	485
15.6	How “(De)Localized” is the Energy of Hydrolysis of ATP?	485
15.6.1	Phosphate Group Energies and Modified Lipmann’s Group Transfer Potentials	485
15.6.2	Atomic Contributions to the Energy of Hydrolysis of ATP in the Absence and Presence of Mg^{2+}	487
15.7	Other Changes upon Hydrolysis of ATP in the Presence and Absence of Mg^{2+}	487
15.7.1	Bond Properties and Molecular Graphs	487
15.7.2	Group Charges in ATP in the Absence and Presence of Mg^{2+}	491
15.7.3	Molecular Electrostatic Potential in the Absence and Presence of Mg^{2+}	492
15.8	Conclusions	493
	References	496

Vol II

Part Three Reactivity, Enzyme Catalysis, Biochemical Reaction Paths and Mechanisms 499

16	Quantum Transition State for Peptide Bond Formation in the Ribosome	501
	<i>Lou Massa, Chérif F. Matta, Ada Yonath, and Jerome Karle</i>	
16.1	Introduction	501
16.2	Methodology: Searching for the Transition State and Calculating its Properties	502
16.3	Results: The Quantum Mechanical Transition State	506
16.4	Discussion	511
16.5	Summary and Conclusions	513
	References	514
17	Hybrid QM/MM Simulations of Enzyme-Catalyzed DNA Repair Reactions	517
	<i>Denis Bucher, Fanny Masson, J. Samuel Arey, and Ursula Röthlisberger</i>	
17.1	Introduction	517
17.2	Theoretical Background	518
17.3	Applications	521

17.3.1	Thymine Dimer Splitting Catalyzed by DNA Photolyase	521
17.3.2	Reaction Mechanism of Endonuclease IV	525
17.3.3	Role of Water in the Catalysis Mechanism of DNA Repair Enzyme, MutY	529
17.4	Conclusions	533
	References	534
18	Computational Electronic Structure of Spin-Coupled Diiron-Oxo Proteins	537
	<i>Jorge H. Rodriguez</i>	
18.1	Introduction	537
18.2	(Anti)ferromagnetic Spin Coupling	538
18.3	Spin Density Functional Theory of Antiferromagnetic Diiron Complexes	539
18.4	Phenomenological Simulation of Mössbauer Spectra of Diiron-oxo Proteins	542
18.4.1	Antiferromagnetic Diiron Center of Hemerythrin	542
18.4.2	Nitric Oxide Derivative of Hr	543
18.4.3	Antiferromagnetic Diiron Center of Reduced Uteroferrin	545
18.5	Conclusion	546
	References	548
19	Accurate Description of Spin States and its Implications for Catalysis	551
	<i>Marcel Swart, Mireia Güell, and Miquel Solà</i>	
19.1	Introduction	551
19.2	Influence of the Basis Set	553
19.3	Spin-Contamination Corrections	556
19.4	Influence of Self-Consistency	558
19.5	Spin-States of Model Complexes	559
19.6	Spin-States Involved in Catalytic Cycles	564
19.6.1	Cytochrome P450cam	564
19.6.2	His-Porphyrin Models	567
19.6.2.1	Reference Data (Harvey)	568
19.6.2.2	Reference Data (Ghosh)	570
19.6.2.3	Other Model Systems	571
19.6.3	NiFe Hydrogenase	574
19.7	Concluding Remarks	579
19.8	Computational Details	579
	References	580
20	Quantum Mechanical Approaches to Selenium Biochemistry	585
	<i>Jason K. Pearson and Russell J. Boyd</i>	
20.1	Introduction	585
20.2	Quantum Mechanical Methods for the Treatment of Selenium	586

20.3	Applications to Selenium Biochemistry	587
20.3.1	Computational Studies of GPx	587
20.3.2	Computational Studies on GPx Mimics	589
20.3.2.1	GPx-like Activity of Ebselen	589
20.3.2.2	Substituent Effects on the GPx-like Activity of Ebselen	596
20.3.2.3	Effect of the Molecular Environment on GPx-like Activity	598
20.4	Summary	600
	References	600
21	Catalytic Mechanism of Metallo β-Lactamases: Insights from Calculations and Experiments	605
	<i>Matteo Dal Peraro, Alejandro J. Vila, and Paolo Carloni</i>	
21.1	Introduction	605
21.2	Structural Information	607
21.3	Computational Details	608
21.4	Preliminary Comment on the Comparison between Theory and Experiment	609
21.5	Michaelis Complex in B1 M β Ls	610
21.5.1	Substrate Binding Determinants	610
21.5.2	Nucleophile Structural Determinants	611
21.6	Catalytic Mechanism of B1 M β Ls	612
21.6.1	Cefotaxime Enzymatic Hydrolysis in CcrA	613
21.6.2	Cefotaxime Enzymatic Hydrolysis in BcII	614
21.6.3	Zinc Content and Reactivity of B1 M β Ls	615
21.6.4	Reactivity of β -Lactam Antibiotics other than Cefotaxime	615
21.7	Michaelis Complexes of other M β Ls	616
21.7.1	B2 Mono-Zn M β L Subclass	616
21.7.2	B3 M β L Subclass	616
21.8	Concluding Remarks	617
	References	618
22	Computational Simulation of the Terminal Biogenesis of Sesquiterpenes: The Case of 8-Epiconfertifin	623
	<i>José Enrique Barquera-Lozada and Gabriel Cuevas</i>	
22.1	Introduction	623
22.2	Reaction Mechanism	627
22.3	Conclusions	639
	References	640
23	Mechanistics of Enzyme Catalysis: From Small to Large Active-Site Models	643
	<i>Jorge Llano and James W. Gauld</i>	
23.1	Introduction	
23.1.1	Factors Influencing the Catalytic Performance of Enzymes	643

23.1.2	Computational Modeling in Enzymology	648
23.2	Active-Site Models of Enzymatic Catalysis: Methods and Accuracy	650
23.3	Redox Catalytic Mechanisms	652
23.3.1	NO Formation in Nitric Oxide Synthase	652
23.3.2	Oxidative Dealkylation in the AlkB Family	654
23.4	General Acid–Base Catalytic Mechanism of Deacetylation in LpxC	658
23.5	Summary	660
	References	662
Part Four	From Quantum Biochemistry to Quantum Pharmacology, Therapeutics, and Drug Design	667
24	Developing Quantum Topological Molecular Similarity (QTMS)	669
	<i>Paul L.A. Popelier</i>	
24.1	Introduction	669
24.2	Anchoring in Physical Organic Chemistry	671
24.3	Equilibrium Bond Lengths: “Threat” or “Opportunity”?	678
24.4	Introducing Chemometrics: Going Beyond r^2	679
24.5	A Hopping Center of Action	681
24.6	A Leap	684
24.7	A Couple of General Reflections	687
24.8	Conclusions	688
	References	689
25	Quantum-Chemical Descriptors in QSAR/QSPR Modeling: Achievements, Perspectives and Trends	693
	<i>Anna V. Gubskaya</i>	
25.1	Introduction	693
25.2	Quantum-Chemical Methods and Descriptors	694
25.2.1	Quantum-Chemical Methods	694
25.2.2	Quantum-Chemical Descriptors: Classification, Updates	697
25.3	Computational Approaches for Establishing Quantitative Structure–Activity Relationships	703
25.3.1	Selection of Descriptors	703
25.3.2	Linear Regression Techniques	705
25.3.3	Machine-Learning Algorithms	706
25.4	Quantum-Chemical Descriptors in QSAR/QSPR Models	710
25.4.1	Biochemistry and Molecular Biology	710
25.4.2	Medicinal Chemistry and Drug Design	712
25.4.3	Material and Biomaterial Science	714
25.5	Summary and Conclusions	715
	References	717

26	Platinum Complexes as Anti-Cancer Drugs: Modeling of Structure, Activation and Function 723
	<i>Konstantinos Gkionis, Mark Hicks, Arturo Robertazzi, J. Grant Hill, and James A. Platts</i>
26.1	Introduction to Cisplatin Chemistry and Biochemistry 723
26.2	Calculation of Cisplatin Structure, Activation and DNA Interactions 726
26.3	Platinum-Based Alternatives 732
26.4	Non-platinum Alternatives 735
26.5	Absorption, Distribution, Metabolism, Excretion (ADME) Aspects 739
	References 740
27	Protein Misfolding: The Quantum Biochemical Search for a Solution to Alzheimer's Disease 743
	<i>Donald F. Weaver</i>
27.1	Introduction 743
27.2	Protein Folding and Misfolding 744
27.2.1	Protein Folding 744
27.2.2	Protein Misfolding 745
27.3	Quantum Biochemistry in the Study of Protein Misfolding 745
27.3.1	Molecular Mechanics 746
27.4	Alzheimer's Disease: A Disorder of Protein Misfolding 747
27.4.1	Alzheimer's – A Protein Misfolding Disorder 748
27.4.2	Protein Misfolding of Beta-Amyloid 748
27.5	Quantum Biochemistry and Designing Drugs for Alzheimer's Disease 750
27.5.1	Approach 1 – Homotaurine 751
27.5.2	Approach 2 – Melatonin 752
27.6	Conclusions 753
	References 754
28	Targeting Butyrylcholinesterase for Alzheimer's Disease Therapy 757
	<i>Katherine V. Darvesh, Ian R. Pottie, Robert S. McDonald, Earl Martin, and Sultan Darvesh</i>
28.1	Butyrylcholinesterase and the Regulation of Cholinergic Neurotransmission 757
28.2	Butyrylcholinesterase: The Significant other Cholinesterase, in Sickness and in Health 760
28.3	Optimizing Specific Inhibitors of Butyrylcholinesterase Based on the Phenothiazine Scaffold 761
28.4	Biological Evaluation of Phenothiazine Derivatives as Cholinesterase Inhibitors 761
28.5	Computation of Physical Parameters to Interpret Structure–Activity Relationships 769

28.6	Enzyme–Inhibitor Structure–Activity Relationships	772
28.7	Conclusions	777
	References	778
29	Reduction Potentials of Peptide-Bound Copper (II) – Relevance for Alzheimer’s Disease and Prion Diseases	781
	<i>Arvi Rauk</i>	
29.1	Introduction	781
29.2	Copper Binding in Albumin – Type 2	783
29.3	Copper Binding to Ceruloplasmin – Type 1	785
29.4	The Prion Protein Octarepeat Region	787
29.5	Copper and the Amyloid Beta Peptide (A β) of Alzheimer’s Disease	789
29.6	Cu(II)/Cu(I) Reduction Potentials in Cu/A β	791
29.7	Concluding Remarks	794
29.A	Appendix	795
29.A.1	Calculation of Reduction Potentials, E°, of Copper/Peptide Complexes	795
29.A.2	Computational Methodology	796
	References	798
30	Theoretical Investigation of NSAID Photodegradation Mechanisms	805
	<i>Klefah A.K. Musa and Leif A. Eriksson</i>	
30.1	Drug Safety	805
30.2	Drug Photosensitivity	806
30.2.1	Photoallergies	807
30.2.2	Photophobia	807
30.2.3	Phototoxicity	807
30.3	Non-Steroid Anti-Inflammatory Drugs (NSAIDs)	808
30.3.1	NSAID: Definition and Classification	808
30.3.2	Pharmacological Action	808
30.3.3	NSAID Uses	809
30.3.4	Side Effects	810
30.4	NSAID Phototoxicity	811
30.5	Theoretical Studies	812
30.5.1	Overview	812
30.5.2	Methodology	814
30.6	Redox Chemistry	815
30.7	NSAID Orbital Structures	817
30.8	NSAID Absorption Spectra	820
30.9	Excited State Reactions	823
30.9.1	Photodegradation from the T ₁ State	825
30.9.2	Possible Photodegradation from Singlet Excited States	826
30.10	Reactive Oxygen Species (ROS) and Radical Formation	827

30.11	Effects of the Formed ROS and Radicals during the Photodegradation Mechanisms	828
30.12	Conclusions	830
	References	831

Part Five Biochemical Signature of Quantum Indeterminism 835

31 Quantum Indeterminism, Mutation, Natural Selection, and the Meaning of Life 837

David N. Stamos

31.1	Introduction	837
31.2	A Short History of the Debate in Philosophy of Biology	839
31.3	Replies to My Paper	842
31.4	The Quantum Indeterministic Basis of Mutations	845
31.4.1	Tautomeric Shifts	845
31.4.2	Proton Tunneling	849
31.4.3	Aqueous Thermal Motion	852
31.5	Mutation and the Direction of Evolution	853
31.6	Mutational Order	855
31.7	The Nature of Natural Selection	857
31.8	The Meaning of Life	863
	References	867

32 Molecular Orbitals: Dispositions or Predictive Structures? 873

Jean-Pierre Llored and Michel Bitbol

32.1	Origins of Quantum Models in Chemistry: The Composite and the Aggregate	874
32.2	Evolution of the Quantum Approaches and Biology	876
32.3	Philosophical Implications of Molecular Quantum Holism: Dispositions and Predictive Structures	882
32.3.1	Molecular Landscapes and Process	882
32.3.2	Realism of Disposition and Predictive Structures	886
32.4	Closing Remarks	893
	References	893

Index 897

List of Contributors

Alya A. Arabi

Dalhousie University
Department of Chemistry
Halifax, Nova Scotia B3H 4J3
Canada
Alya.Arabid@dal.ca

J. Samuel Arey

Federal Institute of Technology – EPFL
Environmental Chemistry Modeling
Laboratory
CH-1015 Lausanne
Switzerland
samuel.arey@epfl.ch

Paul W. Ayers

McMaster University
Department of Chemistry
1280 Main St. West
Hamilton, Ontario L8S 4M1
Canada
ayers@mcmaster.ca

Richard F.W. Bader

McMaster University
Department of Chemistry
Hamilton, Ontario L7L 2T1
Canada
bader@mcmaster.ca

José Enrique Barquera-Lozada

Universidad Nacional Autónoma de
México
Instituto de Química
Coyoacán, Circuito Exterior, Apdo.
Postal 70213
D.F. 04510
México
joseenquebl@yahoo.com

Joan Bertran

Universitat Autònoma de Barcelona
Departament de Química
Bellaterra 08193
Spain
bertran@klignon.uab.es

Michel Bitbol

Université Paris 1
Centre de Recherches en Epistémologie
Appliquée (CREA/Ecole Polytechnique)
32, boulevard Victor
75015 Paris
France
michel.bitbol@polytechnique.edu

Hugo J. Bohórquez

Dalhousie University
Department of Chemistry
Halifax, Nova Scotia B3H 4J3
Canada
Hugo.J.Bohorquez@DAL.CA

Russell J. Boyd

Dalhousie University
Department of Chemistry
Halifax, Nova Scotia B3H 4J3
Canada
russell.boyd@dal.ca

Denis Bucher

University of Sydney
School of Physics
Sydney, NSW 2006
Australia

and

Federal Institute of Technology – EPFL
Laboratory of Computational Chemistry
and Biochemistry
CH-1015 Lausanne
Switzerland
dbucher@physics.usyd.edu.au

Steven K. Burger

McMaster University
Department of Chemistry
1280 Main St. West
Hamilton, Ontario L8S 4M1
Canada
burgers@mcmaster.ca

Roberto Cammi

Università di Parma
Dipartimento di Chimica
Viale delle Scienze 17/A
43100 Parma
Italy
roberto.cammi@unipr.it

Chiara Cappelli

Università di Pisa
Dipartimento di Chimica e Chimica
Industriale
via Risorgimento 35
I-56126 Pisa
Italy
chiara@dcci.unipi.it

Constanza Cárdenas

Pontificia
Universidad Católica de Valparaíso
Laboratorio de Genética e Inmunología
Molecular
Av Brasil 2950
Valparaíso
Chile
conscc@hotmail.com

Paolo Carloni

International School for Advanced
Studies
SISSA-ISAS via Beirut 2-4
34014 Trieste
Italy
carloni@sissa.it

Lung Wa Chung

Kyoto University
Fukui Institute for Fundamental
Chemistry
Kyoto 606-8103
Japan
chung@fukui.kyoto-u.ac.jp

Fernando Clemente

Gaussian, Inc.
340 Quinipiac Street, Building 40
Wallingford, CT 06492
USA
fernando@gaussian.com

Fernando Cortés-Guzmán

Universidad Nacional Autónoma de
Mexico
Instituto de Química
Departamento de Fisicoquímica
Ciudad Universitaria, Coyoacán
D.F. 04510
Mexico
fercor@servidor.unam.mx

Gabriel Cuevas

Universidad Nacional Autónoma de México
 Instituto de Química
 Coyoacán, Circuito Exterior, Apdo.
 Postal 70213
 D.F. 04510
 México
 gecgb@servidor.unam.mx

Matteo Dal Peraro

Federale Institute of Technology-EPFL
 Institute of Bioengineering
 Laboratory for Biomolecular Modeling
 CH-1015 Lausanne
 Switzerland
 matteo.dalperaro@epfl.ch

Katherine V. Darvesh

Mount Saint Vincent University
 Department of Chemistry and Physics
 Halifax, Nova Scotia B3M 2J6
 Canada
 Katherine.Darvesh@msvu.ca

Sultan Darvesh

Dalhousie University
 Departments of Medicine (Neurology),
 Anatomy & Neurobiology and
 Chemistry
 Halifax, Nova Scotia B3H 4J3
 Canada

and

Mount Saint Vincent University
 Department of Chemistry and Physics
 Halifax, Nova Scotia B3M 2J6
 Canada
 sultan.darvesh@dal.ca

Bijoy K. Dey

McMaster University
 Department of Chemistry
 1280 Main St. West
 Hamilton, Ontario L8S 4M1
 Canada
 bijoy@scienide.uwaterloo.ca

Leif A. Eriksson

National University of Ireland
 (NUI Galway)
 School of Chemistry
 University Road
 Galway
 Ireland
 leif.eriksson@oru.se

Laura Estévez

Universidade de Vigo
 Departamento de Química Física
 Lagoas-Marcosende s/n
 36310-Vigo, Galicia
 Spain
 laurilla@uvigo.es

Michael J. Frisch

Gaussian, Inc.
 340 Quinpiac Street, Building 40
 Wallingford, CT 06492
 USA
 frisch@gaussian.com

James W. Gauld

University of Windsor
 Department of Chemistry and
 Biochemistry
 Windsor, Ontario N9B 3P4
 Canada
 gauld@uwindsor.ca

Konstantinos Gkionis

Cardiff University
 School of Chemistry
 Cardiff CF10 3AT
 UK

María J. González Moa

Universidade de Vigo
Departamento de Química Física
Lagoas-Marcosende s/n
36310-Vigo, Galicia
Spain
mjmoa@uvigo.es

Ana M. Graña

Universidade de Vigo
Departamento de Química Física
Lagoas-Marcosende s/n
36310-Vigo, Galicia
Spain
ana@uvigo.es

Anna V. Gubskaya

Rutgers University
Department of Chemistry and Chemical
Biology
Piscataway, NJ
USA
Ganna.Gubska@msvu.ca

Mireia Güell

Universitat de Girona
Institut de Química Computacional and
Departament de Química
Campus Montilivi
17071 Girona
Spain
mireia.guell@udg.edu

Mark Hicks

Cardiff University
School of Chemistry
Cardiff CF10 3AT
UK

J. Grant Hill

Cardiff University
School of Chemistry
Cardiff CF10 3AT
UK

Lulu Huang

Naval Research Laboratory
Laboratory for the Structure of Matter
Washington, DC 20375-5341
USA
huang@ccs.nrl.navy.mil

Marek R. Janicki

McMaster University
Department of Chemistry
1280 Main St. West
Hamilton, Ontario L8S 4M1
Canada
marek.janicki@utoronto.ca

Jerome Karle

Naval Research Laboratory
Laboratory for the Structure of Matter
Washington, DC 20375-5341
USA
peggy@ccs.nrl.navy.mil
jerome.karle@nrl.navy.mil

Noureddin El-Bakali Kassimi

University of New Brunswick
Department of Chemistry
Fredericton, New Brunswick E3B 5A3
Canada
nkassimi@unb.ca

Eugene S. Kryachko

Bogolyubov Institute for Theoretical
Physics
Kiev-143, 03680
Ukraine
e_kryachko@yahoo.com
eugene.kryachko@ulg.ac.be

Xin Li

Kyoto University
Fukui Institute for Fundamental
Chemistry
Kyoto 606-8103
Japan
lixin@fukui.kyoto-u.ac.jp

Yuli Liu

McMaster University
 Department of Chemistry
 1280 Main St. West
 Hamilton, Ontario L8S 4M1
 Canada
 liuy34@mcmastsr.ca

Jorge Llano

University of Windsor
 Department of Chemistry and
 Biochemistry
 Windsor, Ontario N9B 3P4
 Canada
 llano@uwindsor.ca

Jean-Pierre Llored

Centre de Recherches en Epistémologie
 Appliquée (CREA/Ecole Polytechnique)
 32, boulevard Victor
 75015 Paris
 France
 jean-pierre.llored@polytechnique.edu

Marcos Mandado

Universidade de Vigo
 Departamento de Química Física
 Lagoas-Marcosende s/n
 36310-Vigo, Galicia
 Spain
 mandado@uvigo.es

Earl Martin

Mount Saint Vincent University
 Department of Chemistry and Physics
 Halifax, Nova Scotia B3M 2J6
 Canada
 Earl.Martin@msvu.ca

Lou Massa

City University of New York
 Hunter College and the Graduate School
 New York, NY 10065
 USA
 Lmassa@Hunter.CUNY.edu
 lmassa@verizon.net

Fanny Masson

Universität Zürich
 Physikalisch Chemisches Institut
 Winterthurerstrasse 190
 CH-8057 Zürich
 Switzerland
 masson@pci.unizh.ch

Chérif F. Matta

Mount Saint Vincent University
 Department of Chemistry and Physics
 Halifax, Nova Scotia B3M 2J6
 Canada
 cherif.matta@dal.ca
 cherif.matta@msvu.ca

Robert S. McDonald

Mount Saint Vincent University
 Department of Chemistry and Physics
 Halifax, Nova Scotia B3M 2J6
 Canada

Benedetta Mennucci

Università di Pisa
 Dipartimento di Chimica e Chimica
 Industriale
 via Risorgimento 35
 I-56126 Pisa
 Italy
 bene@dcc.unipi.it

Keiji Morokuma

Kyoto University
Fukui Institute for Fundamental
Chemistry
Kyoto 606-8103
Japan
morokuma@fukui.kyoto-u.ac.jp

Ricardo A. Mosquera

Universidade de Vigo
Departamento de Química Física
Lagoas-Marcosende s/n
36310-Vigo, Galicia
Spain
mosquera@uvigo.es

Klefah A.K. Musa

Örebro University
Örebro Life Science Center School of
Science and Technology
701 82 Örebro
Sweden
klefah.musa@oru.se

Marc Noguera

Universitat Autònoma de Barcelona
Departament de Química
Bellaterra 08193
Spain
marc@klingon.uab.es

Manuel E. Patarroyo

Fundación Instituto de Inmunología de
Colombia (FIDIC)
Bogotá D.C.
Colombia
mpatarr@mail.com

Jason K. Pearson

Dalhousie University
Department of Chemistry
Halifax, Nova Scotia B3H 4J3
Canada
kpearson@dal.ca

James A. Platts

Cardiff University
School of Chemistry
Cardiff CF10 3AT
UK
Platts@Cardiff.ac.uk

Paul L.A. Popelier

University of Manchester
School of Chemistry
Oxford Road
Manchester M13 9PL
UK

and

Manchester Interdisciplinary Biocentre
(MIB)
131 Princess Street
Manchester M1 7DN
UK
Paul.Popelier@manchester.ac.uk

Ian R. Pottie

Mount Saint Vincent University
Department of Chemistry and Physics
Halifax, Nova Scotia B3M 2J6
Canada
ian.pottie@msvu.ca

Arvi Rauk

University of Calgary
Department of Chemistry
Calgary, Alberta T2N 1N4
Canada
rauk@ucalgary.ca

Arturo Robertazzi

Università di Cagliari
CNR-INFM SLACS and Dipartimento di
Fisica
S.P. Monserrato-Sestu Km 0.700
I-09042 Monserrato
Italy
robertazzia@gmail.com

Jorge H. Rodriguez

Purdue University
Department of Physics
West Lafayette, IN 47907
USA
jorge-r@physics.purdue.edu

Luis Rodríguez-Santiago

Universitat Autònoma de Barcelona
Departament de Química
Bellaterra 08193
Spain
luis@klingon.uab.es

Ursula Röthlisberger

Federal Institute of Technology – EPFL
Laboratory of Computational Chemistry
and Biochemistry
CH-1015 Lausanne
Switzerland
ursula.roethlisberger@epfl.ch

Debjani Roy

The University of Georgia
Computational Chemistry Annex
Athens, GA 30602-2525
USA
debjani2k5@gmail.com

Lesley R. Rutledge

University of Lethbridge
Department of Chemistry and
Biochemistry
4401 University Drive
Lethbridge, Alberta T1K 3M4
Canada
lesley.rutledge@uleth.ca

Utpal Sarkar

University of Science and Technology
of Lille
UMR CNRS
Laboratory of Physical Metallurgy and
Materials Engineering, LMPGM
8517 Bâtiment C6
59655 Villeneuve Ascq Cedex
France
utpal.sarkar@univ-lille1.fr
utpalchemiitkgp@yahoo.com

Paul von Ragué Schleyer

The University of Georgia
Computational Chemistry Annex
Athens, GA 30602-2525
USA
schleyer@chem.uga.edu

Mariona Sodupe

Universitat Autònoma de Barcelona
Departament de Química
Bellaterra 08193
Spain
Mariona.Sodupe@uab.cat

Miquel Solà

Universitat de Girona
Institut de Química Computacional and
Departament de Química
Campus Montilivi
17071 Girona
Spain
miquel.sola@udg.edu

David N. Stamos

York University
Department of Philosophy
S428 Ross Building, 4700 Keele Street
Toronto, Ontario M3J 1P3
Canada
dstamos@yorku.ca

Marcel Swart

Institució Catalana de Recerca i Estudis
Avançats (ICREA)
Pg. Lluís Companys 23
E-08010 Barcelona
Spain

and

Universitat de Girona
Institut de Química Computacional and
Departament de Química
Campus Montilivi
17071 Girona
Spain
marcel.swart@icrea.es

Ajit J. Thakkar

University of New Brunswick
Department of Chemistry
Fredericton, New Brunswick E3B 5A3
Canada
ajit@unb.ca

Jacopo Tomasi

Università di Pisa
Dipartimento di Chimica e Chimica
Industriale
via Risorgimento 35
I-56126 Pisa
Italy
tomasi@dcc.unipi.it

Alejandro J. Vila

Universidad Nacional de Rosario
Facultad de Ciencias Bioquímicas y
Farmacéuticas
Departamento de Química Biológica
and Instituto de Biología Molecular y
Celular de Rosario (IBR) (CONICET-
UNR)
Suipacha 531
S2002LRK Rosario
Argentina
vila@ibr.gov.ar

Donald F. Weaver

Dalhousie University
Departments of Medicine (Neurology)
and Chemistry
Halifax, Nova Scotia B3H 4J3
Canada
donald.weaver@dal.ca

Thom Vreven

Gaussian, Inc.
340 Quinipiac Street, Building 40
Wallingford, CT 06492

Stacey D. Wetmore

University of Lethbridge
Department of Chemistry and
Biochemistry
4401 University Drive
Lethbridge, Alberta T1K 3M4
Canada
stacey.wetmore@uleth.ca

Ada Yonath

Weizmann Institute of Science
The Helen and Milton A. Kimmelman
Center of Biomolecular Structure
and Assembly
76100
Rehovot
Israel
ada.yonath@weizmann.ac.uk

Part One
Novel Theoretical, Computational, and Experimental
Methods and Techniques

1

Quantum Kernels and Quantum Crystallography: Applications in Biochemistry

Lulu Huang, Lou Massa, and Jerome Karle

1.1

Introduction

Professors Bernard and Alberte Pullman were among the first and most important researchers to apply the notions of quantum mechanics to a great number of molecules of biological importance. It has been often noted that their early work was the beginning of quantum biochemistry, pioneering as they did the application of quantum mechanics to carcinogenic properties of aromatic hydrocarbons. Their quantum computations included the electronic structure of nucleic acids and their mechanisms interacting with various drugs, carcinogens and antitumor compounds. They had success in the interpretation of the role of enzyme constituents important in redox reactions, in calculating stability to ultraviolet radiation, in evaluating the role of functional molecular portions (as opposed to whole molecules) in carcinogen action, and in the evaluation of hydrogen bonding through the amino acid residues as potential pathways for electron transfer. Their landmark book entitled *Quantum Biochemistry* [B. Pullman and A. Pullman, Interscience Publishers (John Wiley & Sons), New York, 1963] has been an inspiration for workers in the research field of the same name as the book title. Their success in quantum biology is all the more impressive today in consideration of the computational difficulty of solving the Schrödinger equation in their time. In this chapter we discuss, the origin of our work in the topic title of this chapter, and certain numerical results of quantum biochemistry made possible since the time of the Pullman's by the enormous increase in computing power that has occurred. Remarkable advances in computing have facilitated the treatment of ever increasing molecular size in both crystallography and quantum mechanics.

1.2

Origins of Quantum Crystallography (QCr)

1.2.1

General Problem of N -Representability

The origins of our work in the field we named quantum crystallography go back to the ideas that originated in the laboratory of Professor William Clinton of the Physics Department at Georgetown University. In a series of papers the Clinton school introduced into crystallography the concept of N -representability.

Over the past many years a voluminous literature concerning the problem of N -representability (Figure 1.1) has arisen [1–11]. Because of the physical indistinguishability of particles, every valid approximation to a solution of the Schrödinger equation must be antisymmetric in the coordinate permutation of fermion pairs. Given such antisymmetric functions, Ψ , one may define reduced density matrices:

$$\rho_p(1 \cdots p, 1' \cdots p') = N(N-1) \cdots (N-p+1) \int \Psi^* \Psi d(p+1) \cdots dN \quad (1.1)$$

The problem of N -representability is that of finding conditions by which to recognize these ρ_p , which are assured to be related to an N -body wavefunction according to the rule of Equation 1.1.

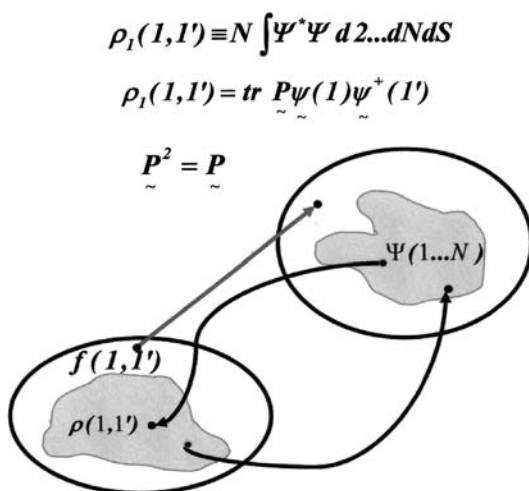


Figure 1.1 Sketch indicating the mapping problem associated with wavefunction representability of density matrices.

Particularly important for the calculation of almost all interesting physical properties are the cases for $p = 2$, and $p = 1$, viz.:

$$\varrho_2(1, 2, 1', 2') = N(N-1) \int \Psi^* \Psi d3 \dots dN \quad (1.2)$$

$$\varrho_1(1, 1') = N \int \Psi^* \Psi d2 \dots dN \quad (1.3)$$

In the case of spinless density matrices, integration occurs over all spins.

For the usual case of Hamiltonians containing at most two-body interactions, the second-order reduced density matrix determines completely the energy of the system.

The problem of finding the conditions that allow the mapping of the objects of Equation 1.1, viz., ϱ_p and Ψ into one another is important mathematically. Moreover, there are important physical and computational aspects to the problem. One sees immediately that, for example, ϱ_2 is, inherently a simpler object than is $\Psi(1 \dots N)$, since it depends only upon the coordinates of two particles, no matter how great is N . Knowledge of N -representable ϱ_2 would allow direct minimization of the energy with respect to the parameters of ϱ_2 , thus eliminating the need for handling an N -body wavefunction. The variation principle, which supplies an upper bound for every approximate ϱ_2 , will hold so long as N -representability of ϱ_2 is satisfied. A practical quantum mechanics might, in such fashion, be framed entirely within the context of density matrices without any explicit computational role played by N -body wavefunctions.

The problem of N -representability is still a subject of current interest. Although very much has been learned the complete problem of N -representability of ϱ_2 has not been solved. Interestingly, the case of N -representability by a single determinant of orbitals is well understood. Idempotency of the one-body density matrix ϱ_1 completely characterizes this case, for which moreover all higher order density matrices are known functionals of ϱ_1 . Of course independent particle models, including the Hartree–Fock and density functional theory cases, are all encompassed within single determinant wavefunctions. The N -representability problem is solved, as far as single Slater determinants are concerned, [6]. Another case for which N -representability is no difficulty occurs for the density itself, that is, $\varrho(1) = \varrho_1(1, 1')|_{1' \rightarrow 1}$, the diagonal elements of the one-body density matrix. It occurs by a theorem of Gilbert [12] that any normalized, well behaved density is N -representable by a single Slater determinant of orbitals. We have shown by calculations with select examples that an exact density is N -representable by a Slater determinant of physically meaningful orbitals [13].

1.2.2

Single Determinant N -Representability

In one case, that characterized by a Slater determinant wavefunction, N -representability of reduced density matrices presents no problem. Such density matrices have

been studied exhaustively and their properties are well understood. We review points of interest.

We take a set of orthonormal molecular spin orbitals $\{\phi_i(i=1 \dots N)\}$ and with them construct a Slater determinant (an antisymmetric function carrying the physical implications of the Pauli principle):

$$\Psi_{\text{det}}(1 \dots N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi(1) & \dots & \phi_1(N) \\ \vdots & & \vdots \\ \phi_N(1) & \dots & \phi_N(N) \end{vmatrix} \quad (1.4)$$

Such a determinant satisfies the normalization condition:

$$\int \Psi^* \Psi \, d1 \dots dN = 1 \quad (1.5)$$

By direct integration over the product of the Slater determinant with itself, the reduced density matrices of every order may be constructed. For example:

$$\mathcal{Q}_{N\text{det}} = N! \Psi_{\text{det}}^* \Psi_{\text{det}} = \begin{vmatrix} \mathcal{Q}(1, 1') & \dots & \mathcal{Q}(1, N') \\ \vdots & & \vdots \\ \mathcal{Q}(N, 1') & \dots & \mathcal{Q}(N, N') \end{vmatrix} \quad (1.6)$$

$$\mathcal{Q}_{2\text{det}} = N(N-1) \int \Psi_{\text{det}}^* \Psi_{\text{det}} \, d3 \dots dN = \begin{vmatrix} \mathcal{Q}(1, 1') & \mathcal{Q}(1, 2') \\ \mathcal{Q}(2, 1') & \mathcal{Q}(2, 2') \end{vmatrix} \quad (1.7)$$

$$\mathcal{Q}_{1\text{det}} = N \int \Psi_{\text{det}}^* \Psi_{\text{det}} \, d2 \dots dN = \mathcal{Q}(1, 1') \quad (1.8)$$

The necessary and sufficient conditions for this one-body density matrix to be N -representable by a single Slater determinant are:

$$\mathcal{Q}_1^2 = \mathcal{Q}_1, \quad \int \mathcal{Q}_1 \, d1 = N, \quad \mathcal{Q}_1^\dagger = \mathcal{Q}_1 \quad (1.9)$$

The density matrix must be idempotent, normalized and hermetian, conditions both simple and of practical utility. McWeeney [8] has shown that a density matrix may be purified to idempotency via an iterative expression and also that an idempotent density matrix can always be factored into a sum of squares of orbitals. The orbitals are not unique in the sense that the one-body density matrix is invariant to a unitary transformation among them. Knowledge of $\mathcal{Q}_{1\text{-det}}$ fixes $\mathcal{Q}_{2\text{-det}}$ and every higher reduced density matrix up to and including $\mathcal{Q}_{N\text{-det}}$, and Ψ_{det} itself. For a two-body Hamiltonian of the usual type:

$$\hat{H} = \sum \hat{h}_i + \sum \hat{h}_{ij} \quad (1.10)$$

the energy:

$$E = \int \hat{h}_1 \mathcal{Q}_{1\text{det}}(1, 1')|_{1' \rightarrow 1} \, d1 + \int \hat{h}_{12} \mathcal{Q}_{2\text{det}}(1, 2) \, d1 \, d2 \geq E_0 \quad (1.11)$$

satisfies the variational theorem. We mention in passing, for the above expression of the energy, that the off-diagonal elements of $\mathcal{Q}_1$ are required, but only the diagonal

elements of ρ_2 . E is, of course, invariant to a unitary transformation among the orbitals. Direct minimization of E , expressed by Equation 1.11, produces the approximate Hartree–Fock energy appropriate to the basis used for expansion of the density matrix.

According to the theorem of Gilbert [12] *every* well-behaved electron density (positive and normalized) is N -representable by a single Slater determinant of orbitals. Of course this is obvious for any Hartree–Fock density, but interestingly the theorem is totally general, and holds equally well for the exact density corresponding to the full Hamiltonian. Every $\rho(1)$ is N -representable by some $\Psi_{\text{det}}(1 \dots N)$.

McWeeney's purification to idempotency [8] may be modified to include conditions of constraint as in Clinton's equations [14]:

$$\mathbf{P}_{n+1} = 3\mathbf{P}_n^2 - 2\mathbf{P}_n^3 + \sum_k \lambda_k \mathbf{O}_k + \lambda_N \mathbf{1} \quad (1.12)$$

In Equation 1.12 the λ 's are Lagrangian multipliers determined from equations of constraint, for example:

$$\mathbf{O}_k = \text{tr } \mathbf{P} \mathbf{O}_k \quad (1.13)$$

$$1 = \text{tr } \mathbf{P} \mathbf{1} \quad (1.14)$$

where $\mathbf{O}_k$ is the matrix representative of an arbitrary quantum operator $\mathbf{O}_k$ and $\mathbf{1}$ is the matrix representative of the normalization operator $\mathbf{1}$. $\mathbf{P}$ is the Löwdin population matrix or density matrix in an orthonormal basis.

Clinton's equations have the physical significance of delivering a one-body density matrix, N -representable by a single Slater determinant, and satisfying chosen quantum conditions of constraint. Applied in context of the X-ray coherent diffraction experiment [15] these equations can deliver the exact experimental electron density. For such a case, the experimental Bragg structure factors $F(K)$ provide conditions of constraint via the Fourier transform relation:

$$F(\mathbf{K}) = \int e^{i\mathbf{K} \cdot \mathbf{r}} \rho(\mathbf{r}) d^3\mathbf{r} \quad (1.15)$$

where the electron density is:

$$\rho(\mathbf{r}) = \rho(\mathbf{r}, \mathbf{r}')|_{\mathbf{r}' \rightarrow \mathbf{r}} \quad (1.16)$$

the diagonal elements of the density matrix. Clinton's equations, applied with an appropriate basis, are capable, consistent with Gilbert's theorem, of delivering physically meaningful orbitals that satisfy the experimental (and therefore exact) density. Within quantum crystallography, this has proven to be one of their important uses.

1.2.3

Example Applications of Clinton's Equations

1.2.3.1 Beryllium

We applied the Clinton equations to a beryllium crystal using the very accurate X-ray scattering factor data of Larsen and Hansen [15]. As may be seen in Figure 1.2 the

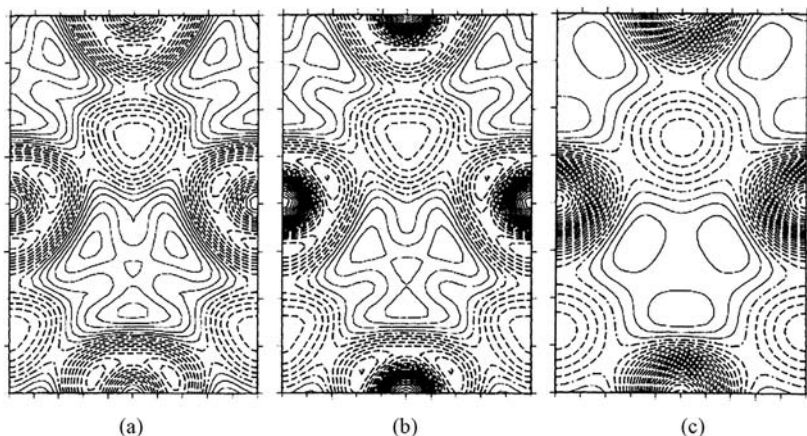


Figure 1.2 Valence density from (a) Dovesi *et al.* [16], (b) this work [15] and (c) Chou, Lam and Cohen [17]. Projections of the tetrahedral and octahedral holes are indicated..

experimental density obtained was quite accurate, as measured by reference to the best theoretical densities that were available for that crystal. The experimental density contours are very similar to the theoretical contours [15–17]. In Figure 1.3 the errors in scattering factor F are plotted as a function of scattering angle [15]. The errors are randomly distributed out to high angles of scattering. At the time of this result, $R = 0.0018$, achieved with an N -representable density matrix, was perhaps the smallest R factor in the literature of crystallography. This established that N -representable density descriptions of actual X-ray scattering data were practicable and of high accuracy.

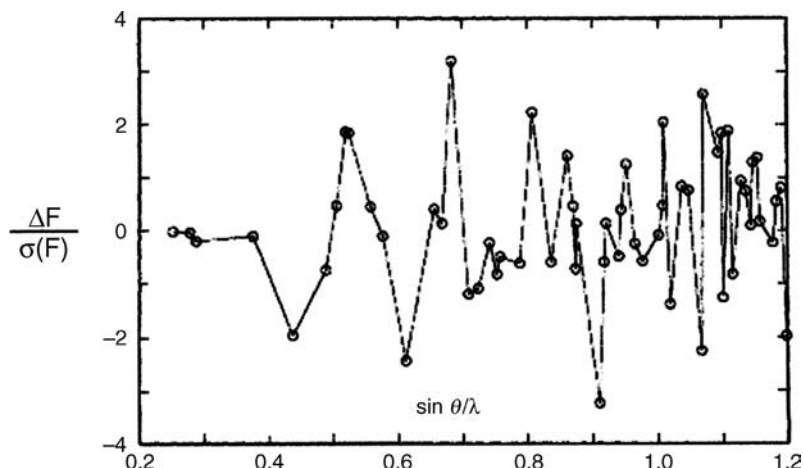


Figure 1.3 Distribution of errors [15]; $R_{\text{wr}} = 0.0018$ and G.O.F. = 1.33.

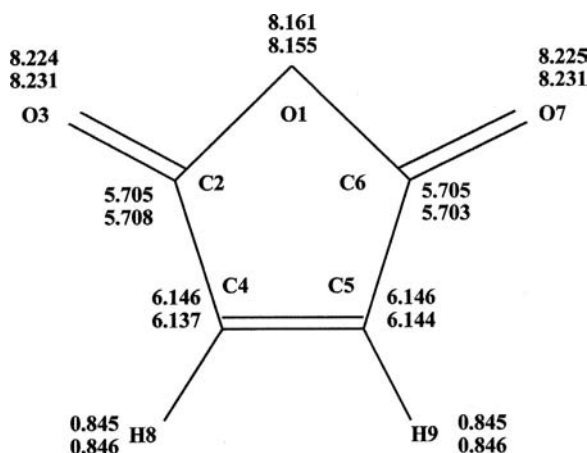


Figure 1.4 Electrons per atom in maleic anhydride. Upper numbers obtained from optimized theoretical calculation with B3LYP/cc-pVTZ. Lower numbers obtained from experimental coordinates and a single point calculation with B3LYP/cc-pVTZ.

1.2.3.2 Maleic Anhydride

This section concerns the application of Clinton's equations to a crystal of maleic anhydride [18], a small, flat molecule, having only nine atoms (Figure 1.4). Data collection and crystallographic refinement for this study were carried out by Louis Todaro. The authors refined the elements of the projector matrix by use of the Clinton iterative equations and the structure factor magnitudes obtained from an X-ray diffraction investigation. The final *R*-factor between the experimental structure factor magnitudes and the theoretical ones from the projector matrix for 6-31G** was less than 1.5%.

A total of 507 independent data were used. The experimental data were collected with CuK α radiation at 110(1) K. A calculation of the resolution of these data yielded a value of about 0.80 Å, and the number of independent elements in the projector matrix was 2250. The total number of data available for the refinement of the elements in the projector matrix was $8 \times 507 = 4056$, and so the ratio of data to independent unknowns was 1.80. After the independent data were corrected for vibrational effects and expanded to include all equivalent reflections for space group $P2_12_12_1$, the following results were obtained.

Tables 1.1 and 1.2 display calculated energies and atomic charges, respectively. Clinton's equations yielded both an experimental density matrix and experimental atomic coordinates. There was no significant difference in the coordinates obtained using Clintons equations and those obtained from an ordinary crystallographic least-squares determination, except for the hydrogen atoms, which are placed differently in X-ray diffraction experiments and in quantum mechanical modeling. The implications for maleic anhydride were that perhaps an accurate and efficient way to combine diffraction data with quantum mechanics is to use the heavy atom coordinates obtained crystallographically, holding them fixed, and then carry out the *ab initio* quantum mechanical calculations for the system. The burden for obtaining

Table 1.1 Energies for maleic anhydride.

	Energies (au) ^{a)}					
	E_{total}	T	NE	EE	NN	$-V/T$
Exp. Coord. ^{b)}	-379.432	377.065	-1439.600	407.672	275.431	2.0063
OPT ^{c)}	-379.435	377.126	-1439.658	407.659	275.438	2.0061

a) Total energy (E_{total}); total electronic kinetic energy (T); total nuclei–electrons attractive potential energy (NE); total electron–electron repulsion energy (EE); total nuclear–nuclear repulsion energy (NN); and the negative of the virial ratio ($-V/T$) the ratio of the potential energy ($V = NE + EE + NN$) to the kinetic energy (T), a ratio that should ideally be exactly 2 according to the “virial theorem” (in a calculation of infinite precision).

b) The single point calculation was performed with the use of the experimental coordinates and B3LYP/cc-pVTZ.

c) OPT refers to geometry optimization with B3LYP/cc-pVTZ.

Table 1.2 Electrons per atom for maleic anhydride.

	Atoms								
	H ₈	C ₄	O ₃	C ₂	O ₁	H ₉	C ₅	O ₇	C ₆
Exp. Coord. ^{a)}	0.846	6.137	8.231	5.708	8.155	0.846	6.144	8.231	5.703
OPT ^{b)}	0.845	6.146	8.224	5.705	8.161	0.845	6.146	8.225	5.705

a) Single point calculations were performed with the use of the experimental coordinates and B3LYP/cc-pVTZ.

b) OPT refers to geometry optimization with B3LYP/cc-pVTZ.

quantum mechanical information is then placed upon use of a sufficiently accurate chemical model, and the problem of atomic coordinates is simply taken from the normal crystallography. This observation had an influence in the creation of the kernel energy method discussed below. The experimental density matrix obtained from Clinton’s equations delivered energies and atomic charges similar to those obtained directly from the density functional theory calculations at the experimental coordinates, the latter of which are shown in Figure 1.4, and are compared to the analogous atomic charges at the DFT optimized coordinates. The overall result is a close correspondence between the N -representable experimental and theoretically calculated charge distribution and energies for the maleic anhydride molecule.

1.3

Beginnings of Quantum Kernels

1.3.1

Computational Difficulty of Large Molecules

Large molecules are a special problem. For example, the computational difficulty of solving the Schrödinger equation increases with a high power of the number of atoms

(or basis functions) in the molecule. In addition, when fixing the elements of the density matrix by a least-squares fit to the X-ray scattering data, it is desirable that the number of data should exceed in good measure the number of matrix elements. But, as the size of a molecule increases the ratio of number of data to number of matrix elements tends to become too small for a reliable determination of the density matrix. The desire to represent increasingly large molecules forced us to consider how to surmount the computational difficulties associated with size. This led to a simple idea, variations of which had occurred to tens of different research groups. That idea was to take a large molecule, break it into smaller pieces, represent the smaller and more tractable pieces, and then put them back together in such fashion as to reconstitute a representation of the original large molecule. One particular method in which this idea is carried out occurs within quantum crystallography.

1.3.2

Quantum Kernel Formalism

The basic formalism that introduces the important idea of the essential molecular pieces, called kernels, [19, 20] is presented in the following paragraphs.

The kernel calculations to be presented here are based on structural data, that is, atomic positions. X-Ray scattering data are used routinely to determine molecular structure, that is, equilibrium atomic arrangements and thermal (disorder) parameters. The same data, when sufficiently accurate, can also be used to obtain the electron density distribution of the unit cell of a crystal [21]. The electron density distribution for a crystal, ϱ , can also be expressed in terms of the trace of a suitable matrix product [13–15, 22–27] according to:

$$\varrho = 2\text{tr } \boldsymbol{\varphi} \boldsymbol{\varphi}^\dagger \quad (1.17)$$

The column matrix $\boldsymbol{\varphi}$ is composed of doubly occupied orthonormal molecular orbitals, giving rise to the factor of 2. Most molecular ground states have doubly occupied orbitals. In other cases, the formalism may be appropriately generalized.

If we write:

$$\boldsymbol{\varphi} = \mathbf{C}\boldsymbol{\Psi} \quad (1.18)$$

Equation 1.17 becomes:

$$\varrho = 2\text{tr } \mathbf{C}\boldsymbol{\Psi}\boldsymbol{\Psi}^\dagger\mathbf{C}^\dagger = 2\text{tr } \mathbf{C}^\dagger\mathbf{C}\boldsymbol{\Psi}\boldsymbol{\Psi}^\dagger \quad (1.19)$$

The value of the trace is insensitive to the cyclic interchange of the position of $\mathbf{C}^\dagger$. The following definitions are made:

$$\mathbf{S} = \int \boldsymbol{\Psi}\boldsymbol{\Psi}^\dagger \, d\mathbf{r} \quad (1.20)$$

where the integration is performed over the individual elements of the product matrix $\boldsymbol{\Psi}\boldsymbol{\Psi}^\dagger$:

$$\underline{\mathbf{R}} = \underline{\mathbf{C}}^\dagger \underline{\mathbf{C}} \quad (1.21)$$

and:

$$\underline{\mathbf{R}} \underline{\mathbf{S}} = \underline{\mathbf{P}}_{\mathbf{a}} \quad (1.22)$$

where $\underline{\mathbf{P}}_{\mathbf{a}}$ is a projector. The subscript $\mathbf{a}$ indicates that unless special steps are taken in forming the matrix $\underline{\Psi}$ the projector matrix will not be symmetric. It can be shown that, as a consequence of the fact that the $\underline{\boldsymbol{\varphi}}$ are composed of elements that are orthonormal:

$$\underline{\mathbf{P}}_{\mathbf{a}}^2 = \underline{\mathbf{P}}_{\mathbf{a}} \quad (1.23)$$

and:

$$\text{tr } \underline{\mathbf{P}}_{\mathbf{a}} = N \quad (1.24)$$

where N is the number of doubly occupied orbitals in the molecule of interest (N should not to be confused with the symbol number of electrons as the context indicates elsewhere in this chapter). Equation 1.23 is the projector property.

It is convenient to have a projector $\underline{\mathbf{P}}_{\mathbf{s}}$ that is symmetric since it reduces the number of elements in the projector that must be evaluated. From Equation 1.22, it follows that:

$$\underline{\mathbf{S}}^{1/2} \underline{\mathbf{R}} \underline{\mathbf{S}} \underline{\mathbf{S}}^{-1/2} = \underline{\mathbf{S}}^{1/2} \underline{\mathbf{P}}_{\mathbf{a}} \underline{\mathbf{S}}^{-1/2} \quad (1.25)$$

This matrix product is a symmetric projector $\underline{\mathbf{P}}_{\mathbf{s}}$ and may be written:

$$\underline{\mathbf{P}}_{\mathbf{s}} = \underline{\mathbf{S}}^{1/2} \underline{\mathbf{R}} \underline{\mathbf{S}}^{1/2} \quad (1.26)$$

It follows from Equations 1.20–1.22 and 1.25 that the electron density can be written:

$$\rho = 2\text{tr } \underline{\mathbf{R}} \underline{\Psi} \underline{\Psi}^\dagger = 2\text{tr } \underline{\mathbf{P}}_{\mathbf{a}} \underline{\mathbf{S}}^{-1} \underline{\Psi} \underline{\Psi}^\dagger \quad (1.27)$$

and:

$$\rho = 2\text{tr } \underline{\mathbf{P}}_{\mathbf{s}} \underline{\mathbf{S}}^{-1/2} \underline{\Psi} \underline{\Psi}^\dagger \underline{\mathbf{S}}^{-1/2} \quad (1.28)$$

It will be seen that in the application of the calculation of fragment densities to obtain kernel densities it is convenient to compute the projector $\underline{\mathbf{P}}_{\mathbf{a}}$. In the further application of quantum crystallography, to adjust the values of the projector with the use of diffraction data from a crystal, it is more suitable to use $\underline{\mathbf{P}}_{\mathbf{s}}$.

There is a third type of projector, $\underline{\mathbf{P}}_{\mathbf{s}\Gamma}$, that is useful because it is a symmetric projector that has fewer elements than $\underline{\mathbf{P}}_{\mathbf{s}}$. It arises from the use of point group symmetry to form symmetry orbitals as a basis for the molecular orbitals. Matrices $\underline{\mathbf{T}}_{\mathbf{s}\Gamma}$ associated with the irreducible representations of the point group of a molecule [28] can be formed that transform atomic orbitals into symmetry orbitals by the operation $\underline{\mathbf{T}}_{\mathbf{s}\Gamma} \underline{\Psi}_{\mathbf{m}}$, where the subscript $\mathbf{s}$ associates $\underline{\mathbf{T}}$ with symmetry orbitals and the subscript Γ associates $\underline{\mathbf{T}}$ with the irreducible representations. The subscript $\mathbf{m}$ denotes the fact that $\underline{\Psi}_{\mathbf{m}}$ is composed of orbitals for a molecule (not the entire unit cell). The coefficients associated with $\underline{\mathbf{T}}_{\mathbf{s}\Gamma} \underline{\Psi}_{\mathbf{m}}$ are denoted by $\underline{\mathbf{C}}_{\Gamma}$,

giving:

$$Q = \sum_{\Gamma} 2\text{tr } C_{\Gamma}^{\dagger} C_{\Gamma} T_{s\Gamma} \left[\sum_{\hat{R}} \hat{R} \Psi_m \Psi_m^{\dagger} \right] T_{s\Gamma}^{\dagger}, \quad (1.29)$$

or:

$$Q = \sum_{\Gamma} 2\text{tr } R_{\Gamma} T_{s\Gamma} \left[\sum_{\hat{R}} \hat{R} \Psi_m \Psi_m^{\dagger} \right] T_{s\Gamma}^{\dagger} \quad (1.30)$$

where $\hat{R}$ represents the symmetry operations of the crystallographic space group of interest and Ψ_m are composed of the atomic orbitals for a molecule. To change R_{Γ} into a symmetric projector, we write an expression equivalent to Equation 1.30:

$$Q = \sum_{\Gamma} 2\text{tr } S_{\Gamma}^{1/2} R_{\Gamma} S_{\Gamma}^{1/2} S_{\Gamma}^{-1/2} T_{s\Gamma} \times \left[\sum_{\hat{R}} \hat{R} \Psi_m \Psi_m^{\dagger} \right] T_{s\Gamma}^{\dagger} S_{\Gamma}^{\dagger -1/2} \quad (1.31)$$

or:

$$Q = \sum_{\Gamma} 2\text{tr } P_{s\Gamma} S_{\Gamma}^{-1/2} T_{s\Gamma} \times \left[\sum_{\hat{R}} \hat{R} \Psi_m \Psi_m^{\dagger} \right] T_{s\Gamma}^{\dagger} S_{\Gamma}^{\dagger -1/2} \quad (1.32)$$

$P_{s\Gamma}$ is symmetric, is associated with symmetry orbitals and:

$$S_{\Gamma} = \int T_{s\Gamma} \Psi_m \Psi_m^{\dagger} T_{s\Gamma}^{\dagger} d\mathbf{r} \quad (1.33)$$

where the integration over all space is performed for all individual elements of the product matrix, $\Psi_m \Psi_m^{\dagger}$.

In the single-determinant approach taken here, the Fourier transforms of Equations 1.28 or 1.32 may be considered to be the basic equations of quantum crystallography. Their Fourier transforms yield the structure factors of crystallographic theory, whose magnitudes are definable in terms of the measured diffraction intensities. The mathematical objective of quantum crystallography is to optimize the fit of the elements of the projector matrix to the experimental structure factor magnitudes and also the fit of some other parameters that occur in the Fourier transform of the right-hand side of Equations 1.28 or 1.32. In addition to the positional coordinates of the atoms, adjustments are made to three scaling factors, which set the average value of the calculated structure factor magnitudes equal to the average of the observed one. Provision may also be made, in quantum crystallography, to adjust the value of thermal parameters attached to the atomic basis orbitals, which have the effect of simulating a smearing of density due to atomic motions.

The fragment calculations that will now be described deliver parts of the R matrix with good accuracy. They may then be assembled into the complete R matrix, and by use of Equation 1.26 the symmetric projector P_s may be formed for use in the quantum crystallography calculations.

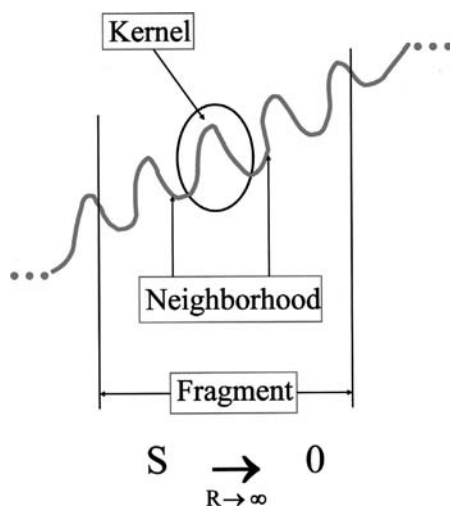


Figure 1.5 Kernel, neighborhood and fragment.

1.3.3

Kernel Matrices: Example and Results

The purpose of kernel calculations is to obtain an accurate $\mathbf{R}$ matrix when *ab initio* calculations of an entire molecule are either not feasible or considered to be too time-consuming. As illustrated in Figure 1.5, a fragment consists of an inner core or “kernel” and several neighboring atoms called a “neighborhood.” The molecule is divided into a suitable number of kernels, which, when recombined, form the complete molecule.

Since the coordinates of the structure of interest are available, it is readily possible to calculate which atoms would occur within a certain chosen distance from all the atoms in a kernel. Such atoms would form the neighborhood. To maintain an electron balance, it may be necessary to attach hydrogen atoms to some of the neighborhood atoms in a fragment. There are various schemes conceivable for choosing the ways in which a molecule may be broken up into kernels and neighborhoods. One of general applicability, which still allows some arbitrary choice, can be based on the rule that all atoms present must be a member of some kernel once and only once. With atomic positions held fixed, the electron density distribution in a fragment is computed. Such a calculation can deliver contributions to an $\mathbf{R}$ matrix from which the portion that concerns the kernel is saved. Those contributions to the $\mathbf{R}$ matrix involving orbitals from a neighborhood atom and an atom in the kernel are saved at the fractional value of one-half, in accordance with the above rule. If all neighborhood atoms occur only as part of a kernel, another one-half value would be added to those contributions already saved at one-half values, when the values associated with the adjoining kernels are calculated. Contributions from pairs of atoms, both in the same kernel, are saved with a coefficient of one. The final $\mathbf{R}$ matrix will be multiplied by the $\mathbf{S}$ matrix to give $\mathbf{P}_a$, and since $\mathbf{S}$ is an overlap matrix, values close to zero will be

obtained for pairs of atoms that are separated by large distances. The pattern of zeros in $\mathbf{S}$ is used to generate zeros in $\mathbf{R}$, justified by the symmetry of $\mathbf{S}$, namely, $\mathbf{S} = \mathbf{S}^\dagger$, and the invariance of $\text{Tr } \mathbf{P}\mathbf{S}$ to the insertion into $\mathbf{R}$ of the pattern of zeros in $\mathbf{S}$. The behavior of $\mathbf{S}$ is the reason why the fragment calculations can give accurate values for the molecule as a whole.

The kernel calculations for a hydrated hexapeptide [29] were performed by defining the kernels as the six peptide residues in the ring with each of the three water molecules associated with the appropriate residues as determined by proximity. The neighborhoods in the fragments were formed by the amino acid residues and associated water molecules, if any, adjoining the one considered as the kernel, for example, residues 3 and 5 were the neighborhood for residue 4 acting as a kernel.

We may write the $\mathbf{R}$ -matrix for the full hexapeptide molecule as:

$$\mathbf{R} = \begin{vmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \cdots & \mathbf{R}_{16} \\ \mathbf{R}_{21} & \mathbf{R}_{22} & \cdots & \mathbf{R}_{26} \\ \vdots & & & \vdots \\ \mathbf{R}_{61} & \mathbf{R}_{62} & \cdots & \mathbf{R}_{66} \end{vmatrix} \quad (1.34)$$

where the subscripts refer to each of the six kernels composed of the six amino acid residues (some associated with water molecules) in the hexapeptide. Each element of the matrix Equation 1.34 is itself a matrix whose dimensions are those of the bases associated with each of the kernels labeled by the subscripts.

A matrix associated with each of the six kernels, $\mathbf{R}_j$ ($j = 1, \dots, 6$), may be formed consistent with the rules of Mulliken population analysis, giving:

$$\mathbf{R} = \sum_{j=1}^6 \mathbf{R}_j \quad (1.35)$$

where $\mathbf{R}_j$ is composed of the sum of two matrices, one whose only nonzero components are $0.5\mathbf{R}_{jk}$ ($k = 1, 2, \dots, 6$) and one whose only nonzero components are $0.5\mathbf{R}_{kj}$ ($k = 1, 2, \dots, 6$). For example, when $j = 4$:

$$\mathbf{R}_4 = \begin{vmatrix} & & & \mathbf{R}_{14}/2 & & \\ & 0 & & \mathbf{R}_{24}/2 & & 0 \\ & & & \mathbf{R}_{34}/2 & & \\ \mathbf{R}_{41}/2 & \mathbf{R}_{42}/2 & \mathbf{R}_{43}/2 & \mathbf{R}_{44} & \mathbf{R}_{45}/2 & \mathbf{R}_{46}/2 \\ & 0 & & \mathbf{R}_{54}/2 & & 0 \\ & & & \mathbf{R}_{64}/2 & & \end{vmatrix} \quad (1.36)$$

The correspondence of Equations 1.34 and 1.35 may be readily verified. The approximation was made that each kernel “overlaps” with a neighborhood that includes only one kernel on either side of the given kernel. This limits the range of k in the $\mathbf{R}_j$ of Equation 1.35 to $k = j - 1, j, j + 1$ instead of $k = 1, 2, \dots, 6$, with $k = 0$

equivalent to $k = 6$ and $k = 7$ equivalent to $k = 1$. Equation 1.36 becomes:

$$\mathbf{R}_4(0) = \begin{vmatrix} & & 0 & & & \\ & 0 & 0 & & 0 & \\ & & \mathbf{R}_{34}/2 & & & \\ 0 & 0 & \mathbf{R}_{43}/2 & \mathbf{R}_{44} & \mathbf{R}_{45}/2 & 0 \\ & 0 & \mathbf{R}_{54}/2 & & 0 & \end{vmatrix} \quad (1.37)$$

where $\mathbf{R}_4(0)$ indicates that $\mathbf{R}_4$ is modified by the introduction of truncated neighborhoods, which introduces a pattern of zeros.

The full molecule R-matrix is approximated by summing over the kernel matrices associated with truncated neighborhoods, that is:

$$\mathbf{R}(0) = \sum_{j=1}^6 \mathbf{R}_j(0) \quad (1.38)$$

which, when written out, is:

$$\mathbf{R}(0) = \begin{vmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & 0 & 0 & 0 & \mathbf{R}_{16} \\ \mathbf{R}_{21} & \mathbf{R}_{22} & \mathbf{R}_{23} & 0 & 0 & 0 \\ 0 & \mathbf{R}_{32} & \mathbf{R}_{33} & \mathbf{R}_{34} & 0 & 0 \\ 0 & 0 & \mathbf{R}_{43} & \mathbf{R}_{44} & \mathbf{R}_{45} & 0 \\ 0 & 0 & 0 & \mathbf{R}_{54} & \mathbf{R}_{55} & \mathbf{R}_{56} \\ \mathbf{R}_{61} & 0 & 0 & 0 & \mathbf{R}_{65} & \mathbf{R}_{66} \end{vmatrix} \quad (1.39)$$

Thus, the electron density distribution for the full molecule is approximately:

$$\varrho(0) = 2\text{tr } \mathbf{R}(0) \left[\mathbf{\Psi} \mathbf{\Psi}^\dagger \right] (0) \quad (1.40)$$

where the last (0) indicates a pattern of zeros in the matrix $\left[\mathbf{\Psi} \mathbf{\Psi}^\dagger \right]$ analogous to the pattern of zeros in $\mathbf{R}(0)$. The pattern of zeros in $\left[\mathbf{\Psi} \mathbf{\Psi}^\dagger \right] (0)$ is the same as in $\mathbf{S}(0)$ and $\mathbf{R}(0)$. Not only are the overlap integrals very small for the product of those elements that are set equal to zero, the product before integration is also very small. We see how a density function for a complete molecule may be obtained approximately from matrices of smaller kernels.

Approximate matrices based on kernels may produce electron densities whose suitability may be further enhanced by ensuring their N -representability [6–8]. This is achieved by requiring the matrices to be normalized projectors. These properties may be imposed on a matrix $\mathbf{R}$ [and also $\mathbf{R}(0)$] by use of Clinton's iterative equations [14] in the form:

$$\mathbf{R}_{n+1} = 3\mathbf{R}_n \mathbf{S} \mathbf{R}_n - 2\mathbf{R}_n \mathbf{S} \mathbf{R}_n \mathbf{S} \mathbf{R}_n + \lambda \mathbf{1} \quad (1.41)$$

subject to the normalization condition given by:

$$\text{tr } \mathbf{R} \mathbf{S} = N \quad (1.42)$$

where $N = 113$ is the number of doubly occupied molecular orbitals for the hydrated hexapeptide. Condition 1.42 requires that:

$$\lambda = [N - \text{tr} (3\mathbf{R}_n \mathbf{S} \mathbf{R}_n \mathbf{S} - 2\mathbf{R}_n \mathbf{S} \mathbf{R}_n \mathbf{S} \mathbf{S})] / M \quad (1.43)$$

where $M = 173$ is the dimension of the Gaussian basis: $\phi_i = \sum_{j=1}^{173} C_{ij} \psi_j$

1.3.4

Applications of the Idea of Kernels

1.3.4.1 Hydrated Hexapeptide Molecule

Isodensity surfaces have been calculated for a hydrated hexapeptide molecule [29], $\text{c[Gly-Gly-D-Ala-D-Ala-Gly-Gly]} \cdot 3\text{H}_2\text{O}$, by use of Equation 1.40 from the Hartree–Fock orbitals for the fully hydrated hexapeptide molecule, the orbitals associated with the $\mathbf{R}(0)$ matrix obtained from the sum over the $\mathbf{R}(0)$ for the six kernels, and the orbitals associated with the $\mathbf{R}(0)$ matrix obtained from the Clinton iterative equations. The three types of density appear to be quite similar. Therefore, only that for the Hartree–Fock orbitals at an isodensity surface of $0.23 \text{ e } \text{\AA}^{-3}$ is shown in Figure 1.6a. To obtain a more quantitative insight into the similarity among the three densities, a series of difference isodensity surfaces were calculated in which differences that did not exceed increasingly larger values were omitted.

Figure 1.6b and c shows the difference isodensity surfaces.

Figure 1.6b and c were obtained from $\mathbf{R}_{\text{HF}} - \mathbf{R}_{\text{K}}(0)$ and $\mathbf{R}_{\text{HF}} - \mathbf{R}_{\text{P}}(0)$, respectively, by use of Equation 1.40 where the subscripts imply Hartree–Fock (HF), a sum over kernels (K), and a projector (P) with a more accurate projector property obtained by

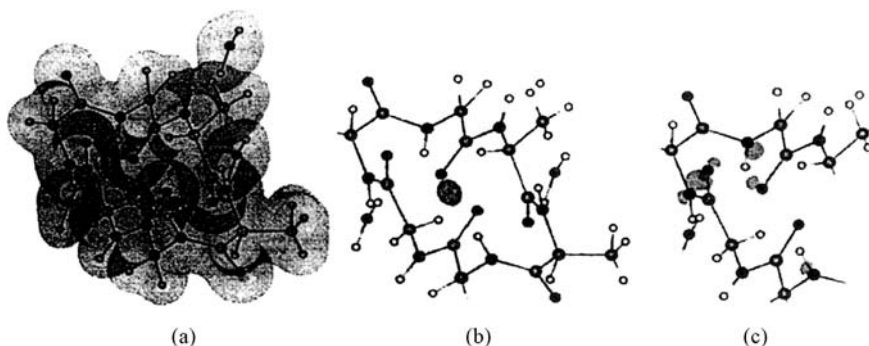


Figure 1.6 (a) Isodensity surface of $0.023 \text{ e } \text{\AA}^{-3}$ for the cyclic hexapeptide trihydrate.

(b) Difference isodensity surface of $5 \times 10^{-4} \text{ e } \text{\AA}^{-3}$ for the cyclic hexapeptide trihydrate. The difference isodensity was obtained from $\mathbf{R}_{\text{HF}} - \mathbf{R}_{\text{K}}(0)$. The small fuzzy region near the center of the diagram, representing the remaining difference isodensity surface, encloses a very small fraction of the

molecular volume. (c) Difference isodensity surface of $3 \times 10^{-3} \text{ e } \text{\AA}^{-3}$ for the cyclic hexapeptide trihydrate. The difference isodensity was obtained from $\mathbf{R}_{\text{HF}} - \mathbf{R}_{\text{P}}(0)$. The small fuzzy regions near the ring represent the difference isodensity surface. They are close to disappearing, enclosing a very small fraction of the molecular volume.

use of Equations 1.41–1.43. For the case of the sum over kernels, a difference isodensity surface of $5 \times 10^{-4} \text{ e } \text{\AA}^{-3}$ is indicated in Figure 1.6b by the small fuzzy region in the center. It encloses a very small fraction of the molecular volume in the center of the molecular framework. The isodensity surface disappears entirely somewhere between 10^{-4} and $10^{-3} \text{ e } \text{\AA}^{-3}$. For the case of $\mathbf{R}(0)$ enhanced to form a more accurate projector matrix, the differences are somewhat larger before they disappear, that is, somewhere between 10^{-3} and $10^{-2} \text{ e } \text{\AA}^{-3}$. The isodensity surface of $3 \times 10^{-3} \text{ e } \text{\AA}^{-3}$ encloses a very small fraction of the molecular volume, as indicated in Figure 1.6c by the tiny fuzzy regions near the ring.

These difference studies indicate that the Hartree–Fock density is closely approximated by the density obtained from the sum over kernels and by the density obtained from enhancing the projector property. This indicates that it is possible to find a projector, $\mathbf{P}(0) = \mathbf{R}(0) \mathbf{S}(0)$, and thus an N -representable matrix of the same simplified form as that of the sum over kernels that gives a good approximation to the Hartree–Fock matrix.

1.3.4.2 Hydrated Leu¹-Zervamicin

Fragment Calculations The fragment calculations for the Leu¹-zervamicin [30] (Figure 1.7a) were performed by defining 19 kernels as the 16 peptide residues,

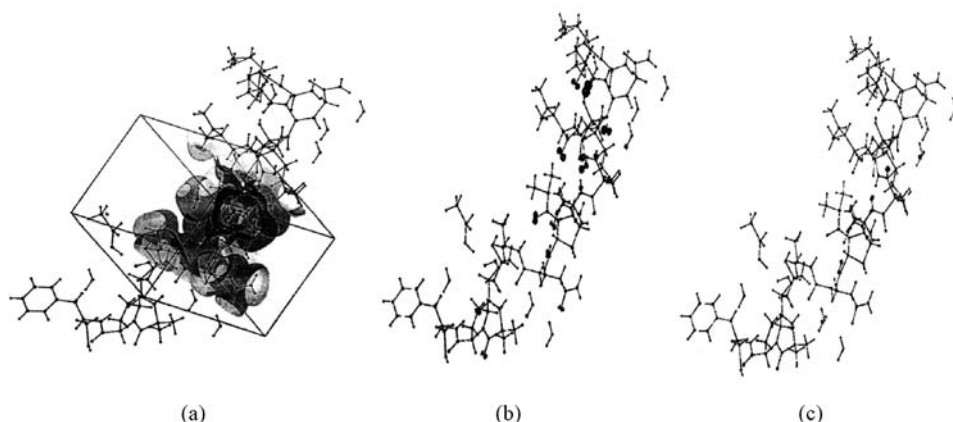


Figure 1.7 (a) Isodensity surface of $0.005 \text{ e } \text{\AA}^{-3}$ within a selected volume for hydrated Leu¹-zervamicin. This isodensity surface was obtained from a Hartree–Fock calculation performed on the entire hydrated molecule and the further use of $\phi_i = \sum_{j=1}^{835} C_{ij} \psi_j$ and $\rho = 2 \sum_{i=1}^{524} \phi_i \phi_i$ applied to the resulting wavefunctions. A ball-and-stick model of the molecule is superimposed on the isodensity surface. (b) Difference isodensity surface of $1.0 \times 10^{-3} \text{ e } \text{\AA}^{-3}$ for the hydrated Leu¹-zervamicin molecule. The difference isodensity

surface was obtained from $P_F - P_K$. Small fuzzy regions represent the remaining difference. They involve a small molecular volume and are evidently small in magnitude. (c) Difference isodensity surface of $-1.2 \times 10^{-3} \text{ e } \text{\AA}^{-3}$ for the hydrated Leu¹-zervamicin molecule. The difference isodensity surface was obtained from $P_F - P_K$. Small fuzzy regions represent the remaining difference. They involve a small molecular volume and are evidently small in magnitude.

two clusters of water molecules and a cluster of a water and an ethanol molecule. In this application, the neighborhoods were formed with atoms within 5 Å of the kernels plus some few additions to assure that all electrons were paired and the number of electron pairs was even. Table 1.3 lists the kernels and their neighborhoods. The numbers refer to the peptide residues in the sequence Ac-Leu-Ile-Gln-Iva-Ile-Thr-Aib-Leu-Aib-Hyp-Gln-Aib-Hyp-Aib-Pro-Phol (Aib: α -aminoisobutyric acid; Iva: isovaline; Hyp: 4-hydroxyproline; Phol: phenylalanimol) description of the chemical content of the Leu¹-zervamicin molecule. The symbols in Table 1.3 correspond to those found in the crystal structure analysis [30]. The crystal structure analysis provided the atomic coordinates used in the calculations reported and also afforded the information from which the selection of the associated solvent molecules was based. The last four columns of Table 1.3 show the number of atoms and the number of basis functions for each kernel and for each neighborhood. Each row of Table 1.3 can be considered to symbolize one individual kernel-neighborhood-fragment calculation. All the calculations of all the rows can be run in parallel on modern supercomputers. The natural parallelization of the calculations is one of the computational advantages of the KEM.

With atomic positions held fixed, the electron density distribution in a fragment is computed. Such a calculation delivers contributions to an $\mathbf{R}$ matrix and an $\mathbf{S}$ matrix from which the portion that concerns the kernel is saved in the form of $\mathbf{P}_k = \mathbf{S}_k^{1/2} \mathbf{R}_k \mathbf{S}_k^{1/2}$ where the subscript k refers to a kernel matrix. The elements that are saved in a kernel projector matrix are described as follows. Those contributions to the $\mathbf{P}$ matrix involving orbitals from a neighborhood atom and an atom in the kernel are saved at the fractional value of one-half. If all neighborhood atoms occur only once as part of a kernel, another one-half value would be added to those contributions to the $\mathbf{P}$ matrix already saved at one-half values, when the values associated with the adjoining kernels are calculated. Contributions from pairs of atoms, both in the same kernel, are saved with a coefficient of 1.

In our previous example (the cyclic hexapeptide trihydrate) we saved the $\mathbf{R}_k(0)$ instead of the $\mathbf{P}_k(0)$ and obtained an $\mathbf{R}(0)$ matrix for the full molecule by combining all the $\mathbf{R}_k(0)$ for the various kernels. The $\mathbf{P}_a(0)$ matrix was then obtained by multiplying the $\mathbf{R}(0)$ matrix by $\mathbf{S}(0)$ according to Equation 1.22. The $\mathbf{P}_k$ are saved here instead and in symmetric form. The $\mathbf{P}_k$ are very good kernel representations and lead to a full $\mathbf{P}_s$ matrix that is a very good projector, an improvement on $\mathbf{P}_a(0)$. For the hexapeptide in Section 1.3.4.1, it was possible to obtain good results by defining the single adjacent peptide residue on both sides of a kernel residue as a neighborhood. As a consequence of the denser packing of residues in Leu¹-zervamicin, more residues were required to form the neighborhoods of each kernel.

The matrix $\mathbf{S}$, defined in Equation 1.20, is a matrix representing the overlap integrals of pairs of orbitals. For pairs of orbitals belonging to atoms that are separated by large distances, the values of the overlap integrals will be close to zero. This behavior of $\mathbf{S}$ is the reason why the fragment calculations can give accurate values for the elements of $\mathbf{P}$ for the molecule as a whole.

Table 1.3 Composition of the 19 fragments, that is, kernels and their corresponding neighborhoods, used in the calculation of the **P**-matrix for the hydrated hexadecapeptide, Leu¹-zervamicin.^{a)}

Kernel	Neighborhood	Kernel		Neighbors	
		No. of atoms	No. of basis functions	No. of atoms	No. of basis functions
1 (Ac-Leu)	2,3,4,5, H6a	27	69	99	275
2 (Ile)	1,3,4,5,6, H7a, Wb3	19	51	124	348
3 (Gln)	1,2,4,5,6,7, N8, H8a Wb2, Wa3, W4	17	53	147	403
4 (Iva)	1,2,3,5,6,7,8, H9a, Wb2	16	44	148	412
5 (Ile)	1,2,3,4,6,7,8,9, H10f	19	51	159	447
6 (Thr)	O1,2,3,4,5,7,8,9,10, Wb2, Wa3, Wb3	14	42	158	446
7 (Aib)	O2,3,4,5,6,8,9,10, H11e, H11 h, Wb2, Wa3	13	37	153	441
9 (Leu)	O3,4,5,7,9,10,11,12	19	51	143	411
9 (Aib)	O4,5,6,7,8,10,11,12,13	13	37	164	465
10 (Hyp)	O5,6,7,8,9,11,12,13,14, Wa1, Wa2	15	47	142	414
11 (Gln)	O7,8,9,10,12,13,14,15, Wa1, Wa2	17	53	164	480
12 (Aib)	8,9,10,11,13,14,15, H16a, H16c, EtOH, Wa1	13	37	141	401
13 (Hyp)	9,10,11,12,14,15,16, EtOH, Wa1, Wa2, W8	15	47	142	410
14 (Aib)	10,11,12,13,15,16, Wa1, Wa2	13	37	117	345
15 (Pro)	11,12,13,14,16	14	42	143	411
16 (Phol)	O12,13,14,15, EtOH, W8	23	67	112	311
EtOH, W8	12,13,16	12	28	116	324
Wa1, Wa2	10,11,12,13,14	6	14	85	253
Wb2, Wa3, Wb3, W4	2,3,4,6,7	12	28	145	411

a) The individual numbers in the first column, associated with the 16 sequential peptide residues, imply the same corresponding residues in column 2. Other numbers in column 2 have letters with them, for example, H for hydrogen, O for oxygen, N for nitrogen, and W for water. EtOH symbolizes ethanol. The structural aspects of these symbols are to be found in Reference [30].

The Hartree–Fock calculations for the entire hydrated Leu¹-zervamicin molecule and for the separate fragment calculations were made by use of the Gaussian 94 program [31] employing the STO-3G basis.

Comparison of Electron Densities Isodensity surfaces have been calculated by use of $\phi_i = \sum_{j=1}^{835} C_{ij}\psi_j$ and $\rho = 2 \sum_{i=1}^{524} \phi_i\phi$ from the Hartree–Fock orbitals for the hydrated Leu¹-zervamicin molecule. Use was made of the $\mathbf{P}$ matrix obtained at once for the full molecule and that obtained from the fragment calculations. The two sources gave electron densities that appeared to be quite similar. Therefore, a portion of the one obtained from the full molecule calculation, as a representative of both types of calculation, is illustrated in Figure 1.7a, at an isodensity surface of $0.005 \text{ e } \text{\AA}^{-3}$. Confinement of the computed volume to a region of interest, as illustrated, saves time and memory when desirable or necessary. A ball-and-stick model of the structure is superimposed. To obtain a more quantitative insight into the similarity of both types of density calculation, a series of difference isodensity surfaces were calculated in which differences that did not exceed increasingly larger values were omitted. Evidently, this is a calculation that can determine and locate the largest differences between the electron densities.

The difference isodensity surfaces shown in Figure 1.7b and c were obtained from $\mathbf{P}_F - \mathbf{P}_K$, where the subscripts imply full molecule (F) and sum over kernel (K) matrices. A difference isodensity surface is shown at $1.0 \times 10^{-3} \text{ e } \text{\AA}^{-3}$ in Figure 1.7b, and $-1.2 \times 10^{-3} \text{ e } \text{\AA}^{-3}$ in Figure 1.7c. Some small fuzzy regions are visible at which there are differences as large as, or larger than, the values of the difference isodensities shown. The fuzzy regions should all disappear at slightly larger difference isodensities. Evidently, the fuzzy regions in Figure 1.7b and c are quite small and highly localized, indicating that the electron density is well represented by the $\mathbf{P}$ matrix obtained from the fragment calculations.

Comments Regarding Kernels and Quantum Crystallography We have presented the basic ideas of quantum crystallography. This entails the treatment of the X-ray scattering experiment in a manner consistent with the requirements of quantum mechanics. In particular, the electron density must be N -representable, that is, obtainable from an antisymmetric wavefunction. We indicate how the projector matrix is ensured to be single-determinant N -representable by imposition of the condition that it be a hermitian, normalized projector. By adopting the approximation that a full molecule can be “broken” into smaller “fragments,” consisting of a “kernel” of atoms and its “neighborhood” of atoms, a simplified representation is obtained that reduces the number of parameters required. The kernels are each “extracted” from their fragments by rules patterned upon those of Mulliken population analysis. An approximate matrix for the full molecule is reconstructed by summing over the kernel matrices and imposing the projection property.

The virtue of introducing the concept of kernel matrices is that their use could allow very large molecules to be studied within the context of quantum crystallography. The fundamental feature that explains the applicability of the kernel approximation, as it is applied here, is the vanishing of orbital overlap as the distance

between orbital centers increases. This has the consequence that elements of the matrix $\mathbf{R}$ that weight the relative importance of such vanishing overlap contributions to the density may be neglected without affecting the density. Thus, a pattern of zeros is introduced into the matrix $\mathbf{R}(0)$, which defines the size of fragments that, in general, will be smaller than the full molecule. The fragments of reasonable size contain the essential information for determining the matrices for kernels. Reconstruction of the full matrix in an approximation that can deliver a good density follows and the projection property maintains the structure of quantum mechanics. The formalism is flexible enough that all the electronic and atomic structural variables may be refined by least-squares methods.

The hexapeptide molecule of this chapter was treated within the context of the *ab initio* Hartree–Fock approximation. However, we point out that the concept of extracting kernel matrices from fragments smaller than a full molecule would be applicable within the context of any method based upon a molecular orbital representation, including extended Hückel, empirical Hartree–Fock, configuration interaction and density functional methods. Our initial exploration of other MO methods bears this out.

1.4

Kernel Density Matrices Led to Kernel Energies

Although our initial interest in the kernel neighborhood fragment approximation of the density matrix concerned its applications within quantum crystallography, we also indicate that it should be useful in the purely quantum mechanical problem of solving the Schrödinger equation. These concepts led us to calculations of kernel energies. Following that the kernel energy method (KEM) evolved, which we now discuss.

Given that the problem of large molecule interactions would be interesting to study by use of the techniques of quantum mechanics, the problem they present is often the considerable size of targets composed of, for example, proteins, DNA, RNA, and so on. That problem is addressed here by using the KEM approximation, whose main features are now reviewed.

In the KEM, the results of X-ray crystallographic coordinates are combined with those of quantum mechanics. This leads to a reduction of computational effort and an extraction of quantum information from the crystallography. Central to the KEM is the concept of the kernel. These are the quantum pieces into which the full molecule is mathematically broken. All quantum calculations are carried out on kernels and double kernels. Because the kernels are chosen to be smaller than a full biological molecule, the calculations are accomplished efficiently, and the computational time is much reduced. Subsequently, the properties of the full molecule are reconstructed from those of the kernels and double kernels. Thus a quantum realization of the aphorism that the whole is the sum of its parts is obtained.

It is assumed that the crystal structure is known for a molecule under study. With known atomic coordinates, the molecule is mathematically broken into tractable

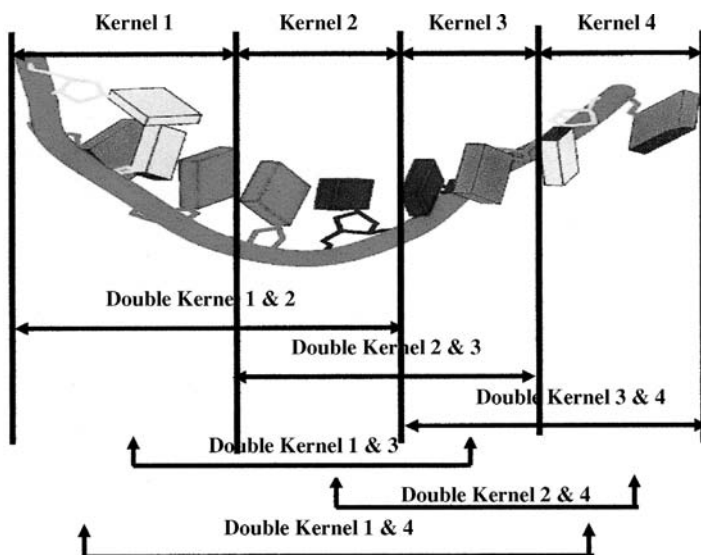


Figure 1.8 Abstract sketch of RNA showing the definitions of the single and double kernels.

pieces called kernels. The kernels are chosen such that each atom occurs in only one kernel. Figure 1.8 shows schematically defined kernels and double kernels, and only these objects are used for all quantum calculations.

The total molecular energy is then reconstructed by summation over the contributions of the double-kernels reduced by those of any single kernels that have been over counted. Two approximations have been found to be useful. In the simpler case, only the chemically bonded double kernels are considered, and the total energy E in this approximation is:

$$E_{\text{total}} = \sum_{i=1, j=i+1}^{n-1} E_{ij} - \sum_{i=2}^{n-1} E_i \quad (1.44)$$

E_{ij} = energy of a chemically bonded double kernel of name ij

E_i = energy of a single kernel of name i

i, j = running indices

n = number of kernels.

In the more accurate case, all double kernels are included, and the total energy is:

$$E_{\text{total}} = \sum_{m=1}^{n-1} \left(\sum_{\substack{i=1 \\ j=i+m}}^{n-m} E_{ij} \right) - (n-2) \sum_{i=1}^n E_i \quad (1.45)$$

E_{ij} = energy of a double kernel of name ij

E_i = energy of a single kernel of name i

i, j, m = running indices

n = number of single kernels.

The purpose of the calculations is to obtain kernel contributions to the energy when it is not computationally feasible to treat the entire molecule as a whole. When a structure of interest has known crystallographic coordinates one may easily define kernels, which altogether represent the entire composite molecule. The use of the single kernels and double kernels indicated above is an approximation that is made to obtain a simplification in the quantum calculation. The validity of this approximation, in the case of various peptides, proteins, DNA and RNA structures, is shown in various works discussed below.

1.4.1

KEM Applied to Peptides

Molecules of biological importance have been chosen for the calculation of molecular energy using the concepts of single kernel and double kernel in the KEM [32]. The examples chosen were sufficiently large to provide significant demonstrations of *ab initio* energy calculations using the kernel energy method, but not so large as to prevent energy calculations of whole molecules using supercomputers. The latter cases were required to provide a standard of excellence against which the approximations using kernels could be judged.

The group of peptides that were selected is shown with their crystal structure geometry in Figure 1.9. Peptides, of course, are of vast biological importance, having the capacity to control many crucial functions of an organism, including cell reproduction, immune response, appetite, and so on. The human organism makes a great many peptides that act as neurotransmitters, hormones and antibiotics. Synthetic peptides are studied as possibly effective drugs. The fundamental biological activity of peptides depends upon their conformation, which is, in turn, determined by the energy of the conformation. Thus, the ability to calculate the molecular energy associated with peptide structure is basic to the study of peptides and their function. In this section, we show how the concept of kernels allows for accurate calculation of peptide energy. All of their crystal structures are known [30, 33–42] and have been used in the energy calculations presented here. Figure 1.9 illustrates various natural and synthetic peptides that vary in size, shape and function.

Table 1.4 shows the energies obtained with Equation 1.44 for 16 different peptides. For one of these, Leu¹-zervamicin [30], we calculated the energy for two different conformations, labeled closed and open. The number of atoms and amino acids in the table range from a minimum of 80 atoms contained in six amino acids to a maximum of 327 atoms contained in a 19 amino acid chain. All energy calculations correspond to the Hartree–Fock approximation using a minimal STO-3G basis, and the effects of solvent were not considered.

The results of Table 1.4 all correspond to a kernel size defined as one amino acid. The KEM requires much less calculation time than would be the case for the full molecule Hartree–Fock calculation in the same basis set without approximation.

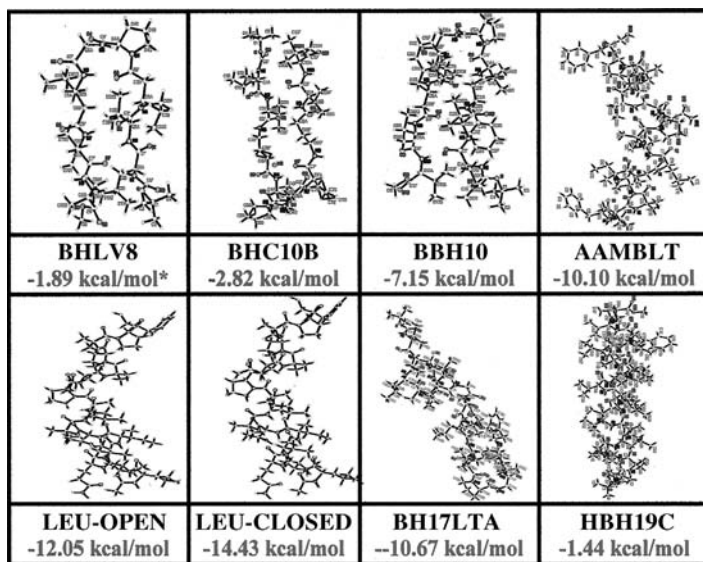


Figure 1.9 Peptide structures from X-ray crystallography. *The energy differences $E_{\text{HF}} - E_{\text{KEM}}$ are from Equation 1.45, which includes all the double kernels.

A distinction has been made between two approximations. In the first, with the use of Equation 1.44, energy contributions are considered only from those double kernels composed of chemically bonded pairs of single kernels, as in the results of Table 1.4. In the second, using Equation 1.45, energy contributions are considered from all double kernels, whether or not they are composed of chemically bonded single kernels. Our anticipation was that including all double kernels would increase the accuracy of the KEM results. Also, with the use of Equation 1.45, if the size of the kernels were increased, it was presumed that it would also increase the accuracy of the KEM approximation. In most cases the use of the Equation 1.44 approximation, as seen in Table 1.4, is fairly accurate. The worst case occurs for HBH19C [34] in 20 kernels for which the difference is $223 \text{ kcal mol}^{-1}$ ($1 \text{ kcal} = 4.184 \text{ kJ}$) out of a total exact value of 6748 au ($4\,234\,370 \text{ kcal mol}^{-1}$) representing about a 0.0053% difference (au stands for “atomic units”). Apparently the approximation based upon the kernels of small size (one amino acid), and including only the chemically bonded double kernels, is a reasonable one.

If the approximation including all double kernels is applied, an increased accuracy is obtained [32]. Also, as is physically reasonable, as the kernel size increases and all double kernels are considered in the calculation, the errors should decrease, as does occur. Thus, judged by the results of peptides represented in Table 1.5, the energy approximations of Equations 1.44 and 1.45 have good accuracy.

The computational results indicate that the kernel energy method is worthwhile. It has yielded results that have small differences. Sixteen calculations have been tabulated for various peptides that have a range of geometries from 4 to 19 residues

Table 1.4 Energy calculation for peptides^{a)}, using Equation 1.44.

Energy	BMA4	ISARAM	ISARIAx	ALAC7ALT	ADPGV7b	BHF4LT	BDPGV7A	BH2L2
Atoms (kernels)	80 (6)	104 (6)	107 (6)	125 (7)	126 (7)	134 (7)	142 (8)	144 (9)
E_{HF} (au)	-1781.44	-2274.28	-2312.95	-2522.80	-2539.50	-3006.70	-2805.72	-2970.77
E_{KEM} (au) ^{b)}	-1781.43	-2274.27	-2312.94	-2522.79	-2539.50	-3006.69	-2805.71	-2970.73
$E_{\text{HF}} - E_{\text{KEM}}$ (kcal mol ⁻¹)	-1.82	-8.16	-6.90	-5.02	-3.39	-8.28	-5.02	-21.46
Energy	BHLV8	BHC10B	BBH10	AAMBLT	Leu-open	Leu-closed	BH17LTA	HBH19C
Atoms (Kernels)	150 (9)	164 (11)	190 (11)	246 (16)	265 (16)	265 (16)	269 (17)	327 (20)
EHF (a.u.)	-3047.10	-3528.00	-3986.31	-5529.32	-5849.50	-5851.57	-5800.36	-6748.41
* E_{KEM} (a.u.)	-3047.07	-3527.96	-3986.28	-5529.26	-5849.44	-5851.50	-5800.32	-6748.05
$E_{\text{HF}} - E_{\text{KEM}}$ (kcal mol ⁻¹)	-15.81	-21.90	-18.39	-33.89	-39.28	-43.55	-24.85	-222.64

a) KEM applied to peptides using Equation 1.44, with 1 kernel $\equiv$ 1 amino acid. (Calculations were performed without solvent at the HF/STO-3G level of theory.)

b) Including only double kernels composed of single kernel pairs chemically bonded to one another, Equation 1.44.

and atoms numbering from 80 to 327. The total energy range is 1781–6748 au. The differences for the energies are quite small as a percentage of the total energy. The total energy was calculated by summing the energies of double kernels. In so doing, the contribution of some single kernels are counted twice and thus the contribution of over-counted single kernels must be subtracted from the total.

The basic assumption is that the energy of any given kernel is most affected by its own atoms and those of the neighboring kernels with which it interacts. A pair of interacting kernels forms a double kernel. Perhaps the most important double kernels are those formed of chemically bonded single kernels. Thus kernels and double kernels are used to define the energy of the full molecule as in Equations 1.44 and 1.45. As a molecule grows in size there are more double kernels and more single kernels, but the basic formula is the same. The total energy is a sum of contributions of double kernels reduced by single kernels that have been over-counted. Tables 1.4 and 1.5 show the energy is well represented by the above kernel energy method. The fragment calculations are carried out on double kernels and single kernels whose ruptured bonds have been mended by attachment of H atoms. A satisfactory occurrence in the summation of energies is that the total contribution of hydrogen atoms introduced to saturate the broken bonds tends to zero. The effect on the energy of the hydrogen atoms added to the double kernels effectively cancels that of the hydrogen atoms added to the pure single kernels that enter with opposite sign.

There are, of course, limitations to the accuracy of the KEM. The basic assumption is that the total energy can be built up so long as the atoms of one kernel are mainly affected by themselves and those of neighboring kernels. The tabulated calculations show that the most important double kernels are those composed of pairs of single kernels that are chemically bonded to one another. For best accuracy, however, all double kernels are calculated.

The effect of kernel size on the accuracy of the energy has been considered. In our calculations, increasing kernel size improves the accuracy of energy results. Based upon the peptides calculated thus far we conclude that increasing kernel size reduces the already small difference that occurs when the size of a kernel is specified to be the size of one amino acid. Including all double kernels gives the smallest difference.

The times for entire molecular calculations have been compared to those based upon Equation 1.44. In Figure 1.10 the full molecule Hartree–Fock case has been fit to a fourth power polynomial, and the results based upon Equation 1.44 have been fit to a linear expression. Clearly, the approximation of Equation 1.44 saves computing time. When the two curves are extrapolated beyond the computational data points represented by Table 1.6, the discrepancy between fourth and first power grows. The main diagram in Figure 1.10 plots the projected times shown in Table 1.7. With 1000 atoms, the computing time for an entire molecule is about 13 hours, and the computing time for the KEM is about 18 minutes. At 10 000 atoms the computing time for an entire molecule is about 145 days, and the computing time for the KEM is about 3.5 hours. The use of the KEM with Equation 1.44 applied to peptides gives good accuracy at a significant saving of computing time. This augers well for application of the same method to even larger molecules.

Table 1.5 Energy calculation for peptides, using Equations 1.44 and 1.45.^{a)}

Energy	BMA4	ISARAM	ISARIAX	ALAC7ALT	ADPGV7b	BHF4LT	BDPGV7A	BH2L2
Atoms (kernels)	80 (4)	104 (3)	107 (3)	125 (4)	126 (4)	134 (4)	142 (4)	144 (4)
E_{HF} (au)	-1781.44	-2274.28	-2312.95	-2522.80	-2539.50	-3006.70	-2805.72	-2970.76
E_{KEM} (au) ^{b)}	-1781.43	-2274.27	-2312.94	-2522.79	-2539.49	-3006.69	-2805.71	-2970.75
Difference (kcal mol ⁻¹)	-4.83	-3.51	-1.26	-5.08	-7.84	-4.96	-7.03	-7.47
E_{KEM} (au) ^{c)}	-1781.43	-2274.27	-2312.94	-2522.79	-2539.50	-3006.70	-2805.71	-2970.76
$E_{\text{HF}} - E_{\text{KEM}}$ (kcal mol ⁻¹)	-2.82	-3.51	-1.26	-3.14	-3.20	-2.70	-3.51	-3.20
Energy	BHLV8	BHC10B	BBH10	AAMBLT	Leu-open	Leu-closed	BH17LTA	HBH19C
Atoms (kernels)	150 (4)	164 (5)	190 (6)	246 (6)	265 (7)	265 (7)	269 (7)	327 (3s)
E_{HF} (au)	-3047.10	-3528.00	-3986.31	-5529.32	-5849.50	-5851.57	-5800.36	-6748.41
E_{KEM} (au) ^{b)}	-3047.08	-3527.98	-3986.30	-5529.29	-5849.46	-5851.52	-5800.34	-6748.41
Difference (kcal mol ⁻¹)	-10.42	-10.35	-12.11	-16.50	-28.30	-29.74	-14.37	-2.63
E_{KEM} (au) ^{c)}	-3047.09	-3527.99	-3986.30	-5529.30	-5849.48	-5851.55	-5800.34	-6748.41
$E_{\text{HF}} - E_{\text{KEM}}$ (kcal mol ⁻¹)	-1.89	-2.82	-7.15	-10.10	-12.05	-14.43	-10.67	-1.44

KEM applied to peptides using Equations 1.44 and 1.45, and with kernel sizes larger than one amino acid. (Calculations were performed without solvent at the HF/STO-3G level of theory.)

The only double kernels included are those made of single kernel pairs that are chemically bonded to one another, Equation 1.44.

All double kernels are included, Equation 1.45.

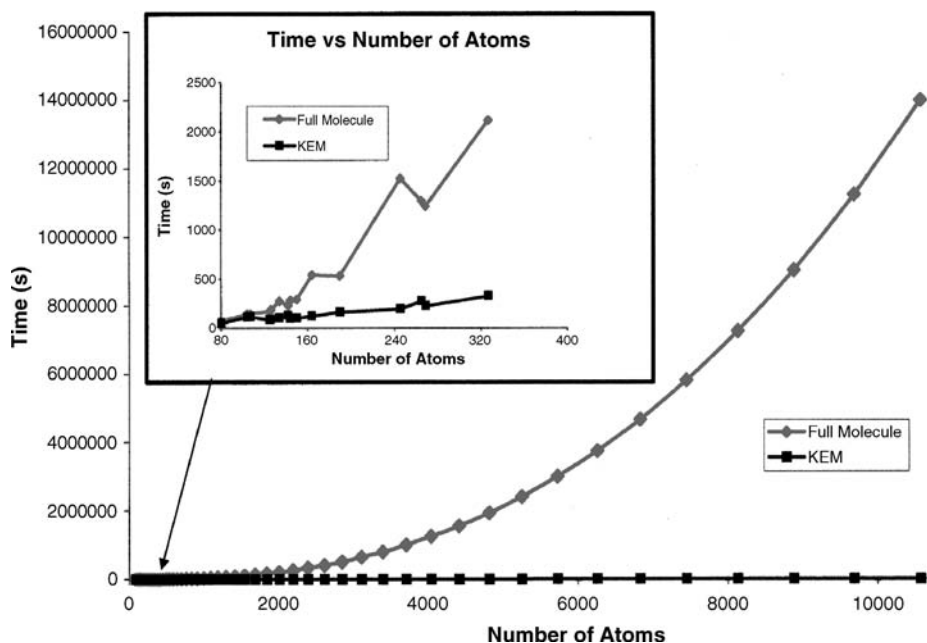


Figure 1.10 Calculation time comparison of full molecule versus KEM. Inset: actual calculation time data for the molecules of Table 1.6. Main figure: projected times obtained from a fourth-order polynomial fit to the HF calculation times for full molecules, and a linear function fit to the KEM calculation times for the same molecules (Table 1.7).

With the use of the known structures of peptides, from crystal structure analysis, it has been shown that it is feasible to make *ab initio* quantum mechanical calculations to good approximation for very large molecules, employing the notion that the whole may be obtained from its parts. In our procedure the parts are the quantum mechanical kernels. The key to such computations is the fragment calculation wherein a molecule is divided into kernels and *ab initio* calculations are performed on each of the kernel fragments and double kernel fragments. The results of our calculations suggest that the larger the kernels the greater the relative accuracy.

1.4.2

Quantum Models within KEM

A model chemistry specifies a quantum method of calculation and a set of basis functions. Given the computational advantages alluded to above, the question arises: What is the effect of the choice of basis functions and quantum methods on the KEM approximation [43] All the previous calculations used to test the approximation were, in the first instance, for reasons of simplicity, based on the use of STO-3G basis functions and HF calculations. It is therefore reasonable to wonder whether the approximation will work equally well with another choice of model chemistry.

Table 1.6 Calculation time (in seconds) for peptides.^{a)}

Time (s)	BMA4	ISARAM	ISARIAX	ALAC7ALT	ADPGV7b	BHF4LT	BDPGV7A	BH2L2
Atoms	80 (6)	104 (6)	107 (6)	125 (7)	126 (7)	134 (7)	142 (8)	144 (9)
(kernels)								
$t_{\text{full-molecule}}$	67	143	148	171	193	276	236	284
$t_{\text{KEM}}^{\text{b)}}$	51	112	118	85	91	106	127	96
Time (s)	BHLV8	BHC10B	BBH10	AAMBLT	Leu-open	Leu-closed	BH17LTA	HBH19C
Atoms	150 (9)	164 (11)	190 (11)	246 (16)	265 (16)	265 (16)	269 (17)	327 (20)
(kernels)								
$t_{\text{full-molecule}}$	296	546	534	1529	1300 ^{c)}	1300 ^{c)}	1241	2122
$t_{\text{KEM}}^{\text{b)}}$	102	126	165	196	274 ^{c)}	274 ^{c)}	226	327

- a) The same supercomputer and the same number of parallel nodes were employed for all calculation times shown here. All energy calculations are in the approximation HF/STO-3G.
b) Only chemically bonded double kernels are included, Equation 1.44; 1 kernel $\equiv$ 1 amino acid.
c) Average time of Leu-open and Leu-closed calculations.

Table 1.7 Comparison between the estimated calculation times for the full molecule and for the KEM.^{a)}

No. atoms	Full molecule in hours	KEM in hours
86	0.022	0.019
94	0.028	0.021
103	0.035	0.023
112	0.043	0.025
122	0.054	0.027
133	0.067	0.030
145	0.083	0.033
159	0.104	0.036
173	0.129	0.040
189	0.161	0.044
206	0.201	0.048
225	0.250	0.053
246	0.311	0.059
268	0.387	0.064
293	0.482	0.071
320	0.601	0.078
349	0.748	0.086
381	0.932	0.094
416	1.160	0.104
454	1.445	0.114
495	1.799	0.126
540	2.240	0.138
590	2.790	0.152
644	3.474	0.167
702	4.327	0.184
767	5.388	0.202

Table 1.7 (Continued)

No. atoms	Full molecule in hours	KEM in hours
837	6.709	0.222
913	8.355	0.245
997	10.405	0.269
1088	12.957	0.296
1187	16.135	0.326
1296	20.093	0.358
1414	25.022	0.394
1544	31.160	0.433
1685	38.803	0.477
1839	48.322	0.524
2007	60.175	0.577
2190	74.935	0.634
2390	93.316	0.698
2609	116.206	0.767
2847	144.711	0.844
3107	180.208	0.929
3392	224.413	1.022
3702	279.460	1.124
4040	348.010	1.236
4409	433.376	1.360
4812	539.681	1.496
5252	672.062	1.645
5732	836.915	1.810
6256	1042.207	1.991
6828	1297.855	2.190
7452	1616.213	2.409
8133	2012.663	2.649
8877	2506.359	2.914
9688	3121.158	3.206
10 574	3886.763	3.526

- a) Comparison of times obtained by fitting polynomials to the actual computing time data for the molecules of Table 1.6, for the full molecule calculation to a fourth-order polynomial and for KEM to a linear function.

Because the previous investigation examined such a wide variety of different peptides, in terms of size, shape, and structure, all with positive results, it seems unlikely that the KEM would depend sensitively on a particular choice. However, to preclude that possibility, we examine here the effect of the choice of model chemistry on the applicability of KEM. The issue is whether KEM is more or less independent of a choice of model chemistry. This question is pursued within the context of both (i) the various choices of basis set and (ii) the use of different quantum chemical methods of calculation. For (i) tests of KEM sensitivity to basis functions have been carried out by applying the KEM approximation repeatedly to the same molecule, ADPGV7b (Figure 1.11) [42], which contains 126 atoms, using various basis functions. For each basis, the energy of the full molecule has been calculated and is labeled

$E_{\text{full-molecule}}$. The difference between the full molecule result and that obtained by KEM in the same basis has been examined. It is of interest to know whether the difference depends in a sensitive way on the choice of basis functions. For example, does that energy difference change systematically with the size and quality of the basis set employed? Alternatively, do the errors fluctuate within limits, not correlated to the size and quality of the basis functions used for the calculations? These questions of basis set dependence are examined in the numerical experiments discussed in Section 1.4.2.1.

Inasmuch as the first KEM paper [32] was restricted to calculations within a HF model chemistry, a question arises as to whether applications of KEM will prove to be valid across a whole spectrum of commonly used quantum methods, characterized by differing levels of accuracy. This is answered by choosing a particular peptide as a test case, namely, Zaib4 (which contains 74 atoms), and by calculating its energy with several different quantum chemical methods. These include HF and DFT calculations, but range widely from there. In the direction of more approximate calculations, semiempirical models are used. In the opposite direction of accuracy, to the same test molecule, several higher-level quantum mechanical chemistry models are applied. It is found that KEM is widely applicable across the spectrum of models tested. Thus, in the Zaib4 study, the above formulas were applied in calculating the molecular energy, to test the accuracy of KEM for various basis functions, as well as chemistry models characterized by different levels of accuracy.

1.4.2.1 Calculations and Results Using Different Basis Functions for the ADPGV7b Molecule

It may be shown that the accuracy of KEM does not depend on a particular choice of basis functions. This is done by calculating the ground-state energy of a representative peptide, ADPGV7b, containing seven amino acid residues, using seven different commonly employed basis function sets, ranging in size from small to medium to large. The study of sensitivity of the KEM approximation to choice of basis functions employed the following basis sets: STO-3G [44, 45], 3-21G [46–51], SV [52, 53], 6-31G [54–63], D95 [64], 6-31G* [65, 66] and cc-pVDZ [67–71]. The accuracy of the KEM does not vary in any systematic way with the size or mathematical completeness of the basis set used, and good accuracy is maintained over the entire variety of basis sets tested. We conclude that the accuracy inherent in the KEM is not dependent on a particular choice of basis functions. The first application, to different peptides mentioned above, employed only HF calculations.

The peptide ADPGV7b of known crystal structure [42] is pictured in Figure 1.11 and broken into four single kernels. The amino acid sequence defining the peptide is as follows: Ac-Val-Ala-Leu-Dpg-Val-Ala-Leu-OMe (Dpg = α , α - di-*n*-propyl glycine). Equations 1.44 and 1.45 were applied repeatedly to the calculation of the energy of the peptide ADPGV7b using each of seven different sets of basis functions. This was done in both the HF approximation and the density functional theory (DFT) approximation, using the standard potential B3LYP. The purpose in both cases was to assess whether the accuracy of the KEM was critically dependent on the choice of basis functions.

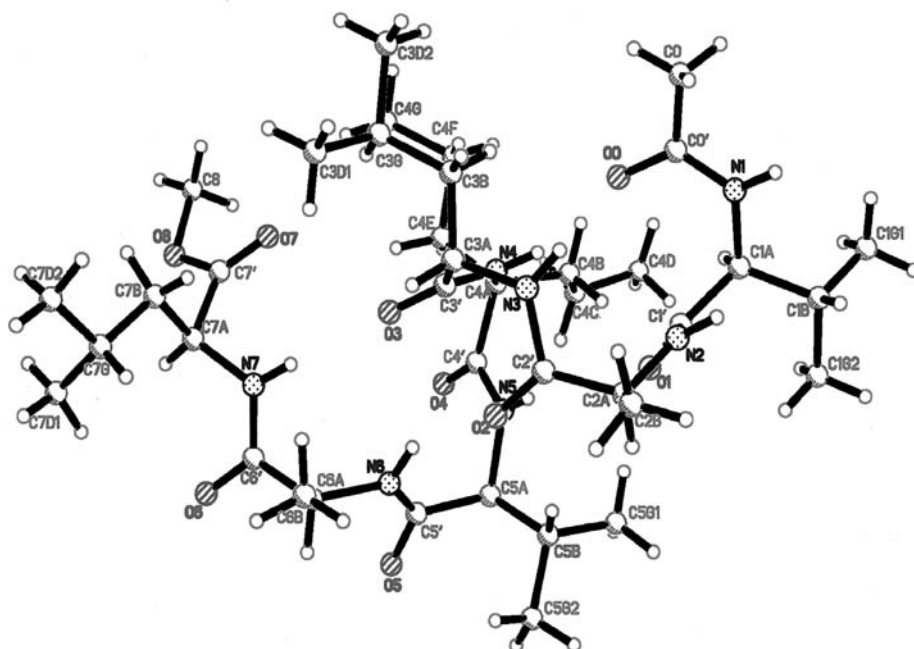


Figure 1.11 ADPGV7b X-ray crystal structure.

Table 1.8 presents the energies obtained with Equation 1.44 for the seven different basis sets. The HF results are in Table 1.8, while the DFT results, which are qualitatively similar, are not shown. The effects of solvent were not considered in this study. All results shown in Table 1.8 correspond to the peptide ADPGV7b, composed of a total of 126 atoms, broken into four kernels. The results for the full molecule calculations are labeled as $E_{\text{full-molecule}}$. A main conclusion that can be drawn from Table 1.8 is that the energy obtained from the KEM is quite accurate for all the basis sets used, and moreover that the accuracy does not correlate in any obvious way with the choice of basis. It may be seen, for example, that the energy differences associated with application of KEM do not correlate with the increasing

Table 1.8 KEM calculation for ADPGV7b, using different basis functions (126 atoms, 4 kernels).

HF/basis	$E_{\text{full-molecule}}$ (au)	E_{KEM} (Equation 1.44) (au)	E_{diff} (kcal mol ⁻¹)
STO-3G	-2539.5022	-2539.4971	-3.20
3-21G	-2557.8857	-2557.8809	-3.01
SV	-2568.7546	-2568.7475	-4.46
6-31G	-2570.9939	-2570.9872	-4.20
D95	-2571.3472	-2571.3383	-5.58
6-31G*	-2572.1191	-2572.1125	-4.19
cc-pVDZ	-2572.2345	-2572.2285	-3.78

mathematical completeness of the basis used for the energy calculation. The same is equally true for the DFT results and the HF results.

Now a distinction is made between two approximations. In the first, with the use of Equation 1.44, energy contributions are considered only from those double kernels composed of chemically bonded pairs of single kernels. In the second, using Equation 1.45, energy contributions are considered from all double kernels, whether or not they are composed of chemically bonded single kernels. As expected, results for the peptide ADPGV7b indicate a general trend in which accuracy is increased when all double kernels are included in the calculation as specified by Equation 1.45. The result is that the already small differences associated with Equation 1.44 are even smaller with the use of Equation 1.45. It is physically reasonable that when all double kernels are considered in the calculation the difference should decrease, as occurs in the tables. However, and this is a main point of interest, the differences associated with the results of Equation 1.45, just as with Equation 1.44, are relatively small and fluctuate rather randomly with the choice of basis set employed in the calculations. This occurs in both the HF and DFT approximations.

1.4.2.2 Calculations and Results Using Different Quantum Methods for the Zaib4 Molecule

The second question (ii) that arises is whether the results obtained with the use of KEM will be accurate only within the HF approximation. Therefore, we also studied whether KEM is applicable across various quantum computational methods, characterized by differing levels of accuracy. The peptide, Zaib4, containing 74 atoms, was used to calculate its energy at seven different levels of accuracy. These include the semi-empirical methods, AM1 and PM5, a DFT B3LYP model, and *ab initio* HF, MP2, CID and CCSD calculations. KEM was found to be widely applicable across the spectrum of quantum methods tested.

The calculations below, which test the sensitivity of the KEM approximation to choice of model accuracy, employ seven different quantum methods as follows: AM1 [72], PM5 [73], HF [74], DFT [75], CID [76], MP2 [77] and CCSD [78]. For this study we have adopted as a test molecule a 74-atom peptide called Zaib4. Figure 1.12 shows a picture of the molecule arising from the X-ray crystal structure. The amino acid sequence defining the Zaib4 peptide is as follows: Z-Aib-Aib-Aib-Aib-OMe. Table 1.9 gives the calculated molecular energy results for the chemistry models tested. All calculations correspond to the crystal structure geometry. The same STO-3G basis functions were used for all *ab initio* quantum mechanical methods listed. $E_{\text{full-molecule}}$ is listed for each chemistry model along the table, and represents the calculated energy of the full molecule taken as a whole without being broken into kernels.

This is the standard of excellence against which KEM results are to be judged. Table 1.9 lists the calculated energies that derive from KEM using the approximations given by Equation 1.45. Also given are the corresponding differences between $E_{\text{full-molecule}}$ and the values calculated with Equation 1.45. Equations 1.44 (not shown) and 1.45 were applied repeatedly to the calculation of the energy of the peptide Zaib4 using each of seven different methods of quantum chemical calculation indicated in

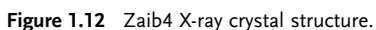


Table 1.9 KEM calculation for Zaib4, using different quantum methods (74 atoms, 3 kernels).

a) Semiempirical methods that consider only the valence electrons.

results of Table 1.9 is that it appears all types of quantum calculations tested, within the limits discussed, are compatible with KEM. The quantum methods displayed in Table 1.9 represent a broad sample of the methodologies commonly used in computational chemistry. Thus, they present a good test of how widely applicable KEM may be for obtaining molecular energies. The numerical values of Table 1.9 indicate that the KEM results are uniformly applicable, for all the model chemistries that have been tested. The errors associated with basing the molecular energy on the approximation related to summing over the kernels, in accordance with Equations 1.44 and 1.45, is generally quite small.

1.4.2.3 Comments Regarding KEM

In judging the accuracy of KEM, the differences of interest are those between the $E_{\text{full-molecule}}$ energy and that predicted by the KEM, both in the same basis set and using the same equations of motion. At least for as the seven basis sets used thus far, it seems that the validity of the KEM approximation does not depend on a particular choice of basis. Therefore, in future applications of KEM, the choice of basis may be made freely, in accordance with those considerations usually apropos of a particular molecular problem, including the absolute accuracy to be achieved, given the computational power, and computational time available, for the task at hand. Turning our attention to the numerical comparisons afforded between the $E_{\text{full-molecule}}$ energies for the various quantum methods and the corresponding energies obtained from KEM approximations, we have seen that they are quite close. It is a favorable result for KEM that it has proved to be applicable with all the quantum methods tested. At least with respect to the limited number of tests that we have been able to carry out, it seems that the validity of KEM will not depend in a sensitive way on either the basis sets or the calculation level of quantum methods used.

1.4.3

KEM Applied to Insulin

1.4.3.1 KEM Calculation Results

An application has been made with the protein insulin [79–81], which is composed of 51 amino acids. Accurate KEM Hartree–Fock energies were obtained for the separate A and B chains of insulin and for their composite structure in the full insulin molecule. A limited basis is used to make possible calculation of the full insulin molecule, which can be used as a standard of accuracy for the KEM calculation.

Insulin is composed of two peptide chains named A and B. The chains are linked by two disulfide bonds, and an additional disulfide is formed within the A chain. The A chain contains 21 amino acids, composed of 309 atoms, including hydrogen, and the B chain contains 30 amino acids, composed of 478 atoms, including hydrogen. Figure 1.13 shows a ribbon diagram of the insulin molecule that gives an impression of the three-dimensional structure of the molecule. The quantum mechanical method chosen for testing the KEM in the case of insulin is that of the Hartree–Fock (HF) equations using atomic orbital basis functions of type STO-3G.

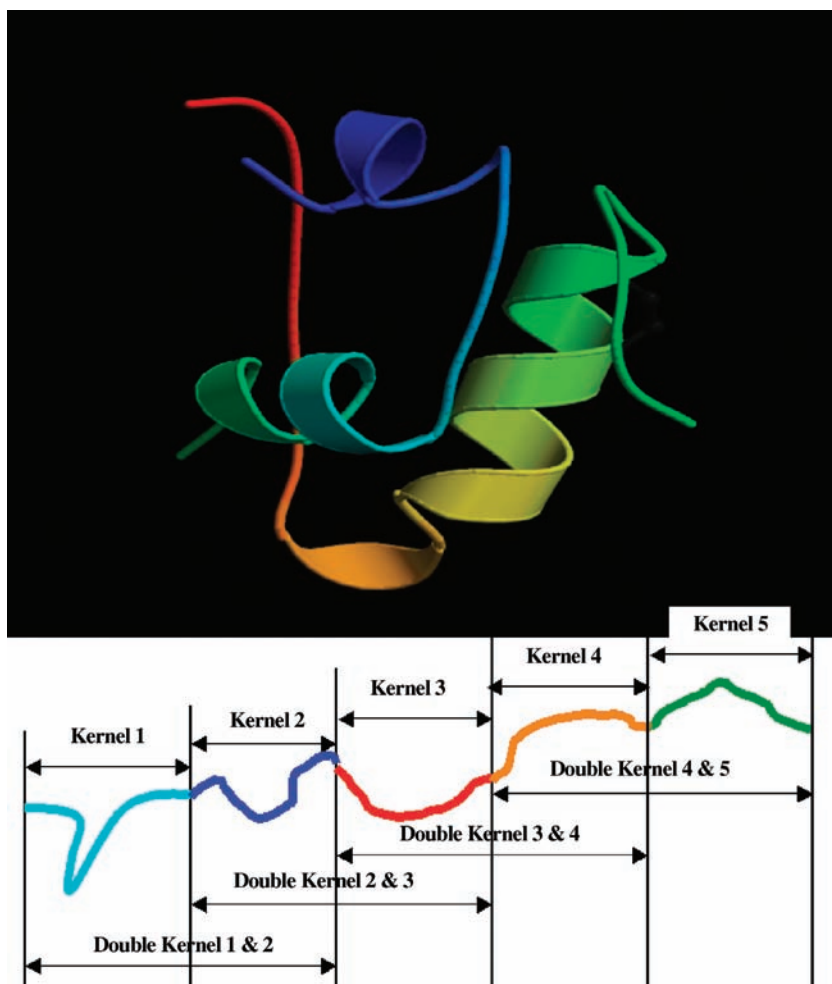


Figure 1.13 The insulin molecule is composed of two chains, A in blue (shown as two shorter helices) and B in green-red (shown as one longer helix). The whole molecule is divided into five kernels as shown. The insulin figure was generated by KING Viewer in the PDB web site.

The full insulin molecule (chains A and B) yields a calculated total energy of $E_{\text{HF}} = -21\,104.7660$ au. The KEM result, $E_{\text{KEM}} = -21\,104.7656$ au (Equation 1.45), differs from this by as little as 0.0004 au. For all three calculations, that is, chain A, chain B and the complete solvated insulin molecule, the energy differences were calculated corresponding to the full molecule result and its approximation by the KEM. The energy differences are relatively small. In all three cases, the Equation 1.45 differences are less than those of Equation 1.44, and are of magnitude 1 kcal mol^{-1} .

Table 1.10 Energy calculation for solvated insulin.

No. of atoms	No. of kernels	E_{HF} (au)	$E_{\text{KEM}}^{\text{a)}}$ (au)	$E_{\text{HF}} - E_{\text{KEM}}$ (kcal mol ⁻¹)
959	6	-26275.4187	-26275.4127	-3.79

a) KEM calculation with HF/STO-3G, using all the double kernels (Equation 1.45).

Table 1.10 considers the case of the full insulin molecule in the presence of solvent molecules. In the crystal, the solvent molecules are present as 56 H₂O and a single 1,2-dichloroethane. The fully solvated insulin contains a total of 959 atoms. All of the atomic positions of the solvent molecules together with those of the full insulin have been determined crystallographically (Protein Data Bank, PDB ID code 1APH), except for hydrogen atoms, which we have added. In the KEM calculations that have included solvent, all atoms of the solvent together have been used to define one additional kernel, over and above the five kernels chosen to represent chain A and chain B of the full insulin molecule. The KEM results are $E_{\text{KEM}} = -26\,275.4013$ au (Equation 1.44) and $-26\,275.4127$ au (Equation 1.45). The results using the KEM are compared with those obtained for the fully solvated molecule, having a total energy of $E_{\text{HF}} = -26\,275.4187$ au. The KEM energies differ from this by 0.0174 au (Equation 1.44) and 0.0060 au (Equation 1.45).

1.4.3.2 Comments Regarding the Insulin Calculations

The electronic structure of protein molecules is still not routinely accessible for study by quantum mechanical methods. Here, it has been shown to be possible using the KEM in the case of the protein insulin. Thus, a quantum mechanical explanation, so useful in application to molecules of moderate size, will prove useful too with protein molecules. Here the KEM, which represents a combination of crystallography and quantum mechanics, while simplifying calculations, has achieved near *ab initio* accuracy in the energy for insulin. This has been demonstrated with the components of insulin called chains A and B, the full insulin molecule, and the fully solvated crystalline insulin molecule. The demonstration was carried out by using the HF approximation in a limited Gaussian basis. The numerical results indicate the validity of the KEM in its application to the various aspects of insulin structure studied in this work.

Table 1.10, which gives the results for the explicit treatment of the solvent molecules that have been crystallized together with insulin, shows that the solvent molecules may be collected into one solvent kernel with results whose accuracy is good. The differences are only of magnitude 10.9428 and 3.7921 kcal mol⁻¹, respectively, using Equations 1.44 and 1.45. The corresponding percentage differences are 0.000 066% and 0.000 023%, respectively. Thus, it is shown here that solvent molecules of crystallization may also be included in the KEM calculations with good accuracy.

The KEM has proven to be applicable to all aspects of the insulin molecule that we have tested [82]. The magnitude of all energy differences obtained between E_{HF} and E_{KEM} are relatively small. Moreover, the energy differences are of the same order of

magnitude as would be expected from the previous work in the case of peptides. We conclude that the KEM calculations are applicable to the energy and electronic structure of proteins.

1.4.4

KEM Applied to DNA

1.4.4.1 KEM Calculation Results

The results for structures from X-ray crystallography and energy differences ($E_{\text{HF}} - E_{\text{KEM}}$ for all of the double kernels) calculated for each of a dozen different DNA systems are displayed in Figure 1.14 [83–93] and Table 1.11. For these DNA systems the number of atoms and the number of kernels involved range from 198

			
BDNA - 251D -0.05 kcal/mol	BDNA - 110D -0.03 kcal/mol	BDNA - 1G6D -0.08 kcal/mol	BDNA - 1IH1 0.75 kcal/mol
			
BDNA - 206D 0.75 kcal/mol	BDNA - 1S9B -0.04 kcal/mol	BDNA - 309D 0.73 kcal/mol	BDNA - 425D 3.54 kcal/mol
			
BDNA - 424D 0.34 kcal/mol	BDNA - 102D -1.20 kcal/mol	ZDNA - 1D48 -7.98 kcal/mol	ADNA - ADH010 1.41 kcal/mol

Figure 1.14 DNA structures from X-ray crystallography; range of molecule size: 197 to 2418 atoms. (DNA diagrams are from the Nucleic Acid Database <http://ndbserver.rutgers.edu/atlas/xray/index.html>.)

Table 1.11 Energy calculation for DNA without solvent.^{a)}

	B-DNA 251D	B-DNA 110D	B-DNA 1G6D	B-DNA 1IH1	B-DNA 206D	B-DNA 1S9B
Atoms (kernels)	198 (3)	197 (3)	330 (5)	394 (6)	395 (6)	466 (7)
E_{HF} (au)	-8127.68	-8143.38	-13614.26	-16287.63	-16270.50	-19082.30
E_{KEM} (au) ^{b)}	-8127.68	-8143.38	-13614.26	-16287.65	-16270.50	-19082.30
$E_{\text{HF}} - E_{\text{KEM}}$ (kcal mol ⁻¹) ^{b)}	0.08	-0.08	0.46	13.02	-5.08	-0.61
E_{KEM} (au) ^{c)}	-8127.68	-8143.38	-13164.26	-16287.63	-16270.50	-19082.30
$E_{\text{HF}} - E_{\text{KEM}}$ (kcal mol ⁻¹) ^{c)}	-0.05	-0.03	-0.08	0.75	0.75	-0.04
	B-DNA 309D	B-DNA 425D	B-DNA 424D^{d)}	B-DNA 102D	Z-DNA 1D48	A-DNA ADH010
Atoms (kernels)	658 (6)	788 (6)	2364 (18)	790 (6)	394 (6)	528 (8)
E_{HF} (au)	-27079.08	-32509.97	-97529.42	-32476.74	-16286.64	-21652.54
E_{KEM} (au) ^{b)}	-27079.09	-32509.96	—	-32476.75	-16286.64	-21652.54
$E_{\text{HF}} - E_{\text{KEM}}$ (kcal mol ⁻¹) ^{b)}	6.85	-5.00	—	1.50	2.07	3.24
E_{KEM} (au) ^{c)}	-27079.08	-32509.98	-97529.42	-32476.75	-16286.63	-21652.54
$E_{\text{HF}} - E_{\text{KEM}}$ (kcal mol ⁻¹) ^{c)}	0.74	3.55	0.34	-1.20	-7.98	1.41

- a) The KEM applied to DNA using Equations 1.44 and 1.45, and with HF/STO-3G.
b) The only double kernels included are those made of single kernel pairs that are chemically bonded to one another.
c) All double kernels are included.
d) 424D has three double helix chains, $E_{\text{HF}} = E_{\text{ab}} + E_{\text{cd}} + E_{\text{ef}}$.

atoms and 3 kernels for the smallest molecule (B-DNA-251D) up to 2364 atoms and 18 kernels for the largest molecule considered (B-DNA-424D). For each DNA molecular system the full molecule Hartree–Fock energy E_{HF} was calculated. This number is the standard against which the accuracy of KEM results is judged. The energies listed as E_{KEM} represent the results obtained by dividing the DNA molecular systems into kernels and then calculating the total energy in the approximations formalized within Equations 1.44 and 1.45 above. The results of Equations 1.44 and 1.45 were calculated separately. Table 1.11 lists for each molecular system the energy differences $E_{\text{HF}} - E_{\text{KEM}}$, for Equation 1.45. (Note that the full molecular energies are usually listed in units au, but the energy differences are listed in the smaller units kcal mol⁻¹.)

The results of Table 1.11 show that the KEM is quite accurate, as one may observe from the energy differences $E_{\text{HF}} - E_{\text{KEM}}$. For Equation 1.44 the absolute magnitude of the energy differences range from a minimum of 0.0795 to a maximum of 13.0105 kcal mol⁻¹. These differences are relatively small, and thus the accuracy of the KEM as implemented in Equation 1.44 is good. The results of Equation 1.45 are even more accurate. For Equation 1.45 the absolute magnitude of the energy differences range from a minimum of 0.0328 to a maximum of 7.9827 kcal mol⁻¹. The Equation 1.45 results are generally expected to be more accurate than the case for Equation 1.44.

1.4.4.2 Comments Regarding the DNA Calculations

The DNA molecular systems of this chapter were treated within the context of the *ab initio* Hartree–Fock approximation. The basis set used for all cases was a limited basis, of Gaussian STO-3G type. A limited basis was chosen to make the energy calculations on full molecular systems (i.e., E_{HF}) as convenient as possible. The numerical values of E_{HF} provided the standard of comparison for the energy values obtained by the KEM. Comparisons between E_{HF} and E_{KEM} have shown that the KEM can be applied to a wide variety of DNA molecular systems with good accuracy. In particular, such calculation accuracy holds true for A-, B- and Z-DNA, the three main types of DNA configuration. The most common configuration of DNA, that is, B-DNA, was examined in ten different molecular systems of variable geometry, and magnitude, as judged by the number of atoms in the system, and was in each case found to be described with good accuracy by the KEM [94].

1.4.5

KEM Applied to tRNA

The quantum mechanical molecular energy of a particular tRNA, of known crystal structure [95], has been calculated with the use of the KEM [96]. The molecule chosen is the yeast initiator tRNA (ytRNA_i^{Met}), designated in the Protein Data Bank as 1YFG and in the Nucleic Acid Database as ID TRNA12 (Figure 1.15).

The structure of this molecule is stabilized by a complicated network of hydrogen bonds that have been identified through crystallography. The numerical results obtained in this work use the Hartree–Fock equations, and a limited basis. Table 1.12



Figure 1.15 Crystal structure of tRNA; 1YFG picture is from the Protein Data Bank (PDB).

lists the results that follow from application of Equations 1.44 and 1.45 to the initiator tRNA molecule 1YFG.

The molecule consists of 2565 atoms, which have been broken into 19 kernels. Thus, the average number of atoms per kernel is about 135, which is of such a size as to be readily calculable, whereas the original number of atoms, 2565, is very much less convenient to treat as a whole. We emphasize that Table 1.12 shows that Equations 1.44 and 1.45 results are quite close. They differ by only -0.0073 au, or -1.79×10^{-3} (kcal mol $^{-1}$ atom $^{-1}$).

Table 1.12 Energy calculation for 1YFG (tRNA) by HF/STO-3G.

No. of atoms	No. of kernels	$E_{\text{KEM}}^{\text{a)}} \text{ (au)}$	$E_{\text{KEM}}^{\text{b)}} \text{ (au)}$	$\Delta E = E_{\text{KEM}}^{\text{b)}} - E_{\text{KEM}}^{\text{a)}} \text{ (au)}$	$\Delta E \text{ per atom (kcal mol}^{-1}\text{)}$
2565	19	-108995.17	-108995.17	-0.0073	-1.79×10^{-3}

a) The double kernels included are only those made of single kernel pairs chemically bonded to one another, and hydrogen bond interaction energies are added to the results of Equation 1.44.
b) All double kernels are included, Equation 1.45.

We turn now to the matter of the hydrogen bonding network for the 1YFG initiator tRNA that has been established by crystallography (see Nucleic Acid Database, NDB ID TRNA12, in Derivative Data: Hydrogen Bonding Classifications, <http://ndbserver.rutgers.edu/atlas/xray/structures/T/trna12/TRNA12-hbc.html>), based upon the experimental distances between putative hydrogen bonding donor and acceptor atoms. The interaction energy between a pair of kernels should be negative if that pair is stabilized by the presence of hydrogen bonds. Moreover, the magnitude of the interaction energy would be a measure of the hydrogen bonding stabilization. The interaction energies between pairs of kernels are data that are automatically generated in application of the KEM. The interaction energy, I , between kernels is defined as:

$$I_{ij} = E_{ij} - E_i - E_j, \quad (1.46)$$

where the symbols on the right-hand side of the equation retain their prior meaning. We found that in every instance, corresponding to the hydrogen bonding network established by crystallography, the interaction energy is negative, which is consistent with a stabilizing hydrogen bonding interaction between the relevant kernels. Thus the energetics available from the KEM provide independent confirmation of the hydrogen bonding network obtained experimentally from crystallography.

1.4.6

KEM Applied to Rational Design of Drugs

1.4.6.1 Importance of the Interaction Energy for Rational Drug Design

The importance of the interaction energy for rational drug design may be envisioned by consideration of Figure 1.16. The efficacy of drugs is based upon a geometrical “lock and key” fit of the drug to the target, complemented by an electronic interaction between the two. As indicated in Figure 1.16 by dashed lines, there will be several interactions between the drug and the kernels that constitute its target. The KEM delivers the *ab initio* quantum mechanical interaction energy between the drug and its target. This is computationally practical for molecular targets containing even tens of thousands of atoms. That is the great advantage of using the KEM for rational drug design. Moreover, not only is the total interaction energy obtained, so too as a natural consequence of the KEM approximation are the individual kernel components of the interaction energy. That is to say, the interaction energy of the drug with each individual kernel in the target is obtained. Thus the contribution from each kernel to the efficacy of binding to the drug, which may be large or small, and attractive or repulsive, may be obtained. In this way the most important interactions between the drug and the kernels of the target become evident.

Here we describe our calculations of the energy of various drug–RNA interactions. All calculations here employ a limited basis and the Hartree–Fock approximation. The definition of the interaction energy between any pair of kernels is Equation 1.46 in the previous section. In this section, we use it to calculate the interaction energies between the drug and RNA.

Knowledge of the list of the double kernel interaction energies is critical to rational drug design. That list determines the total drug–target interaction energy as well as

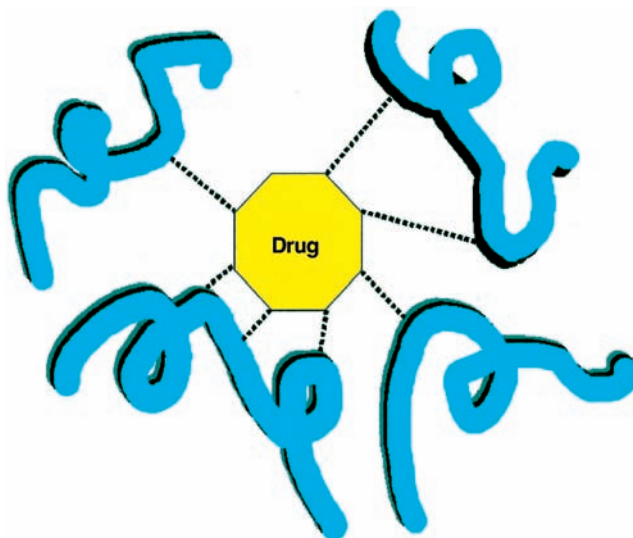


Figure 1.16 Sketch indicating the interaction of a putative drug molecule with its target, a very large medicinal molecular structure. The drug fits geometrically within a reactive “pocket” of the target. The dashed lines indicate interactions with the various kernels that compose the target. The interaction may be either positive or negative; both types of interaction (attractive and repulsive) are expected to occur.

the analysis of exactly which kernels contribute most importantly. Such knowledge may be obtained for the hundreds, or even thousands, of different chemical substitutions at various sites around the drug periphery, and the effect upon the interaction between the drug and the target computed. Such computational information can effectively replace the perhaps thousands of laboratory synthesis experiments needed to obtain related information. Moreover, it would be extremely difficult to obtain, by experimental methods, the double kernel interaction energies that flow naturally from implementation of the KEM to the problem.

1.4.6.2 Sample Calculation: Antibiotic Drug in Complex (1O9M) with a Model Aminoacyl Site of the 30s Ribosomal Subunit

The ribosome is a well-known target for antibiotic drugs. The crystal structure of one such drug, when attached to an A site RNA, is a complex named 1O9M, which has been solved [97] (Figure 1.17). Solvent water molecules included in the crystal structure are not shown in the figure. Utilizing the crystal structure we have calculated by the KEM the relevant energy quantities. These include the total energy of the complex made up of RNA, solvent and drug, and the separate RNA, solvent and drug molecules. We have obtained interaction energies descriptive of the drug–RNA target interaction, and of the hydrogen bonding network within the RNA molecule.

Table 1.13 displays the calculated energy results for the 1O9M drug–RNA target and solvent complex. The total complex, consisting of 1673 atoms, has been “broken”

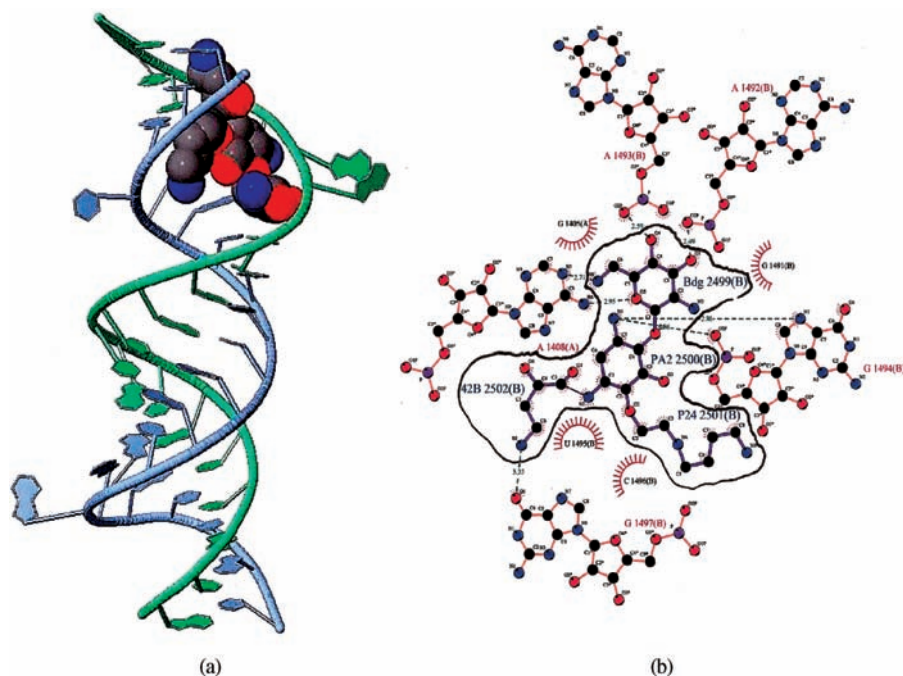


Figure 1.17 (a) Crystal structure of the drug–RNA complex 1O9M (molecule picture generated by Jmol Viewer); (b) drug–RNA interactions in the crystal. (Modified from PDBSum web site, LIGPLOT of interactions involving ligand.)

Table 1.13 Drug–target interaction energies (au) for rational design of drugs (see text for details).

Double kernels ij (RNA & drug)	Single kernel i (RNA)	Single kernel j (drug)	I_{ij} (au)	Kernel i –kernel j (RNA–drug)
–6219.279785	–4402.131144	–1817.149679	0.001038	1–15
–5984.204590	–4167.047454	–1817.149679	–0.007456	2–15
–5964.670898	–4147.520166	–1817.149679	–0.001053	3–15
–6183.414063	–4366.264309	–1817.149679	–0.000074	4–15
–6129.126465	–4311.976759	–1817.149679	–0.000026	5–15
–6129.086914	–4311.937498	–1817.149679	0.000263	6–15
–6038.689453	–4221.539976	–1817.149679	0.000203	7–15
–6219.246582	–4402.096702	–1817.149679	–0.000201	8–15
–5984.210449	–4167.060881	–1817.149679	0.000111	9–15
–5964.679199	–4147.529518	–1817.149679	–0.000002	10–15
–6183.388184	–4366.238058	–1817.149679	–0.000446	11–15
–6129.133301	–4311.980653	–1817.149679	–0.002968	12–15
–6129.113770	–4311.964279	–1817.149679	0.000189	13–15
–6038.665039	–4221.514542	–1817.149679	–0.000818	14–15
–6539.791016	–1817.149679	–4722.624492	–0.016844	16–15

into 16 kernels. Of these kernels, 1–14 represent the RNA target, kernel 15 represents the drug and kernel 16 represents the crystalline water of solvation.

Table 1.13 lists the interaction energies between the drug kernel and the kernels of RNA. The hydrogen atom positions have been energy optimized. The first three columns of the table list the calculated KEM energies for each double kernel and each of its two single kernel components, respectively. The fourth column lists each double kernel interaction energy. The fifth column names the double kernels. The single kernels that make up the RNA target are numbered 1–14. The antibiotic drug is kernel number 15 and the water is kernel 16. The interaction energy of RNA and drug, obtained from the sum of all the 14 RNA kernels and drug kernel interaction energies, is $-0.01\,124$ au, and the interaction energy of RNA in water and drug is $-0.02\,809$ au. We have shown how to begin with a crystal structure, and obtain therefrom quantum mechanical information not otherwise known from the structure alone. Such information includes the energy of the structure, the interaction energy between a drug and its target, and the analysis of such interaction energy in terms of the contribution of each contributing kernel pair. Thus the relative importance of individual kernels to the drug interaction efficacy can be assessed. This forms the basis of a rational drug design improvement from use of a lead drug structure.

1.4.6.3 Comments Regarding the Drug–Target Interaction Calculations

Assume the knowledge of a lead compound that displays the usual list of necessary properties, including adsorption, distribution, metabolism, excretion, and toxicity (ADMET). The critical factor that computational chemistry can contribute is the interaction energy between a putative drug and its target. If the target is a molecular structure containing thousands, or even tens of thousands of atoms, and if an *ab initio* quantum mechanical description of the interaction is to be obtained, then clearly an approximation such as that of the KEM is indicated. Thus, targets composed of peptides, proteins, DNA, RNA and various of their molecular composites can contain enormous numbers of atoms. Because the straightforward computational difficulty of a fully quantum mechanical calculation rises in proportion to a high power of the number of atoms in the molecular system, such calculations have typically been computationally impractical. The use of the KEM alleviates such computational difficulty by means of a formalism that divides a large molecular system into kernels, which are much smaller than the molecular system considered as a whole. Computations with each of the kernels are thus a relatively smaller problem, and can be assigned individually to separate nodes of a parallel processor. Thus a kind of twofold advantage accrues to the KEM, since individual calculations are smaller piecewise than otherwise, and they may be computed in parallel with modern computers designed for that purpose. The entire molecular system is reconstituted from a sum over kernels. What has been shown by the calculations of this chapter is that the KEM may be applied for purposes of rational design of drugs to the large molecules of medicinal chemistry. *Ab initio* results of expected high accuracy, within computational times of reasonable practicality, are obtained. Therefore, in general the KEM will be well suited for obtaining the interaction energy between drug molecules and their target medicinal chemical molecules of large size.

The point that has been made here is that the KEM can be useful for the rational design of drug molecules [98]. The key ideas that result and are useful for drug design are the interaction energy between a drug and its large molecular target, and all the component interaction energies for the individual double kernels.

1.4.7

KEM Applied to Collagen

This discussion combines a collagen molecule of given structure with quantum-mechanical KEM calculations to obtain the energies and interaction energies of a triple helix protein. Knowledge of such energetics allows one to understand the stability of known structures, and the rational design of new protein interacting chains. It is shown that the kernel energy method accurately represents the energies and interaction energies of each of the chains separately and in combinations with one another. This is a challenging problem for the case of large molecular protein chains. However, here the computational chemistry calculations are simplified, and the information derived from the atomic coordinates of the structure is enhanced by quantum mechanical information extracted therefrom.

1.4.7.1 Interaction Energies

The interaction energy among a triplet of protein chains is generalized to:

$$I_{abc} = E_{abc} - (E_a + E_b + E_c) \quad (1.47)$$

where the subscript indices name the triplet of protein chains in question, I_{abc} is the triplet chain interaction energy, E_{abc} is the energy of a triplet of chains, and E_a , E_b and E_c are each the energies of a single protein chain. Again, importantly, the sign of the interaction energy, I_{abc} , indicates whether the triplet of protein chains a, b and c altogether attract (negative I) or repel (positive I). It would be difficult to obtain from atomic coordinates alone the magnitude of the interaction energies that flow naturally from implementation of the KEM. The KEM delivers the *ab initio* quantum mechanical interaction energy between and among protein chains. This may be envisioned to be computationally practical for molecular structures containing thousands, or even tens of thousands of atoms.

1.4.7.2 Collagen 1A89

Collagen is a protein that is essential to the physical structure of the animal body. The molecule is made of three peptide chains that form a triple helix. These are incorporated in a vast number of ways to create structure. Collagen molecular cables provide strength in tendons, resilience to skin, support to internal organs, and a lattice structure to the minerals of bones and teeth. A repeated sequence of three amino acids forms the chains out of which the collagen triple helix is composed. Every third amino acid is glycine. Remaining positions in the chain often contain proline and hydroxyproline.

We selected for study a particular collagen molecule whose molecular structure is known, namely, 1A89 [99], and whose atomic coordinates are readily available in the

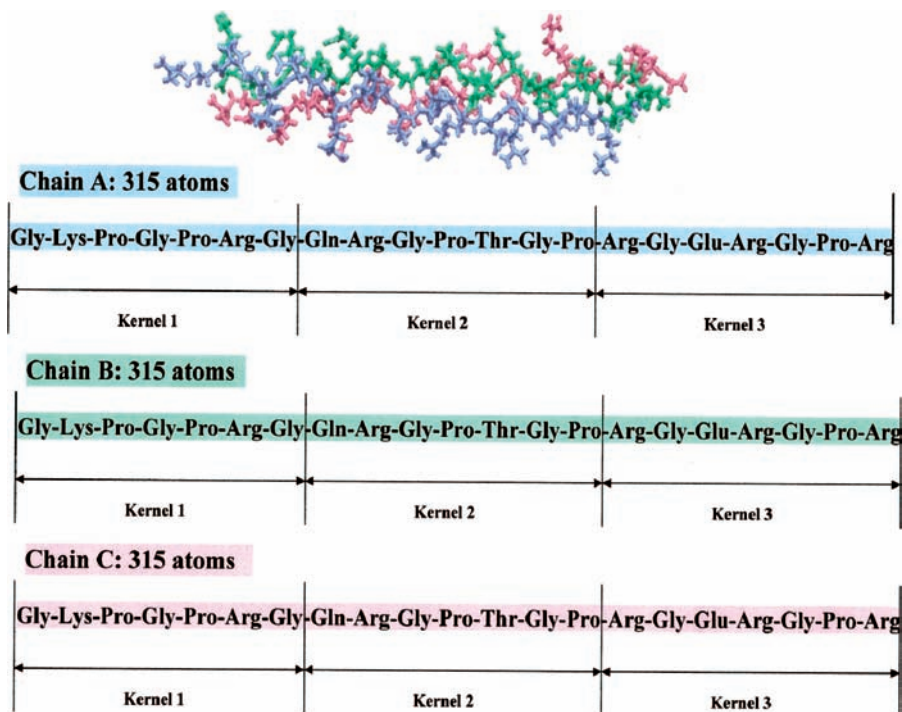


Figure 1.18 Picture of the collagen triple helix 1A89 and the primary structure of each of its individual protein chains broken into kernels.

Protein Data Bank. The atomic coordinates are the starting information from which the KEM proceeds. Clearly, from the structural role it plays in the animal body, collagen must be a stable molecule, with the chains of the triple helix structure adhering to one another. We applied the KEM to the molecular structure 1A89 to establish whether the approximation is sufficiently accurate to reveal the expected adhesion of the collagen triple chains.

Figure 1.18 shows a triple helix of protein chains that make up the collagen molecule that we have studied. Also shown is the amino acid primary structure of the three identical protein chains that make up the helix. Each chain is broken into three kernels, as shown in the figure. The total triplex contains 945 atoms, each chain contains 315 atoms, with kernels 1, 2 and 3 containing 96, 98 and 121 atoms, respectively. The atomic coordinates used in all of the calculations are obtained from the known molecular structure.

Table 1.14 contains the KEM calculations for each of the protein chains considered as a single entity. All calculations of this chapter are of quantum mechanical Hartree–Fock type, using an STO-3G limited basis of atomic orbitals. An “exact” result refers to the Hartree–Fock calculation of an entire molecule, including all of its atoms together, without use of the kernel approximation. The KEM calculated energies are meant to approximate the “exact” results. The difference between the

Table 1.14 Energy calculations for collagen triple helix (1A89) at the HF/STO-3G level of theory.

Chain	Atoms	Kernels	E_{HF} (au)	E_{KEM} (au)	$E_{\text{HF}} - E_{\text{KEM}}$ (au)	$E_{\text{HF}} - E_{\text{KEM}}$ (kcal mol ⁻¹)
A	315	3	-7381.86	-7381.86	0.0000	0.0047
B	315	3	-7382.16	-7382.16	0.0000	0.0260
C	315	3	-7382.83	-7382.83	-0.0002	-0.1027
Triple helix	945	9	-22146.92	-22146.91	-0.0059	-3.7332

two types of calculation is listed in both au and kcal mol⁻¹. One may conclude that the KEM calculation represents well the “exact” result. The percentage difference between the two types of calculation is small. For the single chains A, B and C the percentage differences are $1.0 \times 10^{-7}\%$, $5.6 \times 10^{-7}\%$ and $2.2 \times 10^{-6}\%$, respectively. Notice also that the percentage difference for the entire triple helix is only $2.7 \times 10^{-5}\%$. This level of accuracy is in accord with our previous experiences [32, 43, 82, 94, 96, 98].

Table 1.15 lists the calculation results for the triplex protein chains considered in pairs. The rows and columns are arranged as in Table 1.14, except that a new quantity, the interaction energy between the chains of the pairs, is also listed. As previously, the accuracy of the KEM energies is as expected, with differences for pairs AB, AC and BC of approximately $2.6 \times 10^{-5}\%$, $2.2 \times 10^{-5}\%$ and $2.8 \times 10^{-5}\%$, respectively. Notably, not only do we obtain the chain pair interaction energies but, as expected, the interaction is attractive.

Table 1.16 contains the calculation results for the full triple helix of the collagen structure. As indicated above, the KEM result for the total energy is accurate. The HF and KEM interaction energies of the triple helix are also listed.

Table 1.15 Interaction energy calculations^{a)} of chain pairs at the HF/STO-3G level of theory.

Chains	Atoms/ kernels	E_{HF} (au)	E_{KEM} (au)	I_{HF} (kcal mol ⁻¹)	I_{KEM} (kcal mol ⁻¹)	$I_{\text{HF}} - I_{\text{KEM}}$ (kcal mol ⁻¹)
AB	630/6	-14764.05	-14764.05	-23.1488	-20.7075	-2.4413
AC	630/6	-14764.71	-14764.71	-13.2896	-11.2950	-1.9946
BC	630/6	-14765.01	-14765.01	-8.6151	-6.2123	-2.4028

a) Interaction energies are calculated from: $I_{ab} = E_{ab} - E_a - E_b$.

Table 1.16 Interaction energy calculations^{a)} of collagen triple helix at the HF/STO-3G level of theory.

$E_{\text{HF}(abc)}$ (au)	$E_{\text{HF}(a+b+c)}$ (au)	$E_{\text{KEM}(abc)}$ (au)	$E_{\text{KEM}(a+b+c)}$ (au)	I_{HF} (kcal mol ⁻¹)	I_{KEM} (kcal mol ⁻¹)	$I_{\text{HF}} - I_{\text{KEM}}$ (kcal mol ⁻¹)
-22146.92	-22146.85	-22146.91	-22146.85	-41.48	-37.90	-3.58

a) Interaction energies calculated from: $I_{abc} = E_{abc} - E_{a+b+c}$, where $E_{a+b+c} = E_a + E_b + E_c$.

1.4.7.3 Comments Regarding the Collagen Calculations

The protein molecule chains, and their pair and triplex aggregates, taken from the molecular structure 1A89, in this chapter were treated within the context of the *ab initio* Hartree–Fock approximation. The basis set used was a limited basis, of Gaussian STO-3G type. A limited basis was chosen simply to make the energy calculations as convenient as possible, for a protein structure of this size. Previous numerical experience has shown that the KEM can be applied to a wide variety of molecules with good accuracy, and such expectations were realized in this instance.

We have shown how to begin with a known molecular structure and obtain therefrom quantum mechanical information not otherwise known from the structure alone. With collagen, such information includes the energy of the individual protein chains and their combinations in pairs and as a triplex. Importantly, the interaction energy between chains of a pair, or among those of a triplex are well represented by the KEM. Notably, the KEM approximation is sufficiently accurate to reveal the expected adhesion that must prevail among the collagen triple chains. This forms the basis of an understanding of the structure of collagen in particular, but more generally of a rational design of protein chain interactions [100].

What has been shown by the calculations here is that the KEM may be applied for purposes of obtaining the interaction energy between protein chains for an understanding of known molecular structures and for the rational design of proposed structures of considerable size.

1.4.8

KEM Fourth-Order Calculation of Accuracy

Remarkable accuracy has been achieved in the calculation of the energy of the ground state of the important biological molecule Leu¹-zervamicin [30], whose crystal structure is known and used in the calculations.

Figure 1.19 shows schematically defined kernels, double, triple and quadruple kernels; only these objects are used for all quantum calculations. The total molecular energy is reconstructed therefrom by summation over the contributions of the kernels and multiple-kernels up to the highest order of interaction to be imposed. In this description we extend the KEM to a fourth order of approximation. The aim, of course, is to increase the accuracy of the KEM calculations. Remarkable accuracy, as we indicate below, can be achieved.

1.4.8.1 Molecular Energy as a Sum over Kernel Energies

The formulas for invoking the KEM up to orders of approximation including double, triple and quadruple energies are displayed as Equations 1.48, 1.49 and 1.50, respectively [101]:

$$E_n^{\text{total}} = \sum_{\substack{i=1 \\ i < j}}^{n-1} E_{ij} - (n-2) \sum_{i=1}^n E_i \quad (1.48)$$

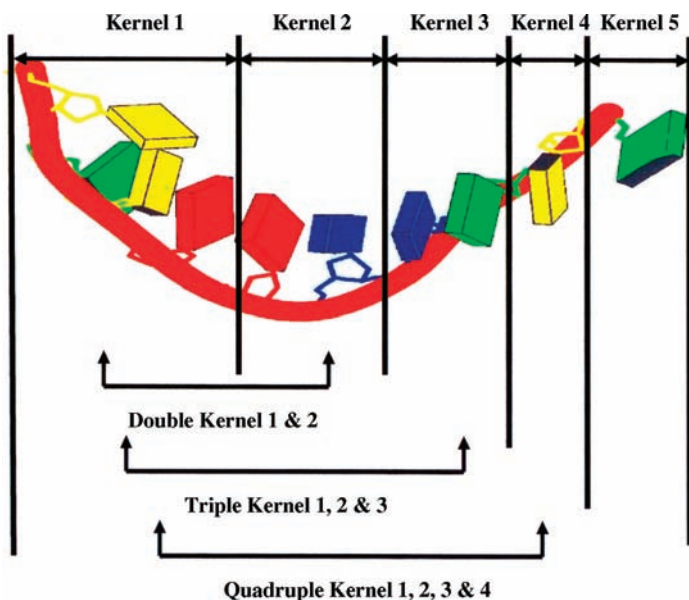


Figure 1.19 Abstract sketch of a molecule showing the definitions of the single, double, triple and quadruple kernels.

$$E_n^{\text{total}} = \sum_{\substack{i=1 \\ i < j < k}}^{n-2} E_{ijk} - (n-3) \sum_{\substack{i=1 \\ i < j}}^{n-1} E_{ij} + \left(\sum_{i=1}^{n-3} i \right) \left(\sum_{i=1}^n E_i \right) \quad (1.49)$$

$$E_n^{\text{total}} = \left(\sum_{\substack{i=1 \\ i < j < k < l}}^{n-3} E_{ijkl} \right) - (n-4) \left(\sum_{\substack{i=1 \\ i < j < k}}^{n-2} E_{ijk} \right) + \left(\sum_{i=1}^{n-4} i \right) \left(\sum_{\substack{i=1 \\ i < j}}^{n-1} E_{ij} \right) - \left(\sum_{j=1}^{n-4} \sum_{i=1}^j i \right) \left(\sum_{i=1}^n E_i \right) \quad (1.50)$$

1.4.8.2 Application to Leu¹-zervamicin of the Fourth-Order Approximation of KEM

We tested the accuracy achievable with the above formulas by application to the important biological molecule Leu¹-zervamicin (Figure 1.20). It is an antibiotic that transports potassium ions across cell membranes. Groups of zervamicin molecules assemble to form channels that serve to allow ion passage. The molecule has hydrophobic side chains extending from the peptide residues on one side and polar side chains extending from other peptide residues on the other side. A side chain of particular interest is a side chain of (residue 11) glutamine. This side chain is attached

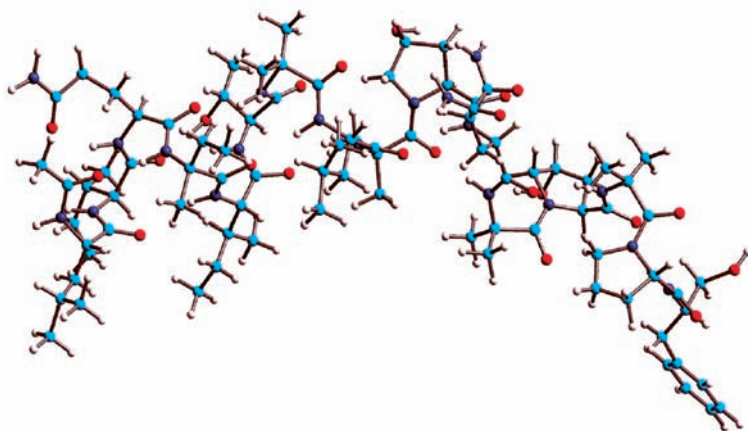


Figure 1.20 Leu¹-zervamicin (closed form) X-ray crystal structure.

on the hydrophobic side of the peptide, as determined from a crystal-structure investigation by Karle *et al.* [30]. The side chain acts as a gate that allows K^+ ions to proceed through a channel membrane. Glutamine 11, by a swinging motion, acts to open and close the channel each time an ion goes through it. If the structural arrangement of the zervamicin molecules in a cell membrane is similar to that in the crystal, it is possible that the zervamicin crystal structures offer a model of a gating mechanism on the level of atomic resolution. Table 1.17 lists the results of calculation related to Leu¹-zervamicin.

The program Gaussian 03 [102] was used to carry out the calculations, related to each of the formulas 1, 2 and 3 applied to the Leu¹-zervamicin molecule. Table 1.17 lists the results in order of increasing accuracy, reading across the table, for the double, triple and quadruple interaction approximations. For each of these approximations, the result listed as the exact energy is the brute force Hartree–Fock calculation on the full molecule using the limited basis of functions STO-3G. The table also lists the corresponding energy calculated by means of the KEM equations. Equations 1.48–1.50 are used to obtain the results listed for the double, triple and quadruple kernel approximations, respectively. The difference between the brute force and the KEM calculations are listed in au. Finally the results are listed on the basis of energy difference per atom in units of kcal mol^{-1} . As expected, the difference between the exact and the KEM results decrease as the order of the KEM approxima-

Table 1.17 HF STO-3G energy calculated up to fourth-order approximation of KEM.

	Single	Double	Triple	Quadruple
E_{KEM} (au)	–5851.8663	–5851.5469	–5851.5686	–5851.5703
E_{exact} (au)	–5851.5703	–5851.5703	–5851.5703	–5851.5703
ΔE (au)	0.296	–0.0234	–0.0017	0.0000
ΔE per atom (kcal mol^{-1})	0.70	–0.06	0.00	0.00

tions taken into account increase from double to triple to quadruple interactions. The numerical differences are -0.0234 , -0.0017 and 0.0000 au for the double, triple and quadruple interactions, respectively.

What are the limits of accuracy based upon the idea of quantum kernels? To answer this question, we calculated the KEM energy to an order of approximation including terms up to a fourth order of interaction among the kernels.

The standard of accuracy for these calculations was the brute force Hartree–Fock calculation of the energy for the full molecule in the same basis as for the KEM calculations. As the results of Table 1.17 show, the accuracy of the KEM increases with each order of approximation. For example, using Table 1.17, consider the differences on the basis of difference per atom. At the level of double kernel interactions (Equation 1.48) the magnitude of difference is of the order of 10^{-2} kcal mol $^{-1}$ atom $^{-1}$, an already small error. If triple kernel interactions (Equation 1.49) are invoked the magnitude of difference is reduced by an order of magnitude to 10^{-3} kcal mol $^{-1}$ atom $^{-1}$. Finally, using quadruple interactions (Equation 1.50) induces a further reduction in difference by an additional four orders of magnitude to 10^{-7} kcal mol $^{-1}$ atom $^{-1}$. For this molecule, at least, and those similar to it one need not contemplate going beyond the quadruple level of accuracy in the KEM approximation. The results here allow one to conclude that *ab initio* accuracy is obtainable for biological molecules within the KEM, carrying the approximation up to quadruple interactions. For large enough molecules, for which brute force calculations are not feasible, the KEM calculations will still be practicable, because the kernels and multiple kernels can be chosen to be very much smaller than the full molecule.

The KEM suggested here for especially high accuracy might find application in many problems in which the object of calculation might concern the quantum mechanics of large molecules. These include the rational design of drugs, peptide folding and the study of weak interactions among biological molecules.

1.4.9

KEM Applied to Vesicular Stomatitis Virus Nucleoprotein, 33 000 Atom Molecule

1.4.9.1 Vesicular Stomatitis Virus Nucleoprotein (2QVJ) Molecule

The vesicular stomatitis virus nucleoprotein (2QVJ) molecule (Figure 1.21) [103] is composed of five chains (A–E). Each chain has 421 residues, 6635 atoms and carries a charge of $+3$ au. The entire molecule contains 33 175 atoms. Each of the five chains is divided into 66 kernels, so we have altogether $66 \times 5 = 330$ kernels, contained in the entire molecule composed of five chains.

To calculate the energies of the molecule and each of its five sub-chains we have used the atomic coordinates of the crystal structure solved for the 2QVJ molecule. However, the crystal structure does not deliver the hydrogen atom coordinates. These have been added automatically to the heavy atoms of the crystal structure using the procedures of the computer program HyperChem [104]. The same amino acid sequence defines each of five chains that make up the full molecule. However, the relative positions occupied by each of the chains in the full molecule will differ, and

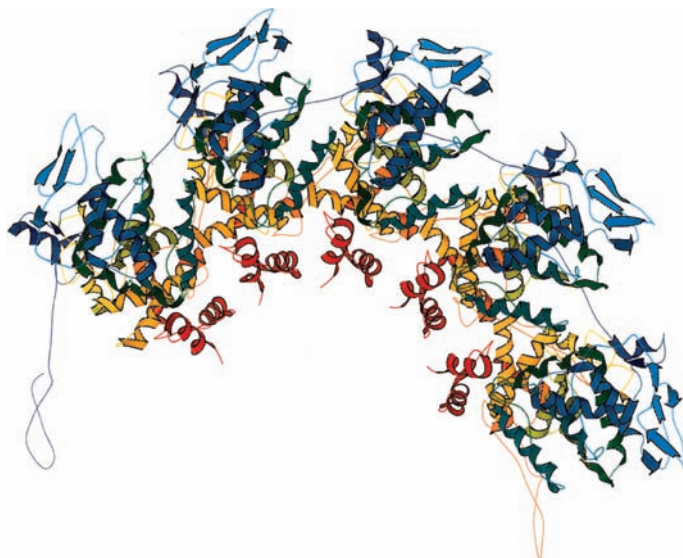


Figure 1.21 The crystal structure of 2QVJ is composed of five similar chains. The picture was generated with KiNG Viewer on the PDB web site.

this in turn will affect slightly the automatic placement of hydrogen atoms from chain to chain. As a result the energy of the chains will differ slightly. In Table 1.18 we list the limited basis (STO-3G) Hartree–Fock energies, calculated using Equation 1.45, for each of the five chains (A–E) that make up the full molecule.

Equation 1.45 was applied to calculate the total energy of each chain and the full molecule 2QVJ; the total energy of the full molecule is $-825\,954.57$ au.

1.4.9.2 Hydrogen Bond Calculations

We determined all hydrogen bond interactions between the chains and calculated their magnitudes using Equation 1.46. All the hydrogen bond calculations reported used 6-31G** basis functions in both the Hartree–Fock and MP2 approximations [105].

1.4.9.3 Comments regarding the 2QVJ Calculations

Within RNA viruses the viral genome RNA is completely enwrapped by a nucleoprotein. Vesicular stomatitis virus is such a case. The nucleoprotein in VSV is a ten-member “cylindrical” oligomer, half of which is the five-member oligomer 2QVJ, which retains a “half cylinder shape,” has been crystallized and is the subject of study

Table 1.18 Energies for individual chains of 2QVJ.

Chain	A	B	C	D	E
Energy (au)	−165191.04	−165192.19	−165189.29	−165193.27	−165189.20

here. As suggested by the authors of the crystal structure [103] the intermolecular interactions among the chains that make up the nucleoprotein play a critical role in providing the structural stability it acquires before encapsulation of the viral RNA. This conclusion follows from the crystal structure study. Knowledge of the crystal structure alone does not dictate the actual magnitude of the inter-chain interaction energies. However, given the crystal structure it becomes possible to extract from it the hydrogen bond donors and acceptors and with that information to calculate the inter-chain hydrogen bond interaction energies. That has been accomplished in this chapter.

To begin, we calculated the total energy of the entire 2QVJ molecule using the basic ideas of the KEM. The coordinates of the atoms used were obtained from the crystal structure at 2.8 Å resolution, except for the hydrogen atoms that were added using HyperChem [104]. For the fairly large number of atoms in the molecule as a whole (33 175 atoms) we calculated the energy in the Hartree–Fock approximation, using a limited basis of Gaussian orbitals. We considered each of the molecule’s five chains separately, breaking each chain into 66 kernels. A total of 330 kernels make up the whole molecule. Each kernel was chosen to contain approximately 100 atoms, which is of practical size. In this way, using Equation 1.45 the energy of each of the five molecular chains was obtained. Equation 1.45 also delivers a total energy for the full molecule of $-825\,954.57$ au.

It is likely that the inter-chain hydrogen bonds are among the most important contributors to the stability of the inter-chain structure of the whole molecule. All of the many possible inter-chain hydrogen bonds have been considered, their geometries displayed and their corresponding energies calculated [105]. The calculated hydrogen bond energies indicate the importance of correlation energy in representing the hydrogen bond interactions. Not only are the MP2 interaction energies quite a bit lower than the Hartree–Fock values, in some cases the Hartree–Fock results indicate repulsion (positive sign) instead of attraction (negative sign).

In summary the quantum calculations of VSV complement the crystal structure determination of the molecule by delivering the energetics that follow from knowledge of the atomic coordinates. One obtains by the KEM an approximation to the total energy of the whole molecule, and the individual chains that make it up. Principal contributors to the chain interactions are the hydrogen bonds between them. All of these hydrogen bond interactions have been calculated in both the Hartree–Fock and MP2 approximations.

1.5

Summary and Conclusions

Quantum mechanics and crystallography are borderline fields of great importance to the infrastructure of biochemistry. They are unified by sharing the electron density as a cornerstone object. X-rays scatter off the electron density. Density functional theory (DFT) obtains quantum properties as a function of the electron density. For a complete description of a molecular system to evolve from the electron density by

means of either experiment, as in X-ray scattering, or theory, as in density functional theory, *N*-representability is a practical mathematical necessity. For this reason, in our analysis of the X-ray scattering experiment, *N*-representability was introduced into the literature of crystallography. Clinton's equations are a practical method for finding a normalized single determinant *N*-representable electron density. Examples of this are given in the applications to crystals of beryllium and maleic anhydride. It may be mentioned in passing that the general importance of *N*-representability is now well recognized. It has been indicated by the US National Research Council to be one of the ten most prominent research challenges in quantum chemistry [106].

Among the earliest studies to recognize the importance of the fact that X-ray densities derive from a wavefunction is that of W. N. Lipscomb and coworkers concerning Hartree–Fock calculations of various models of the molecule of diborane. These were transformed into structure factors that were compared to the experimental ones from X-ray diffraction data. A “best-fit” provided a choice among quantum mechanical models. This work was reviewed in an article by Lipscomb [107].

To obtain *N*-representable densities from the X-ray data of ever larger biological molecules, one finds the number of experimental density matrix elements tends toward an inconveniently large number. Introducing the concept of quantum kernels reduces the size of the density matrices and makes their accurate determination practicable. This was demonstrated by the examples of a cyclic hexapeptide trihydrate and Leu¹-zervamicin.

The case of maleic anhydride suggests that nuclear positions are not much affected by variations of the density matrix. That is to say, the sum-of-spherical-atoms model used to solve the crystal structure gives good nuclear positions. If one simply adopts the atomic coordinates given by the crystal structure determination, and holds them fixed, that leads to the subfield of quantum crystallography called the kernel energy method (KEM). In the KEM a large biological molecule is mathematically broken into smaller pieces called kernels. Practicable calculations are carried out only on kernels (and multiple kernels). Subsequently, a direct summation over the kernels delivers the energy of the full molecule. Happily the kernel representation of the *ab initio* quantum problem is accurate. This point is made by application to peptides, proteins, DNA and tRNA examples. Moreover, the KEM has been tested with various basis functions and quantum methods. The KEM works for the whole variety of chemical models that have been tested. In addition, the computational time of the calculations is reduced by adoption of quantum kernels. At the level of double kernels approximation for calculating the total energy of large molecules, the KEM has a lower limit in both computing time and accuracy as represented by equation 1.44. The KEM calculation is flexible enough to accommodate the addition to equation 1.44 of any particular interaction energies between kernels. Of course increased interaction energies added to equation 1.44 will increase both accuracy and computational time associated with the results. The upper limit of accuracy and computational time is achieved in the representation of equation 1.45. So KEM makes possible a choice, between the lower & upper limits of accuracy and computing time, dependent on needs.

A natural consequence of the KEM is that it represents well the effect of the hydrogen bonding so important in the structure of biological molecules. The case of a

collagen triplex represents this fact. Also, KEM obtains the energetics that underlie a rational design of drugs. An example of this is shown in an antibiotic drug in complex (1O9M) with a model aminoacyl site of the 30S ribosomal subunit.

To push the limits of KEM accuracy and molecular size, two calculations have been put forward. Expressions for the quantum kernel expansion of the energy have been carried to a fourth order of accuracy. Applied to the ground state of the important biological molecule Leu¹-zervamicin, the fourth-order expansion achieves remarkable accuracy. In a quantum mechanical representation of a truly large molecule, the KEM has been applied to the 2QVJ molecule, composed of five chains, each of 421 residues. The entire molecule contains 33 175 atoms.

The density matrices for the large biological molecules that may be built up from the density matrices of kernels are rendered *N*-representable by means of Clinton's equations. The KEM discussed here applies with advantage to a host of problems in which the object of calculation concerns the true quantum mechanics of large molecules. These include the rational design of proteins, the study of protein folding and molecular self-assembly. Furthermore, perhaps it is not too speculative to insist that the quantum mechanics of large biological molecules can with advantage be brought to bear upon important medical problems more frequently than presently occurs. As a medical oncologist, with an interest in a multidisciplinary approach to cancer research, Provenzano has recently observed [108] that the application of quantum mechanics to medical problems is much to be wished for and encouraged. With this remark, surely the Pullmans would agree.

Acknowledgments

We thank the Office of Naval Research for supporting the work at the Naval Research Laboratory (NRL). One of us (L.M.) wishes to thank the U.S. Navy Summer Faculty Research Program administered by the American Society of Engineering Education for the opportunity to spend summers at NRL. Also, L.M. thanks NIH for grant RR-03037 the National Center for Research Resources, and PSC CUNY for grant 69701-00 38.

References

- 1 Von Neumann, J. (1927) *Nachr. Akad. Wiss. Göttingen, Math. Physik. Kl. II a. Math. Physik. Chem. Abt.*, 245–272.
- 2 Dirac, P.A.M. (1931) *Proc. Cambridge Phil. Soc.*, **27**, 240–253.
- 3 Huisimi, K. (1940) *Proc. Phys. – Math. Soc. Japan*, **22**, 264–314.
- 4 Mayer, J.E. (1955) *Phys. Rev.*, **100**, 1579–1586.
- 5 Tredgold, R.H. (1957) *Phys. Rev.*, **105**, 1421–1423.
- 6 Lowdin, P.O. (1955) *Phys. Rev.*, **97**, 1474–1489.
- 7 Coleman, J. (1963) *Rev. Mod. Phys.*, **35**, 668–686.
- 8 McWeeney, R. (1960) *Rev. Mod. Phys.*, **32**, 335–369.
- 9 Fano, K. (1957) *Rev. Mod. Phys.*, **29**, 74–93.
- 10 Terhaar, D. (1976) *Reduced Density Matrices in Quantum Chemistry*, Academic Press, New York.
- 11 Davidson, E.R. (1976) *Reduced Density Matrices in Quantum Chemistry*, Theoretical Chemistry; A Series Of Monographs, vol. 6, Academic Press Inc., London.

- 12 Gilbert, T.L. (1975) *Phys. Rev. B*, **12**, 2111–2120.
- 13 Frishberg, C.A. and Massa, L. (1981) *Phys. Rev. B*, **24**, 7018–7024.
- 14 Clinton, W.L., Galli, A., and Massa, L. (1969) *Phys. Rev.*, **177**, 7–13.
- 15 Massa, L., Goldberg, M., Frishberg, C.A., Boehme, R., and La Placa, S. (1985) *Phys. Rev. Lett.*, **55**, 622–625.
- 16 Dovesi, R., Pisani, C., and Ricca, F. (1982) *Phys. Rev. B*, **25**, 3731–3739.
- 17 Chou, M.Y., Lam, P.K., and Cohen, M.L. (1983) *Phys. Rev. B*, **28**, 4179–4185.
- 18 Huang, L., Massa, L., and Karle, J. (1999) *Int. J. Quant. Chem.*, **73**, 439–450.
- 19 Massa, L., Huang, L., and Karle, J. (1995) *Int. J. Quant. Chem. Quant. Chem. Symp.*, **29**, 371–384.
- 20 Huang, L., Massa, L., Karle, J., and Int, J. (1996) *J. Quant. Chem. Quant. Chem. Symp.*, **30**, 479–488.
- 21 Blessing, R. (ed.) (1990) *Studies of Electron Distributions in Molecules and Crystals, Transactions of the American Crystallographic Association*, vol. 26, Polycrystal Book Service, Dayton, OH.
- 22 Clinton, W.L., Galli, A.J., Henderson, G.A., Lamers, G.B., Massa, L.J., and Zarur, J. (1969) *Phys. Rev.*, **177**, 27–33.
- 23 Frishberg, C. (1986) *Int. J. Quantum Chem.*, **30**, 1–5.
- 24 Cohn, L., Frishberg, C., Lee, C., and Massa, L.J. (1985) *Int. J. Quantum Chem. Symp.*, **19**, 525–533.
- 25 Clinton, W.L. and Massa, L.J. (1972) *Int. J. Quantum Chem.*, **6**, 519–523.
- 26 Clinton, W.L. and Massa, L.J. (1972) *Phys. Rev. Lett.*, **29**, 1363–1366.
- 27 Clinton, W.L., Frishberg, C.A., Goldberg, M.J., Massa, L.J., and Oldfield, P.A. (1983) *Int. J. Quantum Chem. Symp.*, **17**, 517–525.
- 28 Hamermesh, M. (1962) *Group Theory*, Addison-Wesley, Reading, MA.
- 29 Karle, I.L., Gibson, J.W., and Karle, J. (1970) *J. Am. Chem. Soc.*, **92**, 3755–3760.
- 30 Karle, I.L., Flippen-Anderson, J.L., Agarwalla, S., and Balaram, P. (1994) *Biopolymers*, **34**, 721–735.
- 31 Frisch, M.J., Frisch, A., Foresman, J.B. *et al.* (1994) *Gaussian 94*, Gaussian, Inc., Pittsburgh, PA.
- 32 Huang, L., Massa, L., and Karle, J. (2005) *Int. J. Quantum Chem.*, **103**, 808–817.
- 33 Karle, I.L., Perozzo, M.A., Mishra, V.K., and Balaram, P. (1998) *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 5501–5504.
- 34 Karle, I.L., Gopi, H.N., and Balaram, P. (2003) *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 13946–13951.
- 35 Karle, I.L., Gopi, H.N., and Balaram, P. (2002) *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 5160–5164.
- 36 Ravindra, G., Ranganayaki, R.S., Raghothama, S., Srinivasan, M., Gilardi, R.D., Karle, I.L., and Balaram, P. (2004) *Chem. Biodiver.*, **1**, 489–504.
- 37 Karle, I.L., Prasad, S., and Balaram, S. (2004) *J. Peptide Res.*, **63**, 175–180.
- 38 Gopi, H., Roy, R.S., Raghothama, S.R., and Karle, I. (2002) *Helv. Chim. Acta*, **85**, 3313–3330.
- 39 Karle, I.L., Awasthi, S.K., and Balaram, P. (1996) *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 8189–8193.
- 40 Karle, I., Gopi, H.N., and Balaram, P. (2001) *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 3716–3719.
- 41 Karle, I.L., Das, C., and Balaram, P. (2000) *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 3034–3037.
- 42 Vijayalakshmi, S., Rao, R.B., Karle, I.L., and Balaram, P. (2000) *Biopolymers*, **53**, 84–98.
- 43 Huang, L., Massa, L., and Karle, J. (2006) *Int. J. Quantum Chem.*, **106**, 447–457.
- 44 Hehre, W.J., Stewart, R.F., and Pople, J.A. (1969) *J. Chem. Phys.*, **51**, 2657–2664.
- 45 Collins, J.B., Schleyer, P.v.R., Binkley, J.S., and Pople, J.A. (1976) *J. Chem. Phys.*, **64**, 5142–5151.
- 46 Binkley, J.S., Pople, J.A., and Hehre, W.J. (1980) *J. Am. Chem. Soc.*, **102**, 939–947.
- 47 Gordon, M.S., Binkley, J.S., Pople, J.A., Pietro, W.J., and Hehre, W.J. (1982) *J. Am. Chem. Soc.*, **104**, 2797–2803.
- 48 Pietro, W.J., Francl, M.M., Hehre, W.J., Defrees, D.J., Pople, J.A., and Binkley, J.S. (1982) *J. Am. Chem. Soc.*, **104**, 5039–5048.
- 49 Dobbs, K.D. and Hehre, W.J. (1986) *J. Comput. Chem.*, **7**, 359–378.
- 50 Dobbs, K.D. and Hehre, W.J. (1987) *J. Comput. Chem.*, **8**, 861–879.
- 51 Dobbs, K.D. and Hehre, W.J. (1987) *J. Comput. Chem.*, **8**, 880–893.

- 52 Schaefer, A., Horn, H., and Ahlrichs, R. (1992) *J. Chem. Phys.*, **97**, 2571–2577.
- 53 Schaefer, A., Huber, C., and Ahlrichs, R. (1994) *J. Chem. Phys.*, **100**, 5829–5835.
- 54 Ditchfield, R., Hehre, W.J., and Pople, J.A. (1971) *J. Chem. Phys.*, **54**, 724–728.
- 55 Hehre, W.J., Ditchfield, R., and Pople, J.A. (1972) *J. Chem. Phys.*, **56**, 2257–2261.
- 56 Hariharan, P.C. and Pople, J.A. (1974) *Mol. Phys.*, **27**, 209–214.
- 57 Gordon, M.S. (1980) *Chem. Phys. Lett.*, **76**, 163–168.
- 58 Hariharan, P.C. and Pople, J.A. (1973) *Theo. Chim. Acta*, **28**, 213–222.
- 59 Blaudeau, J.P., McGrath, M.P., Curtiss, L.A., and Radom, L. (1997) *J. Chem. Phys.*, **107**, 5016–5021.
- 60 Francl, M.M., Pietro, W.J., Hehre, W.J., Binkley, J.S., DeFrees, D.J., Pople, J.A., and Gordon, M.S. (1982) *J. Chem. Phys.*, **77**, 3654–3665.
- 61 Binning, R.C. Jr. and Curtiss, L.A. (1990) *J. Comput. Chem.*, **11**, 1206–1216.
- 62 Rassolov, V.A., Pople, J.A., Ratner, M.A., and Windus, T.L. (1998) *J. Chem. Phys.*, **109**, 1223–1229.
- 63 Rassolov, V.A., Ratner, M.A., Pople, J.A., Redfern, P.C., and Curtiss, L.A. (2001) *J. Comput. Chem.*, **22**, 976–984.
- 64 Dunning, T.H. Jr. and Hay, P.J. (1976) *Modern Theoretical Chemistry III*, vol. 3 (ed. H.F. Schaefer), Plenum, New York, pp. 1–28.
- 65 Petersson, G.A. and Al-Laham, M.A. (1991) *J. Chem. Phys.*, **94**, 6081–6090.
- 66 Petersson, G.A., Bennett, A., Tensfeldt, T.G., Al-Laham, M.A., Shirley, W.A., and Mantzaris, J. (1988) *J. Chem. Phys.*, **89**, 2193–2218.
- 67 Woon, D.E. and Dunning, T.H. Jr. (1993) *J. Chem. Phys.*, **98**, 1358–1371.
- 68 Kendall, R.A., Dunning, T.H. Jr., and Harrison, R.J. (1992) *J. Chem. Phys.*, **96**, 6796–6806.
- 69 Dunning, T.H. Jr. (1989) *J. Chem. Phys.*, **90**, 1007–1023.
- 70 Peterson, K.A., Woon, D.E., and Dunning, T.H. Jr. (1994) *J. Chem. Phys.*, **100**, 7410–7415.
- 71 Wilson, A., Mourik, T.v., and Dunning, T.H. Jr. (1997) *J. Mol. Struct. (Theochem)*, **388**, 339–350.
- 72 Dewar, M. and Thiel, W. (1977) *J. Am. Chem. Soc.*, **99**, 4499–4450.
- 73 James, J.P. (2001) A major enhancement in computational chemistry accuracy: MOPAC 2002, Stewart Computational Chemistry, Fujitsu Computational Chemistry Seminars.
- 74 Roothan, C.C.J. (1951) *Rev. Mod. Phys.*, **23**, 69–89.
- 75 Kohn, W. and Sham, L.J. (1965) *Phys. Rev.*, **140**, A1133–A1138.
- 76 Pople, J.A., Seeger, R., and Krishnan, R. (1977) *Int. J. Quantum Chem. Symp.*, **11**, 149–163.
- 77 Moller, C. and Plesset, M.S. (1934) *Phys. Rev.*, **46**, 618–622.
- 78 Bartlett, R.J. and Purvis, G.D. (1978) *Int. J. Quantum Chem.*, **14**, 561–581.
- 79 Sanger, F. and Tuppy, H. (1951) *Biochem. J.*, **49**, 463–481.
- 80 Hodgkin, D. (1970) *Verh. Schweiz. Naturforsch. Ges.*, **150**, 93–101.
- 81 Gursky, O., Badger, J., Li, Y., and Caspar, D. (1992) *Biophys. J.*, **63**, 1210–1220.
- 82 Huang, L., Massa, L., and Karle, J. (2005) *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 12690–12693.
- 83 Wahl, M.C., Rao, S.T., and Sundaralingam, M. (1996) *Biophys. J.*, **70**, 2857–2866.
- 84 Leonard, G.A., Hambley, T.W., and McAuleyhecht, K. (1993) *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, **49**, 458–467.
- 85 Soler-Lopez, M., Malinina, L., Tereshko, V., Zarytova, V., and Subirana, J.A. (2002) *J. Biol. Inorg. Chem.*, **7**, 533–538.
- 86 Vargason, J.M., Henderson, K., and Ho, P.S. (2001) *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 7265–7270.
- 87 Tari, L.W. and Secco, A.S. (1995) *Nucleic Acids Res.*, **23**, 2065–2073.
- 88 Valls, N., Uson, I., and Subirana, C.J.A. (2004) *J. Am. Chem. Soc.*, **126**, 7812–7816.
- 89 Qiu, H., Dewan, J.C., and Seeman, N.C. (1997) *J. Mol. Biol.*, **267**, 881–898.
- 90 Rozenberg, H., Rabinovich, D., Frolov, F., and Hegde, R.S. (1998) *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 15194–15199.
- 91 Nunn, C.M. and Neidle, S.J. (1995) *Med. Chem.*, **38**, 2317–2325.
- 92 Egli, M., Williams, L.D., Gao, Q., and Rich, A. (1991) *Biochemistry*, **30**, 11388–11402.

- 93 Shakked, Z., Rabinovich, D., Kennard, O., Cruse, W.B., Salisbury, S.A., and Viswamitra, M.A. (1983) *J. Mol. Biol.*, **166**, 183–201.
- 94 Huang, L., Massa, L., and Karle, J. (2005) *Biochemistry*, **44**, 16747–16752.
- 95 Basavappa, R. and Sigler, P.B. (1991) *EMBO J.*, **10**, 3105–3111.
- 96 Huang, L., Massa, L., and Karle, J. (2006) *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 1233–1237.
- 97 Russell, R., Murray, J., Lentzen, G., Haddad, J., and Mobashery, S. (2003) *J. Am. Chem. Soc.*, **125**, 3410–3411.
- 98 Huang, L., Massa, L., and Karle, J. (2007) *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 4261–4266.
- 99 Delacoux, F., Fichard, A., Geourjon, C., Garrone, R., and Ruggiero, F. (1998) *J. Biol. Chem.*, **273**, 15069.
- 100 Huang, L., Massa, L., and Karle, J. (2007) *J. Chem. Theory Comput.*, **3**, 1337–1341.
- 101 Huang, L., Massa, L., and Karle, J. (2008) *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 1849–1854.
- 102 Frisch, M.J., Trucks, G.W., Schlegel, H.B. et al. (2003) Gaussian 03, Gaussian, Inc., Pittsburgh, PA.
- 103 Zhang, X., Green, T.J., Tsao, J., Qiu, S., and Luo, M. (2008) *J. Virol.*, **82**, 674–682.
- 104 (2007) HyperChem 8.0.3 for Windows, Hypercube, Inc, Gainesville, FL.
- 105 Huang, L., Massa, L., and Karle, J. (2009) *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 1731–1736.
- 106 Stillinger, F.H. et al. (1995) *Mathematical Challenges from Theoretical/Computational Chemistry*, National Academy Press, Washington.
- 107 Lipscomb, W.N. (1972) *Trans. Am. Cryst. Assoc.*, **8**, 79–92.
- 108 Provenzano, A. (2009) *Oncol. Times*, **31**, 3–4.

2

Getting the Most out of ONIOM: Guidelines and Pitfalls

Fernando R. Clemente, Thom Vreven, and Michael J. Frisch

2.1

Introduction

Computational chemistry is extending its boundaries to increasingly large systems, ranging from the study of complex biological processes to the modeling of reactions in condensed phase (solution and solid states). One of the main challenges for the computational chemist in such cases is to find the balance between the accuracy of the results and the computational efficiency. Accurate computational methods scale very unfavorably with the size of the system. Molecular mechanics (MM), semiempirical and independent particle models such as Hartree–Fock (HF) or density functional theory (DFT) methods can be made to scale linearly, but cost still increases substantially with accuracy. Force constant calculations with the same methods scale quadratically or worse, and the more accurate wavefunction-based methods such as coupled cluster scale as N^6 or worse.

In general, the so-called hybrid methods offer a solution to the scaling problem. This approach recognizes and takes advantage of the fact that various regions of the system often play very different roles in the process under investigation. For example, in most enzymatic reactions the bond breaking and forming processes takes place only in the active site, and the effect of the protein environment is usually only steric and/or electrostatic. Similarly, for processes in solution, the role of the solute is clearly very different from that of the solvent. With hybrid methods, each region is treated with a different computational method. It often turns out that expensive computational methods are only required for the chemically active part of the system, while less expensive methods can be used for the supporting regions. The result is that very accurate results can be obtained for a fraction of the computational cost of conventional methods.

Over the years, various hybrid methods have been presented, which are conceptually quite similar but differ in a number of details. Most methods combine a quantum mechanical (QM) method with a MM method, which is generally referred to as QM/MM. Only a few hybrid methods combine QM with QM or more than two different computational methods. Other distinctions between the various methods

involve the description of the interaction between the regions or how the regions are connected when there is covalent interaction between them. We will focus on the ONIOM approach, which can in principle combine any computational methods, including QM with QM and also QM with MM, as well as more than two levels of theory.

In this chapter, we discuss several technical aspects related to the effective use of ONIOM. We first introduce details of the ONIOM method with the purpose of providing the necessary theoretical background and terminology. Then we present some fundamental guidelines for the application of the ONIOM methodology to the study of chemical problems. Finally, we describe in more detail some of the inconsistencies that can result from a naive application of these methods, especially when combining QM with MM methods.

2.2

QM/MM

Three papers, by Warshel and Levitt [1], Singh and Kollman [2] and Field, Bash and Karplus [3], are usually credited with having introduced QM/MM methods. In these papers the QM/MM potential was presented as:

$$E^{\text{QM/MM-EE}} = E^{\nu, \text{QM}} + E^{\text{MM}} + E^{\text{QM-MM}} \quad (2.1)$$

where $E^{\nu, \text{QM}}$ is the QM energy of the QM region, in the field ν generated by the partial charges of the MM region, and E^{MM} is the MM energy of the MM region (containing all the bonded and non-bonded MM terms that involve exclusively centers from the MM region). $E^{\text{QM-MM}}$ describes the interaction between the two regions and has two components. First, if there is covalent bonding between the QM and MM region, it contains the “border crossing” bonded MM terms that involve both QM and MM centers. Second, it contains all the MM van der Waals terms that involve one QM center and one MM center. $E^{\text{QM-MM}}$ does not contain the electrostatic interaction between the QM and MM region since this is already included in $E^{\nu, \text{QM}}$.

Kollman also suggested a simplified potential [2], which was further explored by Thiel [4]:

$$E^{\text{QM/MM-ME}} = E^{\text{QM}} + E^{\text{MM}} + E^{\text{Q, QM-MM}} \quad (2.2)$$

The QM energy, E^{QM} , no longer involves the potential from the MM region. Instead, the electrostatic interaction between the regions is calculated in $E^{\text{Q, QM-MM}}$ by assigning partial charges to the QM atoms, and using the regular expressions for point charge interactions from the MM force field. Thiel referred to the QM/MM potential using Equation (2.1), $E^{\text{QM/MM-EE}}$, as “electronic embedding QM/MM,” and to the potential using Equation (2.2), $E^{\text{QM/MM-ME}}$, as “mechanical embedding QM/MM.” The advantages of electronic embedding are that the wavefunction can be polarized by the charge distribution from the MM region, and that it provides a more accurate description of the electrostatic interaction between the two regions. However, it appears that in many cases the accuracy of the mechanical embedding

version is sufficient, and the simplified expression facilitates the implementation of methods to explore the potential surfaces.

When covalent interaction exists between the QM region and the MM region, the dangling bonds need to be capped in the QM calculation. Analogous to conventional model system calculations, the simplest solution is to use hydrogen atoms, which are then referred to as link atoms (LAs). A further complication of covalent interaction is that there may be partial charges from the MM region very close to the QM region. Since in molecular mechanics force fields the interactions between partial charges are scaled when they are less than three bonds apart, full inclusion of the partial charges in the boundary region in $E^{v, \text{QM}}$ may lead to overestimation of the electrostatic interaction, causing the polarization of the wavefunction to be unphysical. Kollman zeroed the charges that are less than three bonds away from the QM region. Although this will avoid overpolarization, it is rather arbitrary and also may lead to underestimation of the electrostatic interaction between the regions. Alternatives to zeroing the charges are to use delocalized (gaussian) charges instead of point charges [5, 6] or to redistribute the charges close to the boundary (see Reference [7] for more recent work).

2.3 ONIOM

In contrast to the merged Hamiltonian of traditional QM/MM methods, the ONIOM energy expression is written as an extrapolation [8–11]:

$$E^{\text{ONIOM}} = E^{\text{Model, High}} + E^{\text{Real, Low}} - E^{\text{Model, Low}} \quad (2.3)$$

“Real” and “Model” refer to the full system and the QM region, respectively; “High” and “Low” refer to the two levels of theory being combined. ONIOM uses link atoms to saturate the dangling bonds, which together with the QM region form the model system. The extrapolation can be viewed in two ways: (i) extrapolating the high level calculation on the model system by using the two low-level calculations to account for substituent effects or (ii) extrapolating the low-level calculation on the real system by using the two model system calculations to improve the accuracy in the region of the model system. Figure 2.1 illustrates the various components of the ONIOM scheme. When the high level is a QM method and the low-level is MM, then the ONIOM expression (2.3) is essentially the same as the QM/MM-ME expression (2.2), except for some details concerning the bonded MM terms that involve both QM and MM atoms. $E^{\text{Model, QM}}$ is equivalent to E^{QM} , and $(E^{\text{Real, MM}} - E^{\text{Model, MM}})$ in Equation (2.3) describes both the MM region and the interaction between the two regions, similar to $(E^{\text{MM}} + E^{\text{Q, QM-MM}})$ in Equation (2.2). Equation (2.3) can also be written as:

$$E^{\text{ONIOM}} = E^{\text{Model, High}} + S^{\text{Low}} \quad (2.4)$$

where S^{Low} is the substituent effect – the contribution from the low-level calculations to the ONIOM energy (E^{ONIOM}):

$$S^{\text{Low}} = E^{\text{Real, Low}} - E^{\text{Model, Low}} \quad (2.5)$$

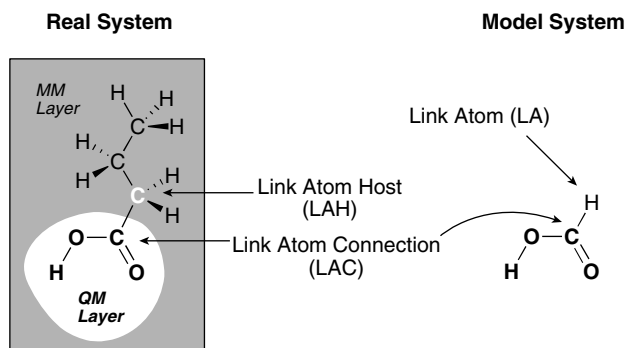


Figure 2.1 Components of the ONIOM scheme using butanoic acid as an example.

ONIOM uses link atoms to cap the dangling bonds that result from covalent interaction between the regions. For the potential surface to be well defined, the link atom must not introduce additional degrees of freedom. In ONIOM, we place the link atom on the line between the atom it is connected to (the link atom connection, LAC) and the atom it replaces (the link atom host, LAH), and obtain the LAC-LA distance by scaling the LAC-LAH distance [8, 12]. Besides the correct number of degrees of freedom, this scheme has the advantage that compression/elongation effects of the low-level calculation on the real system on the LAC-LAH bond are transferred to the model system calculations, through compression/elongation of the LAC-LA bond.

The acronym ONIOM stands for “Our own N-layered Integrated molecular Orbital molecular Mechanics.” As the name indicates, ONIOM can also be used for more than two layers, for example:

$$E^{\text{ONIOM}} = E^{\text{Real, Low}} - E^{\text{Mid, Low}} + E^{\text{Mid, Medium}} - E^{\text{Small, Medium}} + E^{\text{Small, High}} \quad (2.6)$$

Here a part of the real system has been chosen as a mid-sized system and a part of the mid-sized system has been chosen as the small system. The low-level calculation on the real system is then corrected for the difference between the low and medium levels of theory on the mid-sized system and then for the difference between the medium and high levels of theory on the small system. This is denoted ONIOM (High:Medium:Low). It is also possible to define ONIOM models with disjoint model systems, as discussed by Tschumper [13].

In the basic ONIOM scheme described above, the interaction between the atoms in the model system and the atoms that are only in the real system is described at the low level of theory. For ONIOM(QM:QM) this includes electrostatic effects on the electronic structure of the model system by the real system, albeit at the lower level of theory. For the case of ONIOM(QM:MM), this scheme corresponds to mechanical embedding, and does not include electrostatic effects of the real system atoms on the electronic structure, since the latter is computed only for the model system. To account for this effect, we extended the formalism of ONIOM(QM:MM) to include electronic embedding [14–16]. Because the model system needs to be identical for both the QM and MM calculation, we include the

environment charges in both, and do not change the real system calculation from Equation (2.3):

$$E^{\text{ONIOM(QM:MM)}-\text{EE}} = E^{v, \text{model, QM}} + E^{\text{real, MM}} - E^{v, \text{model, MM}} \quad (2.7)$$

To avoid overpolarization of the wavefunction, we may scale charges close to the QM region. Because these charges will then be scaled in both the $E^{v, \text{model, QM}}$ and $E^{v, \text{model, MM}}$ terms the balance will not change. The charge interactions that are over-counted or under-counted at the QM level in the $E^{v, \text{model, QM}}$ will be balanced at the MM level in the $E^{v, \text{model, MM}}$ term.

ONIOM is implemented in the Gaussian package for electronic structure calculations [17]. Most methods that are available in the package can be used in ONIOM, for either two- or three-layer calculations.

2.4

Guidelines for the Application of ONIOM

In this section we present a series of rules that can be followed for the successful application of ONIOM to the chemical problem of interest. We start with the basic considerations about the types of problems that can be studied with ONIOM. This will be followed by guidelines that ensure the accuracy of the ONIOM results. Finally, we discuss several technical issues that are specific to ONIOM(QM:MM) calculations.

Note that in some of our own work we have broken one or more of these rules. However, these studies have usually the goal of pushing ONIOM to its limits, and to investigate the behavior and performance for the most difficult cases. Also, in those studies, we usually need to carry out the full calculations at the high level of theory to compare the ONIOM result with the calculation it is intended to simulate (a high level calculation on the real system). This limits the size of the systems that we can consider and, with small systems, it is not always possible to follow all the guidelines below. Of course, in production calculations ONIOM is intended for use on larger systems that do not present the same problems.

■ Rule 1: Identify the Part of the System where the Action Takes Place

The essence of ONIOM is that it allows different parts of the system to be treated at different levels of theory. This implies that ONIOM will only be useful when one can identify parts that do require different levels of theory. Fortunately, this is often the case. For example, the bond breaking and forming in most chemical reactivity problems takes place in only a small part of the system. This region would then be treated at an appropriately high level of theory, while the remainder of the system can be treated with a

more approximate method. Similarly, many molecular properties, such as chemical shifts and normal modes, are generally quite local, and a suitable partitioning can be devised [18].

In some cases a localized active site cannot be identified, because all of the molecule participates equally in the process or property. Some properties, such as optical rotation, result from the entire system. Similarly, the electronic excitation in an extended polyene is fully delocalized over the molecule. In such cases, ONIOM cannot be applied (although substituent effects on such a chromophore can generally be included at a lower level of theory, provided the part of the system where the excitation takes place is included entirely in the high-level region).

Of course, it is not always easy to determine how the system needs to be partitioned. Often, atoms that are not directly involved in the process of interest but which are located near the active site must be treated accurately, either by including them in the model system or by choosing a low level of theory that is sufficiently accurate. The S-value test can be used to check whether the combination of partitioning and models are providing the desired accuracy. An example is given in Section 2.5.

■ **Rule 2: Only use ONIOM for Relative Energies, and Keep the Partitioning the Same for the Various Systems**

Through the specification of the computational methods that are combined and the partitioning of the system, ONIOM becomes a new model chemistry. Just as one does not directly compare absolute energies that are obtained with different conventional levels of theory, for example HF and DFT, one can also not compare, for example, the absolute ONIOM(HF:Amber) energy with the absolute HF energy. Only relative ONIOM energies, like the difference between reactant and product, are meaningful. Calculations with a correct ONIOM partitioning should closely approximate the relative energies as obtained at the (conventional) high level of theory. Thus, for example:

$$E^{\text{ONIOM(HF:XX)}} \neq E^{\text{HF}} \quad (2.8)$$

$$\Delta E^{\text{ONIOM(HF:XX)}} \approx \Delta E^{\text{HF}} \quad (2.9)$$

The same consideration applies to the partitioning, since it is a parameter in the definition of the ONIOM level of theory as well. Therefore, only ONIOM energies obtained with the same partitioning can be compared.

■ **Rule 3: The Low-Level Method Must be Accurate Enough to Describe the Substituent Effect**

The ONIOM calculation attempts to describe the entire system very accurately, but only using an expensive method for a small part. The rest of the system is treated with a lower level method. The low-level method takes care of two contributions to the total ONIOM energy. First, it describes the region of the molecule that is not included in the model system. Second, it describes the interaction between this region and the high level region. This implies that the interaction between the two regions is always included at the lower of the two levels of theory when the two levels are combined.

If the role of the low-level region is purely steric, a molecular mechanics force field or a semiempirical method often works satisfactorily. When there are electronic effects between the regions other than simple electrostatics, such as charge transfer, then the low level must be an appropriate QM method. Many benchmark studies have been published on ONIOM, which are a good starting point for determining partitionings and method combinations in a new study. We want to make a few general comments here:

- *ONIOM(QM:QM) schemes treated with very small high-level regions often perform best when the two methods are close in the same hierarchy, for example ONIOM(MP2:HF), ONIOM(B3LYP:BLYP) and ONIOM(MP4:MP2). The larger the model system, the less this is an issue.*
- *The high-level method must be better than the low-level method in every way. If we were to use ONIOM(CASSCF:B3LYP), the low-level method would include dynamic correlation, while the high-level method does not. This will cause part of the ONIOM extrapolation to be in the wrong direction, and the result could not be systematically better than the conventional CASSCF calculation on the model system alone, and would likely be significantly worse.*
- *A somewhat different way of thinking is needed when deciding the low-level method. The low-level method should be chosen as one able to describe properly the supporting role of the low-level region, but not necessarily the complete process under investigation. As a result, methods that otherwise would have never been considered as a conventional method to describe the process under investigation may be suitable as the low-level method in ONIOM. The clearest example is the application of*

ONIOM(QM:MM) to chemical reactivity problems. The standard MM methods are not able to describe the bond breaking and forming. However, as low-level method in ONIOM, MM methods are extremely powerful.

- *We can also search in a systematic way for the low-level method. Formally, the goal of an ONIOM calculation is to approach the result that would have been obtained if the full system were to be treated conventionally with the high-level method:*

$$\Delta E^{\text{ONIOM(High:Low)}} \approx \Delta E^{\text{High}} \quad (2.10)$$

The error is defined as:

$$\varepsilon = \Delta E^{\text{ONIOM(High:Low)}} - \Delta E^{\text{High}} \quad (2.11)$$

which can be rewritten as:

$$\varepsilon = \Delta S^{\text{Low}} - \Delta S^{\text{High}} \quad (2.12)$$

where S is the substituent value (S-value) and defined as:

$$S^{\text{level}} = E_{\text{real}}^{\text{level}} - E_{\text{model}}^{\text{level}} \quad (2.13)$$

Thus, the best low-level method is the one that gives a ΔS -value as close as possible to that at the high level of theory. The S-value allows a systematic investigation of the performance of low-level methods for a particular type of problem. Examples of using the S-value are given in Section 2.5 and in some of our previous studies [19, 20].

■ **Rule 4: A Link Atom Host (LAH) can Only be Replaced by a Single Link-Atom**

If an atom in the low-level region has two bonds with atoms in the high-level region, then it would either have to be replaced by two link atoms, which would lie nearly on top of one another, or by one link atom, which would leave a bond dangling (uncapped) in the model system. See also Rule 8.

■ **Rule 5: Avoid Having Link Atoms that are Involved in the Primary Process. They Should be as Far Away from the Process as Practically Possible**

ONIOM uses link atoms to cap the model system whenever there are covalent bonds between the regions. Obviously, a hydrogen atom is not the same as the atom that is being replaced, and an

error is introduced in its contribution to the ONIOM energy or other property. However, this does not necessarily introduce an error in the total ONIOM result, provided either or both of the following conditions are met:

- First, the model system enters the ONIOM expression twice ($E_{\text{Model}}^{\text{High}}$ and $E_{\text{Model}}^{\text{Low}}$), with opposite sign. If the errors resulting from the link atom are the same in both terms they will cancel in the ONIOM expression. Of course, the error is generally not exactly the same for the methods that are combined in ONIOM, and to what extent the errors cancel depends on the compatibility of those methods. When the levels of theory are similar or close in the same hierarchy (e.g., MP4 and MP2) the errors usually cancel well, and do not compromise the ONIOM accuracy. When the levels of theory are very different (e.g., DFT and MM), the errors may not cancel well. Thus, selecting compatible levels of theory minimizes the error resulting from the link atom.
- In many cases, though, the levels of theory are not compatible with respect to the link atom error, for example in the aforementioned example of combining a QM method with a MM method. This brings us to the second condition that minimizes this error: When the link atom is further away from the part of the system where the changes take place, the error will be the same in both reactant and product (or any two states of the system for that matter). In that case, the link atom errors that are still left in the ONIOM expression will cancel between the different species that are being investigated. In other words, as long as the link atom does not affect the process, there will be no error.

Combining the above two statements: When the levels of theory are very compatible, the link atom can be close to the reaction center, because any error will cancel between $E_{\text{Model}}^{\text{High}}$ and $E_{\text{Model}}^{\text{Low}}$. This is generally the case in QM:QM schemes. When the levels of theory are very different, for example in QM:MM calculations, it is better to have the link atom further separated from the process that is looked at. This will in practice often be the case anyway, because when the methods are very different, the model system will generally be larger (see Section 2.5 for more details).

■ Rule 6: Only Cut Non-Polar (C–C) Bonds

The error is minimized when only non-polar C–C bonds are replaced with C–H in the model system. The difference in the effect

of a non-polar C—C bond and a non-polar C—H bond on the rest of the model system is minimal. In contrast, breaking a C—O bond will replace the electron withdrawing effect of the oxygen in the real system with the more neutral C—H in the model system, which is more likely to affect the rest of the model system. As in Rule 5, the larger the model system and the further the link atoms are from the active site, the less sensitive the results are to these considerations.

■ Rule 7: Only Cut Single Bonds

The electronic structure of the model system must mimic that of the high-level region in the entire system. Clearly, replacing a fragment that is doubly bonded to the model region with a single link atom can dramatically change the electronic structure. One could in principle replace a doubly-bonded fragment with a divalent atom such as oxygen or beryllium, but this would introduce an error because of electronegativity differences and charge transfer.

■ Rule 8: Do Not Cut Through Cyclic Structures

This is similar to Rule 6, which warns against introducing polar link atoms. Cyclic structures are often strained, and the description at the two levels of theory can be very different, resulting in an error in the total ONIOM result. However, in many applications, it is not possible to avoid cutting through cyclic structures, for example when surfaces or other materials are studied. In those cases, the errors can be minimized by enlarging the model system, or by increasing the level of theory used for the low-level method in ONIOM.

■ Rule 9: In ONIOM(QM:MM) Calculations, the Bond Breaking and Forming Must Take Place at Least Three Bonds Away from the MM Region

For a detailed description we refer to Reference [14], and the example in the following section. Bonded terms in molecular mechanics methods depend on at most four centers. Since the molecular mechanics contribution to the potential must be continuous, the bond breaking and forming in the QM region must not affect the non-canceling MM terms. This is ensured by having at least three bonds between the MM region and the changing bonds in the QM region.

- **Rule 10: In QM:MM Calculations with Electronic Embedding, the Total Charge in the MM Region Must be Constant Throughout the Reaction Path**

With ONIOM(QM:MM) potentials employing electronic embedding, there are some considerations related to the charges in the various regions of the system. As a rule of thumb, the partial charges assigned to the centers in the molecular mechanics region should be stationary throughout the reaction path. In Section 2.5, we illustrate this with an example and explain when this rule may or may not be broken.

- **Rule 11: Run Test Calculations on Appropriate Systems**

It is not reliable to rely on intuition alone to determine the ONIOM partitioning and method. This is especially true with QM:QM combinations. It is always a good idea to test the partitioning and combination of methods by doing ONIOM calculations on examples that are small enough to permit the high-level calculation on the real systems to be carried out, and the predictions of the ONIOM model compared with the theoretical model it is intended to emulate.

If improvement in the ONIOM model is necessary, there are two ways to achieve this. One can increase the level of theory that is used for the low-level, and one can increase the size of the model region.

The partitioning is often more closely related to the chemistry than the method combination. An efficient strategy is, therefore, to first explore which partitionings can be used for a particular chemical problem using combinations of inexpensive computational methods. This will provide good candidates for the partitioning, which can then be considered in detail with the intended high-level method and potentially matching low-level methods. This avoids spending much computational time in the investigation of partitionings that are not chemically sensible.

- **Rule 12: There Must be a Significant Difference in Computational Cost between the Methods and a Significant Difference in Size between the Real and Model Systems**

If the model region is large (close to the size of the entire system) the evaluation of the $E_{\text{Model}}^{\text{High}}$ term will be nearly as costly as the

conventional calculation on the entire system. Similarly, if the low-level method is expensive, the evaluation of $E_{\text{Real}}^{\text{Low}}$ will be nearly as costly as the conventional calculation on the entire system. When this is the case, the calculation using the ONIOM scheme will not give a substantial reduction in computational cost and, in some cases, can even increase the cost.

2.4.1

Summary

The message from the guidelines above is that ONIOM is extremely powerful, but must be used with some care. Merely combining any two (or three) levels of theory using a random partitioning does not automatically give accurate results. However, when following chemical intuition, common sense and these guidelines, ONIOM will give highly accurate results for a fraction of the cost compared to conventional methods. Keeping this final rule in mind throughout the study will ensure reliable results and the use of ONIOM to its full potential.

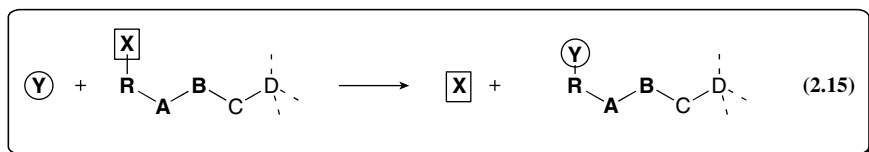
2.5

The Cancellation Problem

Chemical reactions involve processes of bond breaking and bond formation. Thus, in general, the number and type of bonds in the reactant species are different than those in the product species. This represents a challenge for molecular mechanics (MM) methods since the MM force field expressions explicitly depend on the number and type of bonds in the system through the bonded energy terms (bond stretches, angle bends, torsions). Equation (2.14) shows an example of a typical MM force field, in particular AMBER [21]:

$$E^{\text{MM}} = \sum_{\text{bonds}} K_r (r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{\text{eq}})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \\ + \sum_{i < j} \left[s_{ij}^{\text{vdw}} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + s_{ij}^q \frac{q_i q_j}{\epsilon r_{ij}} \right] \quad (2.14)$$

The MM energy function of the reactant thus involves different terms than that of the product and there is not a conversion of the former into the latter. The MM potential energy surface connecting the two structures is discontinuous and chemical reactions cannot typically be studied with MM methods. In contrast to MM methods, quantum mechanical (QM) methods do not make use of predefined bonds, and the energy of a chemical system only depends on the positions of the nuclei and the electronic state of the system. The QM potential energy surface connecting reactants and products is continuous, making the computational study of chemical reactions with QM methods possible.



MM Bonded terms that change from reactants to products:

Stretch **X-R**
 Bend **X-R-A**
 Torsion **X-R-A-B**

Stretch **Y-R**
 Bend **Y-R-A**
 Torsion **Y-R-A-B**

Figure 2.2 MM bonded terms changing in a general substitution reaction [(2.15)].

As mentioned above, the combination of QM and MM methods via ONIOM allows the study of very large systems in a computationally efficient way. The processes of bond breaking and bond formation that occur in a chemical reaction are treated with a QM method, while the supporting effect from the rest of the system is included at the MM level. A major decision in setting up an ONIOM(QM:MM) calculation is the partitioning of the system into the QM and MM layers. We have already discussed some of the rules to follow for the selection of the layers, but here we will focus on what we define as the cancellation problem. This problem simply refers to how to determine whether the potential energy surface described by the chosen ONIOM (QM:MM) partitioning is continuous.

Consider a general substitution reaction such as that shown in Figure 2.2. In this reaction, atom X in the reactant is substituted by atom Y in the product at the reaction center R. From the computational efficiency point of view, it would be desirable that the QM layer of our ONIOM(QM:MM) calculation be the smallest possible portion of the system. The choice of the inclusion of the three atoms directly involved in the bond breaking and bond formation processes (X, Y and R in Figure 2.2) is obvious, but how far from the reaction center R can we reasonably place the partitioning between the QM and MM layers?

Since X and Y would generally be two different atoms, the change in the number and/or type of bonds from reactants to products corresponds to a change in the MM bonded terms that are computed in the MM energy calculation. Thus, when setting up the ONIOM partitioning, the first goal must be to make sure that all these MM terms (Figure 2.1) are not included in the contribution from the MM region to the final ONIOM energy, S^{MM} (Equation (2.5)), so the bond breaking/bond forming process is exclusively treated at the QM level. Looking at the expression of the S^{MM} energy, it is equivalent to say that the terms shown in Figure 2.2 must be identical in the MM calculations of both the real and the model systems, so they cancel. Figure 2.3 shows three different ways to partition the system into QM and MM ONIOM layers at 1, 2 and 3 bonds away from the reaction center R. Clearly, from the picture, only when the MM layer is three or more bonds away from the reaction center R – partitioning 3 in Figure 2.3 – are the MM bonded terms involving the reactive atoms identical in the real and model system calculations, and therefore would cancel one another in the computation of the S^{MM} energy.

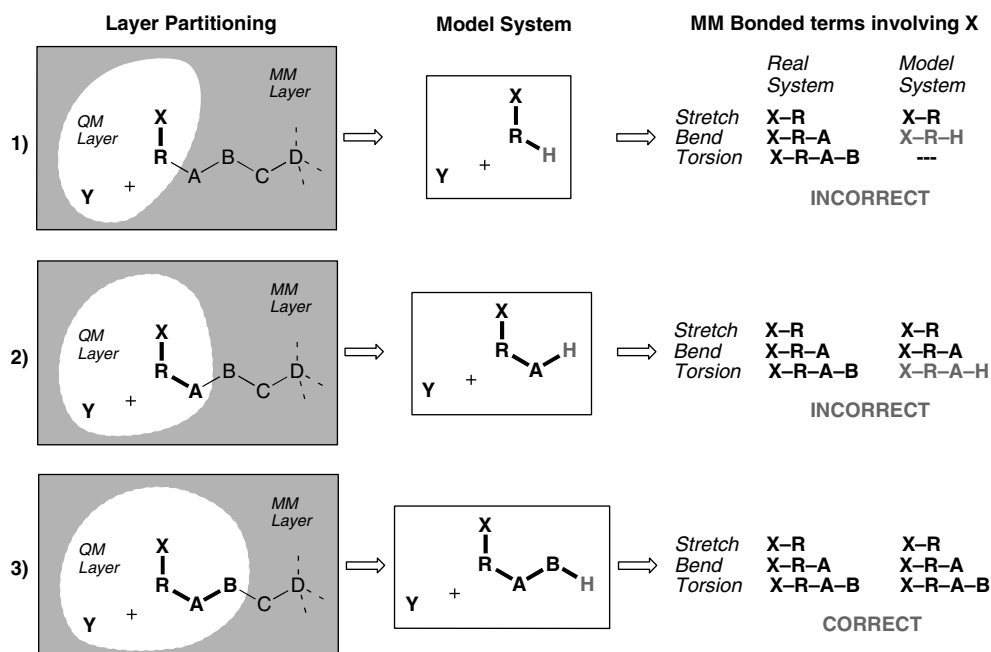


Figure 2.3 MM bonded terms included in the calculation of the model systems of three different ONIOM partitionings for the study of a general substitution reaction [(2.15) in Figure 2.2].

An ONIOM partitioning three or more bonds away from the reactive center (R in Figure 2.3) guarantees that the bond breaking/bond forming event is fully described at the QM level, because the MM force fields include bonded interactions only among atoms which are within three bonds of each other. The region of the system that is treated exclusively at the MM level, on the other hand, does not involve any changes in the number and/or type of bonds. The potential energy surface connecting reactant and product is then continuous. In other words, the ONIOM partitioning must be such that the energy of a given structure would be independent of the connectivity used (bonds as in the reactant versus bonds as in the product).

Although the safest way to guarantee the complete cancellation of the MM bonded terms in the reaction is to partition the layers three or more bonds away from the reactive center it may sometimes be desirable to use a smaller model system to speed up the calculations. If the three-bond rule is violated by choosing a smaller model system, then one must be careful that there is still proper cancellation of MM bonded terms between the calculations on different structures. This must be checked carefully, as described below.

The computation of the S^{MM} for different proposed partitionings and different connectivity patterns is a fast, useful way to determine whether a particular partitioning would provide the correct cancellation of the MM bonded terms. The S^{MM} energy values for a given structure computed using either the reactant connectivity or the

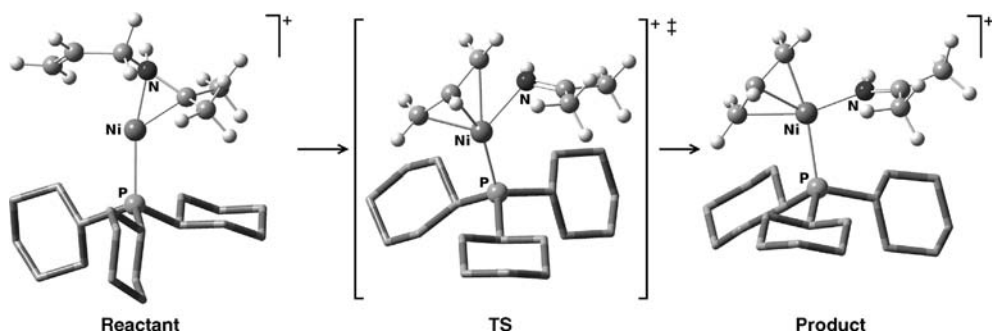


Figure 2.4 Oxidative addition step of the C–N bond activation reaction of N-allyliminium ions catalyzed by $(\text{Cy})_3\text{P}\text{Ni}$ (Torrent *et al.* [31]). High layer displayed in ball-and-stick model; low layer displayed in tube model. Hydrogen atoms in the low layer are hidden for clarity.

product connectivity must be identical if all the MM bonded terms involved in the reaction properly cancel.

To show the analysis of the S^{MM} values, we have chosen the reaction in Figure 2.4, from a computational study originally published by Torrent *et al.* [22]. In this publication, the reaction was studied at the ONIOM(B3LYP/6-31G(d):UFF) level with mechanical embedding. The QM layer chosen by the authors included all the atoms of the N-allyliminium cation, the metal center (Ni) plus the phosphorous atom of the phosphine ligand, while the three cyclohexyl groups of the phosphine were included in the MM layer. In the reactant state, the iminium cation is bound to the metal by the C and N atoms of the iminium group while the allyl fragment is not directly bound to the metal center. The reaction proceeds by scission of the C–N bond linking the iminium and allyl fragments. This is accompanied by a reorganization of bonds around the Ni atom; in the product, the allyl carbons are directly bound to the Ni atom, while the imine is only bound to the Ni atom through the N atom. By analogy with the general picture shown in Figure 2.3, the Ni atom in Figure 2.4 is thus the reactive center (R) of this system.

The authors chose to place the MM layer only two bonds away from the reactive center, the Ni atom, which is one bond closer than the “safe partitioning” introduced earlier in this section. Is this partitioning reasonable for this particular case? Could the size of the QM layer have been reduced any further, to be just one bond away from the Ni atom?

Computation of the S^{MM} value for a given system is a simple procedure, which consists of two MM energy calculations, one on the real system and another one on the model system, and taking the difference between the two (Equation (2.5)). We computed four different S^{MM} values for the TS structure in Figure 2.4, which correspond with two different connectivities (bonds as in the reactant or bonds as in the product) for two different ONIOM partitionings (Figure 2.5). The results are compiled in Table 2.1.

For the MM layer two bonds away from the reactive center (Ni), the S^{MM} values of the TS structure (Figure 2.4) are identical whether the reactant or product connectivity is used. All the MM energy components give the same S^{MM} value for

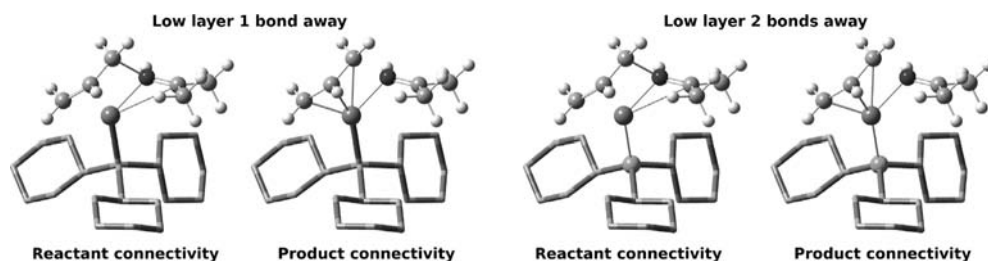


Figure 2.5 Models used for computation of the S^{MM} energy of the TS structure from Figure 2.4 using two different connectivities (reactant vs. product) and two different partitionings of the ONIOM layers (1 or 2 bonds away from the Ni atom).

both connectivities, which confirms that the bonding terms involved in the reaction properly cancel in this case. As a result, the computed ONIOM energy for the TS structure is independent of the connectivity pattern (reactant or product) used as input. In this particular system, then, and based on the analysis of the S^{MM} values, it is possible to reduce the size of the QM layer to have the MM layer only two bonds away from the reactive center.

In contrast, for the case of a MM layer only one bond away from the Ni atom the S^{MM} values using reactant or product connectivity markedly differ by more than 0.03 E_h (Table 2.1) This means that, at the TS geometry (Figure 2.4), the ONIOM potential energy surface using the product connectivity lies $\sim 20 \text{ kcal mol}^{-1}$ (1 kcal = 4.184 kJ) above the one using the reactant connectivity. The ONIOM energy of such structure depends on the connectivity and, therefore, this partitioning is inappropriate for the computational study of the reaction shown in Figure 2.4.

Table 2.1 S^{MM} energies (in E_h) for the TS structure in Figure 2.4.

	Low layer one-bond away			Low layer two-bonds away		
	Real	Model	S^{MM}	Real	Model	S^{MM}
Reactant connectivity						
Stretch	0.5128	0.5065	0.0064	0.5128	0.5066	0.0062
Bend	0.5723	0.4787	0.0936	0.5723	0.5289	0.0434
Torsion	0.0137	0.0104	0.0033	0.0137	0.0104	0.0033
Out-of-plane	0.0006	0.0006	0.0000	0.0006	0.0006	0.0000
Van der Waals	0.0474	0.0385	0.0089	0.0474	0.0366	0.0108
Total	1.1469	1.0347	0.1122	1.1469	1.0831	0.0638
Product connectivity						
Stretch	0.1261	0.1197	0.0064	0.1261	0.1198	0.0062
Bend	0.8207	0.6958	0.1249	0.8207	0.7773	0.0434
Torsion	0.0159	0.0126	0.0033	0.0159	0.0126	0.0033
Out-of-plane	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Van der Waals	0.0325	0.0234	0.0091	0.0325	0.0217	0.0108
Total	0.9952	0.8514	0.1438	0.9952	0.9315	0.0638

2.6

Use of Point Charges

We earlier introduced the two alternative ways of evaluating the electrostatic effect of the supporting MM region in an ONIOM(QM:MM) calculation: mechanical embedding (ME) and electronic embedding (EE). In both cases, point charges must be assigned to all the atoms in the system in a way that is representative of the electron distribution. In ME, the electrostatic effects involving the atoms that are only in the real system are only included in the MM calculation on the real system. In EE, the electrostatic interaction between the atoms that are only in the real system and the model system layer is included in the model system calculations as the effect of the point charges associated with the MM atoms. Thus, these charges can polarize the QM electron density.

Assignment of appropriate point charges can be a sensitive issue in ONIOM(QM:MM) calculations, in both ME and EE versions when a major reorganization of electron density takes place in going from reactant to product. Of course, charges must be assigned for pure MM calculations; the difference in ONIOM(QM:MM) calculations is that there can be a significant change in charge distribution in the QM calculation, whereas purely MM calculations permit no changes in the charge distribution.

We illustrate the sensitivity to the choice of point charges by computing the binding energy of a water molecule embedded in a protein, the bacteriorhodopsin (bR) system published by Vreven and Morokuma [15]. Figure 2.6 shows the vicinity of the binding site of the retinal chromophore in bR. For this example, the binding energy of water molecule W401 will be computed as the direct energy difference between the unbound and bound forms using fixed geometries. This is obviously a crude approximation to the binding energy of this molecule since we are not taking into consideration any solvation or relaxation effects of the system in going from bound to unbound, but it serves our purpose to show the sensitivity of the energy evaluations to the choice of point charges. As for the QM layer in our calculations, only the water molecule W401 and the carboxylate group from residue ASP85 are included. Both ME and EE versions of ONIOM are used in these calculations with the

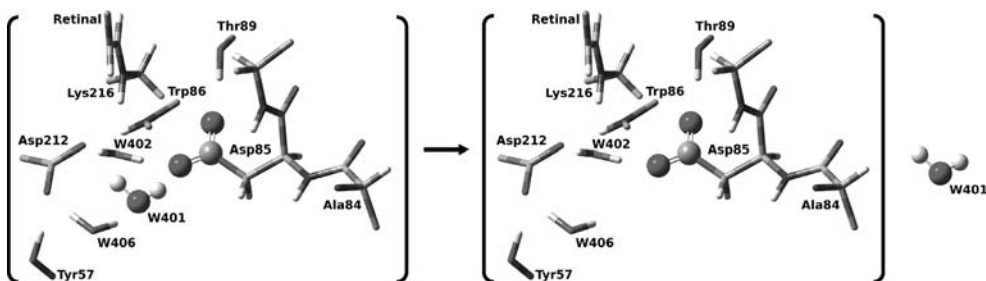


Figure 2.6 Model used for calculation of the binding energy of W401 near the retinal binding site in bacteriorhodopsin using the ONIOM(QM:MM) scheme. High layer displayed in ball-and-stick model; low layer displayed in tube model. Protein is truncated in the figure for better visibility.

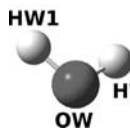


Figure 2.7 Model used for the derivation of atom charges for both bound and unbound states of the W401 + ASP85 system.

B3LYP/6-31G(d) level of theory for the QM layer and the AMBER force field for the MM layer.

All the atoms in the peptide chain and water molecules of this system have been initially assigned the corresponding atom charges according to the standard values used by the AMBER force field. We have also derived two other sets of charges for the atoms in the QM layer including the complete protein residue ASP85 (Figure 2.7). This model was constructed by taking the fixed geometries from the complete bR system and capping the N and C ends of ASP85 with acetyl and *N*-methylamino groups (also at positions taken from the adjacent residues to ASP85 in the bR system, i.e., ALA84 and TRP86). We performed single point B3LYP/6-31G(d) energy calculations on the bound (shown in Figure 2.7) and unbound forms, with independent calculations for W401 and ASP85, to obtain the electrostatic potential at the same QM level used in the ONIOM calculation.

The RESP program [23] was used to perform the fitting of the atomic charges to the QM electrostatic potential. Table 2.2 shows the three different sets of point charges used in our ONIOM calculations. Columns 2 and 3 give the standard atom charges for ASP and WAT residues in AMBER. “RESP” charges were derived from the B3LYP/6-31G(d) electrostatic potential to satisfy a total charge of -1 in the Asp...Water (bound) or Asp (unbound) portion of the model. In the “RESP constrained” charges, the global values for the MM region (excluding the LAH), the link atom host (LAH) and QM region are additionally constrained to be identical between bound and unbound states, and set to the average of the “RESP” values in each case (see last four rows in Table 2.2).

The three sets of charges shown in Table 2.2 illustrate three different approaches. AMBER charges represent the approach of using identical charges for both bound and unbound states. These charges were developed for their application in molecular dynamics simulations, providing an overall good description of the electrostatic effects for an average of configurations. The use of this type of charges is limited in the study of chemical reactions, since the reorganization of chemical bonds from the reactant to the product species generally requires different charges for either species.

Table 2.2 Point charges for the atoms in the system Asp85 + W401 (Figure 2.7) and the global charges for MM region (excluding LAH), the link atom host (LAH) and the QM region.

	AMBER		RESP		RESP constrained	
	Bound	Unbound	Bound	Unbound	Bound	Unbound
N	-0.5163	-0.5163	-0.4081	-0.4207	-0.3657	-0.4364
CA	0.0381	0.0381	-0.0976	-0.0710	-0.1316	-0.0469
C	0.5366	0.5366	0.4655	0.4424	0.4263	0.4612
O	-0.5819	-0.5819	-0.4993	-0.4974	-0.4871	-0.5071
CB	-0.0303	-0.0303	-0.1148	-0.1202	-0.1175	-0.1175
CG	0.7994	0.7994	0.6011	0.6124	0.6662	0.5487
OG1	-0.8014	-0.8014	-0.6276	-0.6574	-0.6398	-0.6408
OG2	-0.8014	-0.8014	-0.6838	-0.7144	-0.7221	-0.6931
HB1	-0.0122	-0.0122	0.0240	-0.0196	0.0197	-0.0168
HB2	-0.0122	-0.0122	0.0575	0.0565	0.0506	0.0670
HA	0.0880	0.0880	0.1259	0.1185	0.1369	0.1102
H	0.2936	0.2936	0.2581	0.2708	0.2537	0.2715
OW	-0.8340	-0.8340	-0.8531	-0.7324	-0.8609	-0.7324
HW1	0.4170	0.4170	0.3606	0.3684	0.3623	0.3684
HW2	0.4170	0.4170	0.3916	0.3640	0.4090	0.3640
MM	-0.1663	-0.1663	-0.0740	-0.1205	-0.0972	-0.0972
LAH	-0.0303	-0.0303	-0.1148	-0.1202	-0.1175	-0.1175
QM	-0.8034	-0.8034	-0.8112	-0.7593	-0.7853	-0.7853
Total	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000

The two sets of RESP charges (“RESP” and “RESP constrained”) represent the approach of using charges that respond to the changes in the electrostatic potential. This is the general case for the study of chemical reactions. The only difference between the two RESP approaches presented here is that in the “RESP constrained” charges the sum of the point charges associated with the atoms in the QM region do not change from the reactant to the product state (even though the individual atomic charges are different), and the same applies to the MM region and the LAH (Table 2.2). Given that the MM region only acts as support, the motivation behind using the “RESP constrained” charges is that the overall electrostatic potential applied from the MM region onto the QM region should remain approximately constant, while still responding to the changes in the local electrostatic potential occurring in the chemical reaction. The electrostatic terms are a large contribution to the energy of the system (especially when the total charge of the system is different than zero), so small changes in the total charge of either (or both) layers may translate into large deviations in the energy calculation. Looking at the “RESP” charges in Table 2.2, the charge on the LAH only changes by $\sim 0.005 e$ from bound to unbound states. The global changes in the MM and QM regions are $\sim 0.05 e$, that is, there is a charge migration of $\sim 0.05 e$ from the QM to the MM region when going from the bound to the unbound state. So, how important is it to constrain the charges of layers

Table 2.3 Binding energies (kcal mol^{-1}) of W401 computed at the ONIOM(B3LYP/6-31G(d): AMBER)^{a)} level of theory. The binding energies were computed at fixed geometries using three different sets of charges for the atoms in W401 and ASP85.

Embedding ^{b)}	AMBER	RESP	RESP constrained
EE – (+10)	31.6	27.0	27.9
EE – (+3)	31.6	30.6	27.8
EE – (neutral)	31.6	32.1	27.8
EE – (–10)	31.7	37.4	27.9
ME – (+10)	32.0	33.5	28.0
ME – (+3)	31.9	33.5	27.9
ME – (neutral)	32.0	33.5	27.9
ME – (–10)	32.1	33.6	28.0

a) EE = Electronic embedding; ME = mechanical embedding.

b) Total charge of the system given in parentheses.

to constant values? Does 0.05 e charge migration between layers really affect the results?

The total charge of the bacteriorhodopsin system used in our calculations is +3. Our ONIOM calculations for the binding energy of W401 were carried out at four different values for the total charge of the system: +10, the bR system plus seven sodium cations; +3, the bR system with no added counterions; neutral, the bR system plus three chloride ions; –10, the bR system plus thirteen chloride ions. The counterions were placed at arbitrary positions 30 Å away from W401 oxygen atom. This covers a range of common situations for calculations on biological systems, which often have a non-zero total charge. The expected behavior is that the binding energy of this water molecule would be dominated by the interactions in the local environment, and thus it should be fairly independent of the presence of counterions or any other charges placed far away from its binding site. Our results are collected in Table 2.3, while Figure 2.8 shows the binding energies of W401 as a function of the total charge of the system. The dependence of the binding energy with respect to the total charge of the system when using ONIOM-EE with the RESP charges is striking. The binding energies deviate by more than 5 kcal mol^{-1} in a range of 10 charge units, and this is just in our example where the charge migration between layers is merely 0.05 e . The slope of the line in Figure 2.8 is proportional to the charge migration. In contrast, the ONIOM-EE binding energies using point charges with global constant values for the MM and QM layers (“AMBER” and “RESP constrained”) are independent of the total charge of the system. It is thus crucial in ONIOM with electronic embedding to avoid any charge migration from the QM to the MM layer (or vice versa) to obtain reliable calculations.

On the other hand, the calculations using ONIOM with mechanical embedding show no dependency on the total charge of the system, even if the “RESP” charges are used. The electrostatic interactions between the MM and QM layers and within the MM layer are exclusively computed at the MM level, which is much less sensitive to small variations in the point charges.

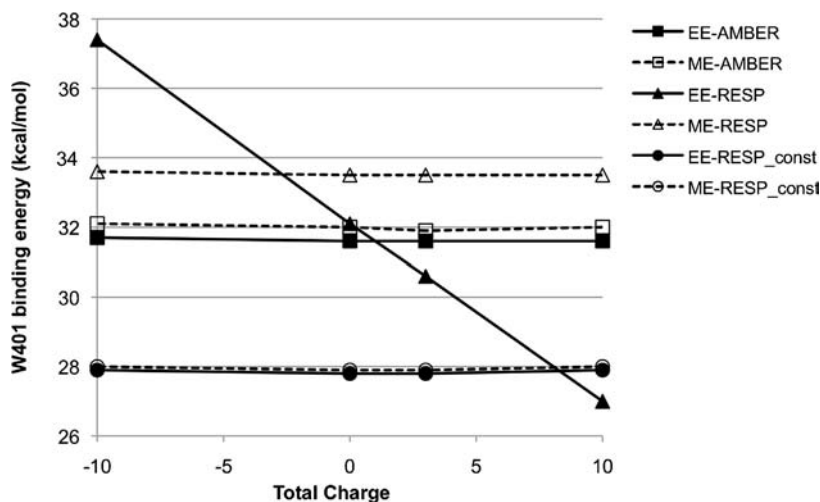


Figure 2.8 Binding energies (kcal mol^{-1}) of W401 as a function of the total charge of the system computed at the ONIOM(B3LYP/6-31G(d):AMBER) level of theory with both mechanical and electronic embedding, and using three different sets of charges.

Note the remarkably consistent results between mechanical and electronic embedding when the “RESP constrained” charges are used. These charges were derived from the electrostatic potential at the same level of theory used in the QM part of the ONIOM calculation, and they were allowed to respond to the changes in the electrostatic potential from the bound state to the unbound state. This shows that the apparent sensitivity of the results using electronic embedding to the presence or absence of charges far away from the active site was entirely an artifact of the inconsistency in MM charges between the reactant and product, even though the difference in total charges appeared to be small.

Finally, the discrepancy ($\sim 5 \text{ kcal mol}^{-1}$) between the ME calculations using “RESP” and “RESP constrained” is the result of the small size of the QM region in our calculations. In a production calculation, this difference should be taken as a strong warning that a larger model system should be considered.

2.7 Conclusions

We have presented guidelines that facilitate the effective use of the ONIOM model. The example in Section 2.5 illustrates how one can deviate from one of those rules to speed up the calculations if one uses the *S*-value test to ensure that proper cancellation still occurs. The test procedure illustrated there can also be applied in other cases, including ONIOM(QM:QM) calculations, to test the accuracy of possible partitions of the real system. Finally, we have also illustrated the issue of consistent charge assignments, which is specific to ONIOM(QM:MM) with electronic embedding, but which is important in this common use of ONIOM.

References

- Warshel, A. and Levitt, M. (1976) Theoretical studies of enzymatic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, **103**, 227–249.
- Singh, U.C. and Kollman, P.A. (1986) A combined *ab initio* quantum-mechanical and molecular mechanical method for carrying out simulations on complex molecular-systems – applications to the $\text{CH}_3\text{Cl} + \text{Cl}^-$ exchange-reaction and gas-phase protonation of polyethers. *J. Comput. Chem.*, **7**, 718–730.
- Field, M.J., Bash, P.A., and Karplus, M. (1990) A combined quantum-mechanical and molecular mechanical potential for molecular-dynamics simulations. *J. Comput. Chem.*, **11**, 700–733.
- Bakowies, D. and Thiel, W. (1996) Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J. Phys. Chem.*, **100**, 10580–10594.
- Amara, P. and Field, M.J. (2003) Evaluation of an *ab initio* quantum mechanical/molecular mechanical hybrid-potential link-atom method. *Theor. Chem. Acc.*, **109**, 43–52.
- Das, D., Eurenium, K.P., Billings, E.M., Sherwood, P., Chatfield, D.C., Hodoscek, M., and Brooks, B.R. (2002) Optimization of quantum mechanical molecular mechanical partitioning schemes: gaussian delocalization of molecular mechanical charges and the double link atom method. *J. Chem. Phys.*, **117**, 10534–10547.
- Lin, H. and Truhlar, D.G. (2005) Redistributed charge and dipole schemes for combined quantum mechanical and molecular mechanical calculations. *J. Phys. Chem. A*, **109**, 3991–4004.
- Dapprich, S., Komáromi, I., Byun, K.S., Morokuma, K., and Frisch, M.J. (1999) A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *J. Mol. Struct. (Theochem)*, **461–462**, 1–21.
- Svensson, M., Humbel, S., Froese, R.D.J., Matsubara, T., Sieber, S., and Morokuma, K. (1996) ONIOM: A multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and $\text{Pt}(\text{P}(t\text{-Bu})(3))(2)^+ \text{H}^{-2}$ oxidative addition. *J. Phys. Chem.*, **100**, 19357–19363.
- Humbel, S., Sieber, S., and Morokuma, K. (1996) The IMOMO method: integration of different levels of molecular orbital approximations for geometry optimization of large systems: test for *n*-butane conformation and $\text{S}(\text{N})_2$ reaction: $\text{RCl} + \text{Cl}$. *J. Chem. Phys.*, **105**, 1959–1967.
- Vreven, T. and Morokuma, K. (2000) On the application of the IMOMO (integrated molecular orbital + molecular orbital) method. *J. Comput. Chem.*, **21**, 1419–1432.
- Derat, E., Bouquant, J., and Humbel, S. (2003) On the link atom distance in the ONIOM scheme. An harmonic approximation analysis. *J. Mol. Struct. (Theochem)*, **632**, 61–69.
- Hopkins, B.W. and Tschumper, G.S. (2003) Multicentered approach to integrated QM/QM calculations. Applications to multiply hydrogen bonded systems. *J. Comput. Chem.*, **24**, 1563.
- Vreven, T., Byun, K.S., Komáromi, I., Dapprich, S., Montgomery, J.A. Jr., Morokuma, K., and Frisch, M.J. (2006) Combining Quantum Mechanics Methods in ONIOM. *J. Chem. Theory Comput.*, **2**, 815–826.
- Vreven, T. and Morokuma, K. (2003) Investigation of the $\text{S}_0 \rightarrow \text{S}_1$ excitation in bacteriorhodopsin with the ONIOM(MO:MM) hybrid method. *Theor. Chem. Acc.*, **109**, 125–132.
- Fermann, J.T., Moniz, T., Kiowski, O., McIntire, T.J., Auerbach, S.M., Vreven, T., and Frisch, M.J. (2005) Modeling Proton Transfer in Zeolites: Convergence Behavior of Embedded and Constrained Cluster Calculations. *J. Chem. Theory Comput.*, **1**, 1232–1239.
- Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman,

- J.R., Montgomery, J.A. Jr., Vreven, T., Kudin, K.N., Burant, J.C., Millam, J.M., Iyengar, S.S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G.A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J.E., Hratchian, H.P., Cross, J.B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev, O., Austin, A.J., Cammi, R., Pomelli, C., Ochterski, J.W., Ayala, P.Y., Morokuma, K., Voth, G.A., Salvador, P., Dannenberg, J.J., Zakrzewski, V.G., Dapprich, S., Daniels, A.D., Strain, M.C., Farkas, O., Malick, D.K., Rabuck, A.D., Raghavachari, K., Foresman, J.B., Ortiz, J.V., Cui, Q., Baboul, A.G., Clifford, S., Cioslowski, J., Stefanov, B.B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R.L., Fox, D.J., Keith, T., Al-Laham, M.A., Peng, C.Y., Nanayakkara, A., Challacombe, M., Gill, P.M.W., Johnson, B., Chen, W. Wong, M.W. Gonzalez, C., and Pople, J.A. (2004) Gaussian 03, Revision E. 01, Gaussian, Inc., Wallingford CT.
- 18 Karadakov, P.B. and Morokuma, K. (2000) ONIOM as an efficient tool for calculating NMR chemical shielding constants in large molecules. *Chem. Phys. Lett.*, **317**, 589–596.
- 19 Vreven, T. and Morokuma, K. (1999) The accurate calculation and prediction of the bond dissociation energies in a series of hydrocarbons using the IMOMO (integrated molecular orbital + molecular orbital) methods. *J. Chem. Phys.*, **111**, 8799–8803.
- 20 Vreven, T. and Morokuma, K. (2002) Prediction of the dissociation energy of hexaphenylethane using the ONIOM(MO:MO:MO) method. *J. Phys. Chem. A*, **106**, 6167–6170.
- 21 Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., and Kollman, P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179.
- 22 Torrent, M., Musaev, D.G., and Morokuma, K. (2000) Theoretical study of the mechanism of oxidative addition of allyl-ammonium and -iminium salts to low-valent metal complexes. Rationalization of selective C–N and N–H bond activation. *Organometallics*, **19**, 4402–4415.
- 23 Bayly, C.I., Cieplak, P., Cornell, W.D., and Kollman, P.A. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.*, **97**, 10269–10280.

3

Modeling Enzymatic Reactions in Metalloenzymes and Photobiology by Quantum Mechanics (QM) and Quantum Mechanics/Molecular Mechanics (QM/MM) Calculations

Lung Wa Chung, Xin Li, and Keiji Morokuma

3.1

Introduction

Metalloenzymes and photobiology, which use the metal(s) and photons (energy of the light) to facilitate otherwise difficult reactions in a specific manner, are unique and essential in our living organisms. Such reactions include oxygen storage and carriers by myoglobin (Mb) and hemoglobin (Hb), formation of a secondary messenger nitric oxide by nitric oxide synthase (NOS) and visualizing dynamics of living cells by fluorescent proteins (FPs). Therefore, reaction mechanisms in metalloenzymes and photobiological systems are of great importance and interest, and have been studied extensively by both experiment and theoretical calculations. In particular, combining the recent advancement of the experimental techniques with advanced quantum chemistry calculations has found to shed some light on mechanistic pathways. In this chapter, we summarize recent new insights into the reaction mechanisms of metalloenzymes and photobiology through quantum mechanics (QM) and quantum mechanics/molecular mechanics (QM/MM) calculations. Important heme-containing enzymes, cobalamin-dependent enzymes, fluorescent proteins and firefly luciferase are the focus of this chapter. The reader is referred to excellent reviews [1–7] related to our discussions.

The outline of this chapter is as follows. To model complex systems in metalloenzymes and photobiology, multiple computational means are often employed. First, we briefly introduce common computational strategies (methods and models) in Section 3.2. Recent studied key reaction mechanisms of selected metalloenzymes and photobiology are then discussed in detail in Sections 3.3 and 3.4, respectively.

3.2

Computational Strategies (Methods and Models)

3.2.1

Quantum Mechanical (QM) Methods

Among various approximations in quantum mechanical methods, the density functional theory (DFT) [8] method has recently been most widely used to model active-sites in metalloenzymes containing transition metal complexes, and ground-state chromophores in photobiology. The DFT method can give reasonable accuracy with lower computation cost than traditional *ab initio* wavefunction methods. Therefore, the DFT method has been widely applied to study energetics of large systems, which cannot be computed by highly-accurate *ab initio* methods. The foundation of the DFT method is the Hohenberg–Kohn theorem, which states that the ground-state properties of the system are determined by the electron density $\rho(\mathbf{r})$, where $\mathbf{r}$ is the three spatial coordinates. Unfortunately, the exact functionals for exchange and correlation are still unknown and have been approximated/parameterized in different ways. The hybrid B3LYP functional [9] is a popular functional in quantum chemistry calculations. B3LYP gives an average error of $4.14 \text{ kcal mol}^{-1}$ ($1 \text{ kcal} = 4.184 \text{ kJ}$) for the G3/05 test set containing enthalpies of formation, ionization energies, electron affinities, proton affinities and hydrogen-bonded complexes, but gave an error of 9 kcal mol^{-1} for large molecules having 28 or more valence electrons [10]. In contrast with the *ab initio* methods, increasing basis set size can lead to larger error for some functionals [10]. Recently, the accuracy of the popular functionals, for example, B3LYP, has been questioned, partly in relation to the medium- and long-range interaction [11]. Consequently, several new functionals, such as the M06 series, have been developed and parameterized to circumvent these problems [11].

Accurate experimental data for transition metal complexes are too limited to compare with computational methods. In particular, evaluating accurate energetics of first-row transition metal complexes is very challenging in quantum chemistry calculations [4a, 12], due to their low-lying excited-states (near-degeneracy). The error for the metal–ligand bond strengths was reported to be about $3\text{--}5 \text{ kcal mol}^{-1}$ by the B3LYP method [4a,b, 13]. The highly-accurate single-reference coupled cluster method (CCSD(T)) [14], which may require very large basis sets, can be applied only to very small systems. Therefore, the B3LYP method still remains as one of the feasible methods to study reaction mechanism, particularly for the large systems, including metalloenzymes [4a,b,i,m, 13]. Recently, re-parameterization of the B3LYP functional (so-called B3LYP*) was suggested to improve the energy order of different spin states [13a, 15].

The energetics derived from DFT functionals for bioinorganic chemistry were sometimes found to be inaccurate and to depend on the functionals. To obtain more reliable results, more advanced and complicated multi-reference methods, such as multi-reference self-consistent field (MRSCF) [16] and multi-reference perturbation theory (MRPT) [17] were suggested to be employed for bioinorganic

chemistry [18, 19]. However, possibly due to the limited active space, it was reported to be “irregular in its behavior” and gave large, unexpected errors [4m, 13a]. In addition, complicated multi-reference calculations are not of the black-box type and should be performed with great caution. Such drawbacks hampered wide application of methods like CASPT2 (complete active space multiconfigurational second-order perturbation theory) to bioinorganic chemistry, but would become more popular by means of more advanced methods [20] and increasing computational power.

Chemical reactions in photobiological systems involve excited-state potential energy surfaces (PESs), which remain challenging for quantum chemistry calculations [7c,g, 21]. Several excited-state methods, including time-dependent (TD) DFT [7c, 21d, 22], coupled cluster based methods and analogs (EOM-CCSD, CC2 [23] and SAC-CI [24]), multi-reference methods – such as complete active space self-consistent field (CASSCF), CASPT2 and multi-reference configuration interaction (MRCI) – are commonly used. Different methods have individual merits and weaknesses. The TD-DFT method can handle relatively large systems, but it is not often very accurate or reliable, and has a larger error for charge-transfer (CT) state, Rydberg state and ionic states of large π -systems as well as transition metal complexes [7c, 21d, 22e, 25]. The coupled cluster response theory methods can give more accurate excitation energies, with much higher computational cost, when the ground-state wavefunction of the system is dominated by a single determinant. Multi-reference methods more reliably describe excited states (if no intruder state [17a] exists for CASPT2), but the limited and affordable active space cannot be applied to very large systems. Recently, Thiel and coworkers performed benchmark calculations for the valence excited state (by TD-DFT, DFT/MRCI, CASPT2, CC2, CCSD and CC3) with a set of 28 medium-sized neutral organic molecules that are representative examples of the organic chromophores [26]. TD-B3LYP has an error of 0.27 and 0.45 eV for the vertical singlet and triplet excitation energies, respectively. The CC3 method gives similar results to multi-state CASPT2, while CC2 and CCSD have larger deviation.

As discussed above, the energetics of medium- and large-sized transition-metal complexes in metalloenzymes or excited-state chromophores in photobiology often cannot be easily quantitatively and conclusively evaluated by quantum chemistry calculations. The qualitative and semiquantitative computational results are more reliable when they are also supported by calculations with larger basis sets or other different reliable methods. In addition, these should not be regarded as “black-box” calculations, as desired electronic configurations need to be examined. In this connection, d orbital spin density analysis is one of the most straightforward means to depict electronic configurations for bioinorganic chemistry. For example, spin densities of roughly 3.7–3.8, 4.0–4.2, 3.0–3.3 were found to reside on the metal center in high-spin iron(II), (III) and (IV) complexes [27]. Such trend provides a fast and preliminary guidance. Further checking of the stability of the wavefunction and/or spin contamination is also desirable. Examination of the molecular orbitals (MOs) is essential to understand electronic configurations of the transition metal complexes and chromophores. Not to be too pessimistic, new qualitative or semiquantitative

insights into poorly-understood reactions in metalloenzymes or photobiology are sometimes discovered by theoretical calculations.

3.2.2

Active-Site Model

Before hybrid QM/MM methods were widely applied, the minimum and key parts of transition-metal complexes or chromophores were employed to be studied by the gas phase calculations (the so-called active-site model). As pointed out previously [13a], the active-site model has several merits. First, small and adequate models often give similar results to those for a very large model. Second, different feasible mechanisms can be carefully examined by efficient active-site model calculations. Third, the reliability of QM calculations for the active-site model is much more easily controlled than huge and complex QM/MM calculations. In addition, when the available X-ray crystal structures are not good (e.g., low resolution, not in an active form or missing the substrate), active-site model calculations may be a better way to calculate intrinsic reactivity of the critical active-site part. Moreover, in many current QM/MM calculations, the reaction pathway is approximated by relaxed scan calculations along one (or two) assumed reaction coordinate(s), or the calculated transition state is not verified by the very expensive Hessian calculations. Well-characterized reaction pathways using active-site model calculations could help to judge and provide more support to QM/MM calculations. Furthermore, comparing the active-site model and QM/MM calculations, which is highly recommended, can delineate the detailed protein effects.

To introduce the protein effects onto the active-site model, implicit solvent calculations, normally with low dielectric constant ($\epsilon = 4$) (or 20–80 for more polar surrounding) and a spherical probe with radius of 1.4 Å, are often carried out [28]. In addition, to suppress excessive flexibility of active-site models, a few of the atoms are sometimes frozen during the optimization, which gives structures closer to the protein. However, it might cause some strain if the active-site changes the structure (induced-fit model [29]) during the reaction [30]. Moreover, such a constrained approach cannot give accurate free energy via Hessian calculations, and should underestimate the protein effect (the geometric effect), when we compare active-site model calculations with QM/MM calculations.

3.2.3

QM/MM Methods

Pioneered by Warshel and Levitt in 1976 [31], combined quantum-mechanics/molecular-mechanics (QM/MM) approaches have been developed and have become a common protocol to study the reaction mechanisms by including the explicit solvent and whole protein [32]. The reader is referred to excellent QM/MM reviews [33]. In QM/MM calculations, the key and chemically interesting parts are described by a highly-accurate QM method and the rest of the protein and solvent are treated by very fast classical force fields. There are two general

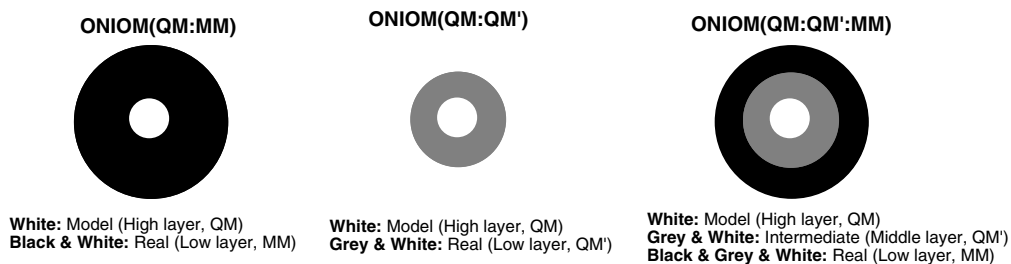


Figure 3.1 Schematic diagrams for two- and three-layer ONIOM methods.

approaches to evaluate the total energy (gradient and Hessian) of the system in the QM/MM calculations: additive and extrapolation schemes. In the additive scheme, the total energy of the system is a sum of the internal energies of the QM part (E_{QM}) and MM part (E_{MM}) and the QM-MM interaction energy (E_{QM-MM}) (Equation 3.1). Alternatively, Morokuma and coworkers have developed the extrapolated approach, the so-called our “Own N-layer Integrated molecular Orbital molecular Mechanics” (ONIOM) [34] (Equation 3.2 and Figure 3.1). QM-MM interactions are classically calculated by the MM force field in the original ONIOM mechanical embedding formalism (ONIOM-ME) (Equation (3.2)). Uniquely, such an extrapolated scheme can be extended to ONIOM (QM:QM') and ONIOM(QM:QM':MM) calculations (Equations 3.3 and 3.4 and Figure 3.1), in which QM-QM' interactions are described by the lower QM method (i.e., QM'). (The ONIOM method and its implementation in the Gaussian program are reviewed with practical examples in Chapter 2.):

Additive scheme:

$$E_{QM/MM} = E_{QM} + E_{MM} + E_{QM-MM} \quad (3.1)$$

Extrapolated schemes (ONIOM):

$$E_{\text{ONIOM(QM:MM)}} = E_{QM,\text{model}} + E_{MM,\text{real}} - E_{MM,\text{model}} \quad (3.2)$$

$$E_{\text{ONIOM(QM:QM')}} = E_{QM,\text{model}} + E_{QM',\text{real}} - E_{QM',\text{model}} \quad (3.3)$$

$$E_{\text{ONIOM(QM:QM':MM)}} = E_{QM,\text{model}} + E_{QM',\text{int}} - E_{QM',\text{model}} + E_{MM,\text{real}} - E_{MM,\text{int}} \quad (3.4)$$

To allow polarization of the QM wavefunction by the MM part, so-called electronic embedding (EE) [35], the MM fixed point charges as one-electron terms are incorporated in the QM Hamiltonian. The electronic embedding formalism was also developed for the ONIOM(QM:MM) and ONIOM(QM:QM') calculations [34h, 36]. However, polarization of the MM part by the QM part is not taken into account in the electronic embedding scheme. The unique three-layer ONIOM (QM:QM':MM) calculations allow mutual electronic polarization and charge transfer in the medium layer treated by the lower QM' method (Equation 3.4).

In contrast, an introduction of the mutual polarization and charge transfer along the QM/MM boundary has recently been developed by classical means [37]. Alternatively, for a better and more reliable description of the mutual polarization and charge transfer in the additive scheme for large QM systems, the extrapolated ONIOM(QM:QM') scheme could be incorporated in the additive scheme (Equation 3.5). In this so-called mixed scheme, the most important part is described by the more reliable QM method, and the interactions (including mutual polarization and charge transfer) between the most important part and its surroundings are treated by a less expensive QM method (i.e., QM', such as HF or a semiempirical method).

Mixed scheme:

$$E_{\text{QM/QM'}/\text{MM}} = E_{\text{QM,model}} + E_{\text{QM',int}} - E_{\text{QM',model}} + E_{\text{MM}} + E_{\text{QM'-MM}} \quad (3.5)$$

3.2.4

QM/MM Model and Setup

The preparation and setup for QM/MM calculations should be carefully performed, since errors caused by improper setup are hard to be estimated and corrected. The QM/MM models and setups have been tested and discussed by different groups [33i, j, 34f, i, 38]. We comment here on several key points:

- 1) Since it is difficult to distinguish electron density for carbon, nitrogen and oxygen atom in the modest resolution of the X-ray crystal structures, the possible orientation of histidine, asparagine and glutamine residues should be examined (e.g., by WhatCheck [39], MolProbity [40] or visual inspection) to obtain better hydrogen bonding or less steric repulsion.
- 2) Hydrogen bond networks may be important in the reaction. The optimal hydrogen bond networks of the proteins could be attained by some software, for example, PDB2PQR, HBUILD and WhatIf [41]. The protonation states of the titratable residues are also important and can be estimated by solving the Poisson–Boltzmann equation [42] or by empirical methods [43].
- 3) Since the force field is not always good for the active-site model, the active-site should be kept frozen during the MM optimization and dynamic simulations. The active-site model is then optimized by the subsequent QM/MM optimization. To avoid dramatic and artificial changes of the PESs during the QM/MM optimization, it is recommended that the entire system is further divided into the optimized MM region and the frozen MM region in the QM/MM optimization.
- 4) It is best for the QM-MM boundary to be far from the reaction site and at least three bonds from the reaction center to avoid discontinuity of the PESs [34i]. The ideal QM-MM boundary should be an inert C_{sp3}–C_{sp3} bond, if possible. (See Chapter 2 for a tutorial on how to avoid the pitfalls in an ONIOM calculation.)

3.3

Metalloenzymes

3.3.1

Heme-Containing Enzymes

3.3.1.1 Binding and Photodissociation of Diatomic Molecules

Subtle regulation of binding of exogenous gaseous diatomic molecules, O₂, CO and NO, to heme proteins, such as myoglobin, hemoglobin, oxidoreductases and soluble guanylate cyclase (sGC), plays an important role in maintaining their functions, such as storage, transportation, redox reactions and as sensors [44]. Many QM (mainly DFT) [45–47] and QM/MM calculations [47–49] have been performed to understand how these diatomic molecules bind or photodissociate, how proteins differentially bind with them, and what is the nature of this bonding. Generally, the calculated structural parameters for the deoxy (unligated) forms and the ligand-bound (ligated) forms by the DFT method are similar to the X-ray structures. The experiments and calculations also showed that the singlet Fe–CO bond is linear, while the singlet Fe–O₂ and doublet Fe–NO bonds are bent, due to orbital interaction between Fe(II) d π electrons and π^* orbital of the ligand [50]. However, the splitting energies for the unligated and ligated forms, as well as the binding energies, were found to be fairly dependent on the functionals.

Recently, Harvey and coworker adopted two simplified models, [Fe(CH₃N₂)₂(OH₂)] and [Fe(C₃H₅N₂)₂(OH₂)], to mimic the iron-porphyrin imidazole complex and performed very highly-accurate CCSD(T) energy calculations with Dunning's correlation consistent basis sets [45q]. In comparison with other DFT studies, pure functionals (e.g., BP86) were found to overestimate the stability of the low-spin state and the binding energy. On the other hand, hybrid functionals, such as B3LYP, give closer energies to the CCSD(T), except for the strongest binding NO ligand. B3LYP was shown not to be reliable for the binding and the trans-influence effect of NO, which was attributed to spin-contamination [45m, 45q]. Likewise, Radoń and Pierloot have performed expensive CASPT2 calculations with active spaces up to 16 electrons in 15 orbitals to calculate the binding of O₂, CO and NO, and compared the CASPT2 results with DFT functionals [45s]. CASPT2 calculations were suggested to reproduce the estimated experimental binding energy [45r,s],¹⁾ and OLYP functional was recommended to give closer energies to the CASPT2 method [45s]. Although very large active spaces were used to describe static correlation for the iron and ligands, these calculations do not cover the static correlation of the high-lying π and low-lying π^* orbitals of the porphyrin. One may need an even larger active space to examine any large effect of the static correlation of these orbitals in these calculations.

DFT calculations also showed that the ground state of ferrous carbonmonoxy-heme complexes is a closed-shell singlet state, while that for the oxy-heme complexes

1) extensive CASPT2(16e,14o) calculations found two van der Waals adducts in triplet and quintet states, which are even lower in energy than the singlet oxy-heme adduct, by 0.3–2.3 kcal mol^{−1} [45r].

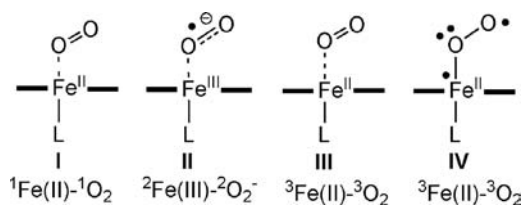


Figure 3.2 Several key models for the O_2 binding.

has been suggested to be an open-shell singlet [i.e., $\text{Fe(III)}-\text{O}_2^-$] with strong antiferromagnetic coupling between the ferric and superoxide radical anion. In fact, the nature of the O_2 binding is controversial. Several models, Pauling's model [I, $^1\text{Fe(II)}-^1\text{O}_2$], Weiss's model (II, $^2\text{Fe(III)}-^2\text{O}_2^-$), McClure's model [III, $^3\text{Fe(II)}-^3\text{O}_2$] and Goddard's model (IV, ozone model), have been proposed to explain the nature of the O_2 binding to heme (Figure 3.2) [51, 52]. Interestingly, as shown in Figure 3.3, different theoretical methods supported different O_2 binding models [48e, 52]. Jensen, Roos and Ryde have performed multistate (MS) CASPT2 calculations with 14 electrons in 13 orbitals for the ferrous oxy-porphyrin complex, which was shown to be multiconfigurational [45j]. They concluded that the CASSCF wavefunction is mainly contributed to by a mixture of both the Pauling and Weiss electronic configurations with roughly similar weight [45k]. In this regard, Shaik and coworkers have studied the nature of O_2 binding in myoglobin by QM

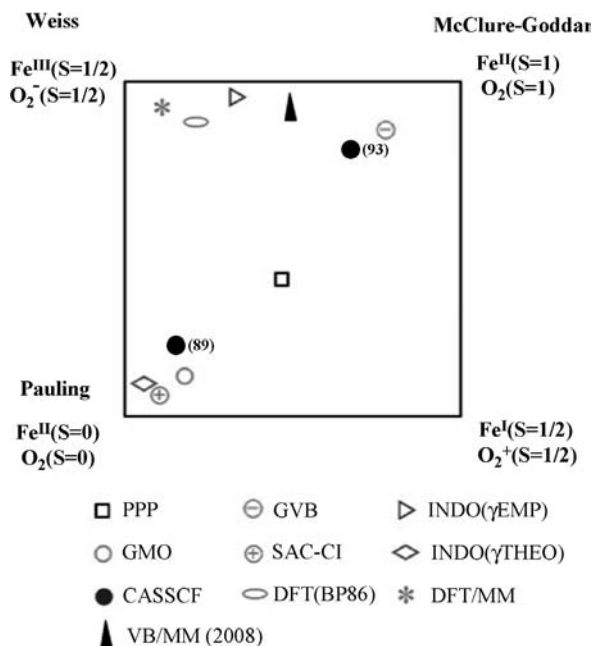


Figure 3.3 Nature of the O_2 binding by different computational methods. (Modified from Scheme 2 in Reference [48e].)

(B3LYP and CASSCF) and QM/MM calculations [48e]. The selected key CASSCF configurations for the bonding were further transformed into generalized valence bond (GVB) orbitals and eventually to VB structures. These calculations concluded that the wavefunction for O₂ binding state in myoglobin has VB coefficients of 0.71 from Weiss, 0.56 from McClure-Goddard and 0.18 from Pauling configurations [48e]. The electronic configurations for the doublet ground state of Fe-NO porphyrin complexes are even more complicated, as different functionals give two different configurations. Based on the calculated spin density, the five- and six-coordinated Fe-NO porphyrin complexes were suggested to be Fe⁺-NO⁻ and Fe^{~+0.5}-NO^{~-0.5} configurations, respectively, based on BP86, OLYP and CASSCF methods [45c, 45s].

To rationalize the discrimination against CO binding and in favor of O₂ binding in myoglobin (by about 4 kcal mol⁻¹) [1a], a distal histidine was proposed to push the favorable linear Fe-CO coordination to the bent conformation by steric hindrance. However, DFT calculations showed that bending or tilting the Fe-CO bond process is not high in energy [4d,k, 45a-c]. DFT and QM/MM calculations further supported that the more polar Fe-O₂ coordination, which has some character of the Fe(III)-O₂⁻, forms a stronger hydrogen bond with the distal histidine than the Fe(II)-CO coordination [4k, 45e, 47, 48a,c]. Differential hydrogen bond stabilization by the distal pocket was supported as a dominant factor for the ligand discrimination in myoglobin by the theoretical calculations [1a, 53].

Interestingly, Estrin and coworkers have shown the cooperative proximal effects (charge relay, orientation of the proximal histidine and strength of the Fe-histidine bond) on the O₂ binding in myoglobin and leghemoglobin by PBE and QM/MM (PBE/AMBER) calculations [47]. The calculated O₂ binding increases with the presence of an amide, which forms a hydrogen bond with the proximal histidine and thus increases charge transfer from the histidine to the bound O₂. The stronger hydrogen bond was attributed to favor more π back-donation from Fe(II) to π^* (O=O). However, the calculated binding energy decreases with the presence of an acetate and phenol, which further enhances the charge transfer. This result was rationalized by competition with σ -bonding between the imidazole and O₂. In addition, the calculated binding energy was found to increase (by 3.1 kcal mol⁻¹) by changing the orientation of the histidine from the staggered to the eclipsed conformation, especially in the presence of hydrogen bond with the amide. Changing orientation led to a shorter Fe-imidazole distance. Furthermore, a fairly linear correlation between the calculated binding energy and the Fe-imidazole distance was found. The shorter distance gives the higher binding energy. QM/MM calculations further showed that the distal histidine effect on the O₂ binding in leghemoglobin is smaller than myoglobin, because the histidine forms a hydrogen bond with one tyrosine and thus forms a weaker hydrogen bond with the O₂ (about 4 kcal mol⁻¹ based on the mutation calculation) in leghemoglobin. Instead, the proximal histidine in the optimized leghemoglobin is more staggered (78.4°), which was suggested to enhance the binding, than the optimized myoglobin (60.1°). This QM and QM/MM study showed interesting cooperative effects of the distal and proximal pocket on the binding, but the origin of the enhancement was not further explored. Notably, the binding affinity of the exogenous ligand is determined by two factors: (i) splitting energy between the most stable quintet state and the singlet state in the deoxy-heme and (ii) the intrinsic

binding affinity of the ligand to the singlet oxy-heme. Therefore, the former factor should not be overlooked and could also be important in the binding energy. Interestingly, Parrinello and coworkers showed that the out-of-plane displacement of the iron strongly destabilizes the singlet and triplet PESs of the deoxy-heme, compared to the most stable quintet form [45c]. Thus, the energetics of the different spin states was implicated to be modulated by the proteins via the interaction of Fe and the proximal histidine.

Cooperativity of binding in hemoglobin through conformational change of the quaternary protein structure (allostery [54]) has also been examined by QM and QM/MM calculations [45o, 48d]. It is difficult to experimentally decompose the origin of the allostery quantitatively. Recent QM/MM(B3LYP/AMBER) calculations demonstrated that the computed CO binding affinity for R-state (relaxed) is higher than that for T-state (tense) by about 7 kcal mol^{-1} , irrespective of the α - or β -subunits [48d]. More than 60% ($4.4\text{--}4.5 \text{ kcal mol}^{-1}$) of the binding energy comes from the protein–protein and protein–heme interactions. The singlet-state heme in the T-state was further found to have about $5\text{--}6 \text{ kcal mol}^{-1}$ more strain than the R-state, which is mainly due to the strain of the pyrroles of the heme during ligation.

Apart from the ligand binding, TD-B3LYP and QM/MM calculations have been performed to study the early stages of photodissociation of CO and O₂ in hemoglobin, myoglobin or neuroglobin [46, 49b]. Upon irradiation, the ground-state singlet ligated form is excited to the π, π^* state, followed by a rapid decay to another excited state E₁ in about 50 fs and then a decay via intersystem crossing (ISC) to give the high-spin deoxy-form on the picosecond time scale (Figure 3.4) [55]. Owing to the computational cost of the excited-state geometry optimization, the excited-state PESs were estimated from the ground-state (B3LYP) geometry optimization by elongating the Fe–ligand distance followed by the TD-B3LYP calculations. TD-B3LYP calculations showed that the lowest-energy π, π^* state (1 A' or so-called Q₁ state) was characterized as a local excitation (LE) on the porphyrin. The two lowest-energy π, π^* states, Q₁ and Q₂ states (Q-band), were shown not to be repulsive along the Fe–CO bond distance [46a,b, 49b]. When the Fe–CO bond was elongated to be about $2.0 \text{ \AA}$, the Q₁ and Q₂ states can cross with 5 A'' and 3 A' states (E-band) via two avoided

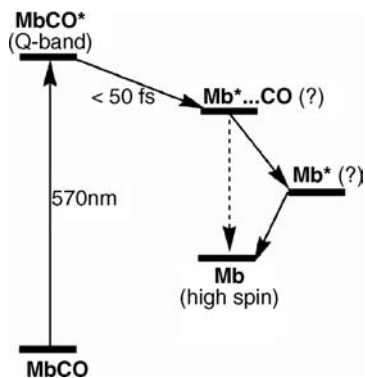


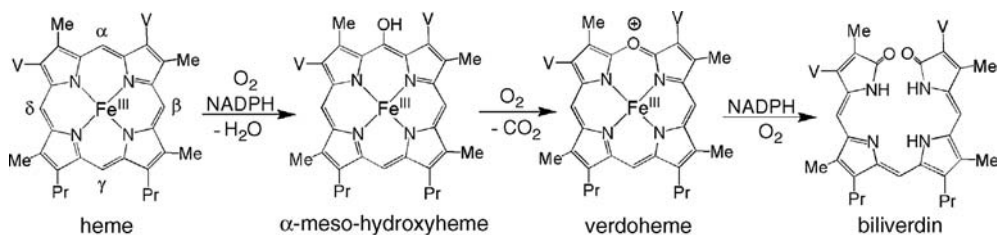
Figure 3.4 Schematic of the photodissociation of CO in myoglobin.

crossings with a small barrier of about 0.12–0.2 eV, which was suggested to be accessible via vibrational excitation [46a,b]. The excited-state E-states, which have an electron in the back-donative $\text{Fe}_d\text{-}\pi^*(\text{CO})$ antibonding orbital, are strongly repulsive and responsible for the bond dissociation. When the Fe–CO was elongated to about 2.5 Å, the E-band showed considerable charge transfer character from Fe to CO, imidazole, and nitrogen of the porphyrin [46b, 55b]. Kitagawa and coworkers have performed DFT and TD-DFT calculations and attributed a high barrier of about 0.6 eV to reach the charge transfer and strong Fe–CO σ^* repulsive states to non-photo-dissociable five-coordinate carbonmonoxy-heme [46d].

To explain the low quantum yield ($\sim 28\%$) of photodissociation of O_2 , compared to photodissociation of CO with quantum yield of near unity in myoglobin, bending of the nonlinear Fe– O_2 bond to afford the side-on conformer was proposed to cross with other excited states, providing a de-excitation pathway, while bending the linear Fe–CO bond did not cause such a crossing [46c]. Owing to Fe_d and $\pi^*(\text{O}_2)$ bonding character and a greater mixing of Fe_d and $\pi^*(\text{porphyrin})$ orbitals, the Q state is repulsive along the Fe– O_2 bond elongation and leads to direct O_2 dissociation without crossing with other excited states. However, the excited-state side-on conformer could also be obtained with a barrier of about 0.4 eV. Once the excited-state side-on conformer is formed, several lower-energy excited states were suggested to cross, which eventually lead to the ground-state end-on and side-on conformers, rather than the O_2 photodissociation.

3.3.1.2 Heme Oxygenase (HO)

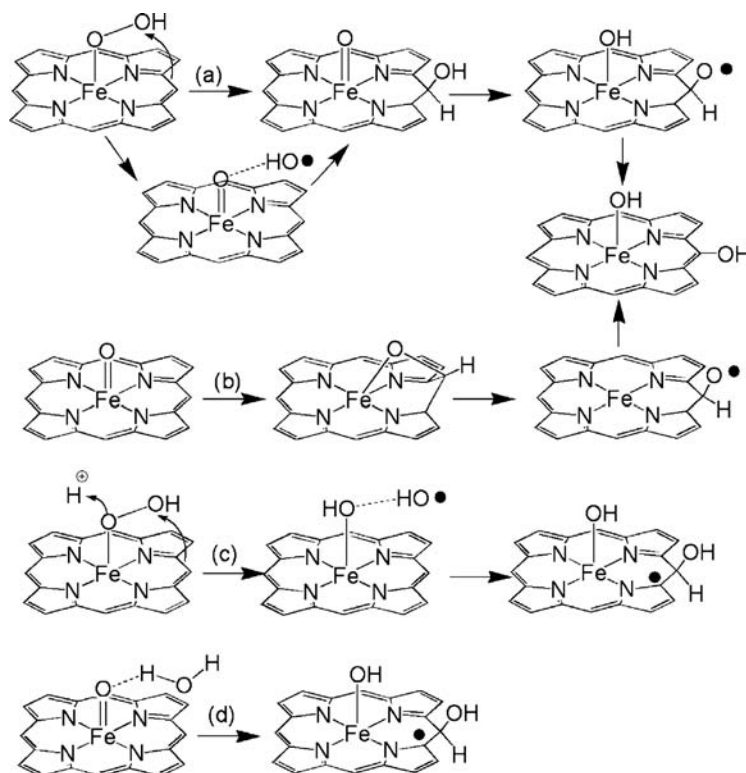
Heme oxygenase (HO) catalyzes the first step of heme metabolism, in which heme (iron-protoporphyrin IX) itself as the substrate is regioselectively oxidized at its α -meso carbon position to give biliverdin, carbon monoxide and free iron by consuming three O_2 molecules and seven electrons (Scheme 3.1) [1c,e]. All three products derived from HO are physiologically important: iron homeostasis, the formation of CO in the brain as a neurotransmitter, and the formation of the antioxidant biliverdin (and bilirubin). The first step in HO is regioselective oxygenation of heme to give α -mesohydroxyheme (Scheme 3.1). The ferryl-oxo complex, so-called compound I, and ferric-hydroperoxo complex have been proposed to be the active oxidant for the first step [1c,e].



Scheme 3.1 Heme degradation by heme oxygenase (V = vinyl, Pr = $\text{CH}_2\text{CH}_2\text{CO}_2\text{H}$).

Recently, the groups of Yoshizawa and Shaik independently carried out B3LYP and QM/MM calculations to elucidate the possible mechanism of the first step [56].

Yoshizawa's calculations showed that rate-determining concerted direct attack of either the distal hydroxyl group of the ferric-hydroperoxo complex or the oxo of compound I to the α -meso carbon of heme has a very high reaction barrier (about 39.9–42.9 kcal mol⁻¹), due to significant deformation of the heme in the concerted transition states [pathways (a) and (b) in Scheme 3.2] [56a]. The very high barrier for the concerted pathway was also supported by Shaik and coworkers [56b]. In addition, they found that the stepwise pathway via an initial homolytic O–O cleavage followed by a low-barrier radical addition of the hydroxyl radical to the α -meso carbon (<1 kcal mol⁻¹) is much lower in energy than the concerted pathway, by more than 25 kcal mol⁻¹. Furthermore, strong hydrogen bonding and electrostatic interaction of the heme with the hydroxyl radical were found to be positioned above the α -meso carbon, which is important for the regioselective oxidation.



Scheme 3.2 Reaction mechanisms for the first-step heme degradation in HO studied by the DFT and QM/MM calculations [56].

To examine the feasibility of the proposed acid-assisted concerted electrophilic hydroxylation [pathway (c) in Scheme 3.2], Shaik and coworkers subsequently included a protonated water cluster, $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$, $n = 0-2$, and $\text{NH}_4^+(\text{H}_2\text{O})_2$, in the proposed concerted and stepwise pathways [56c]. Although the introduction of general acid catalysts lowers the reaction barriers for the concerted pathway to about

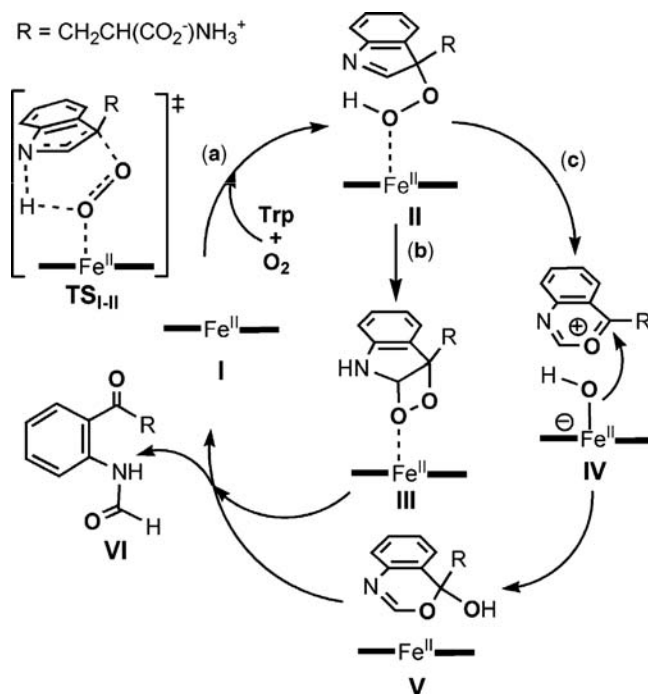
25.6–31.6 kcal mol⁻¹, they were calculated to be still higher in energy than the stepwise pathways via acid-assisted homolysis of the O–O bond, by 8–15 kcal mol⁻¹. By using H₃O⁺(H₂O)₂ as the acid catalyst, the calculated solvent kinetic isotope effect (SKIE) for the stepwise pathway (2.2) is quite similar to the measured value (2.3). In contrast, all concerted pathways gave large deviation between the calculated and measured SKIE, which further supports the stepwise pathway. However, neither pathway can reproduce the measured large inverse secondary α -deuterium kinetic isotope effect (*sec*-KIE) at the α -meso position. This discrepancy between the theory and experiment was rationalized by the absence of the electric field of the protein, electrostatic interaction with the protein, or steric constraints by the pocket in gas-phase calculations [56c].

Alternatively, Yoshizawa and coworker have proposed the concerted water-assisted oxo mechanism [pathway (d) in Scheme 3.2], in which the oxo ligand abstracts a hydrogen atom of one water molecule concerted with the radical addition of the resultant hydroxyl radical [56d]. The calculated barrier was reduced to be about 14 kcal mol⁻¹ by the B3LYP calculations, owing to the lesser deformation of heme in this concerted transition state. ONIOM(B3LYP:AMBER) calculations were also performed to study a possible route for the acid-catalyzed formation of the ferryl-oxo intermediate (compound I) from the ferric-hydroperoxy intermediate in the protein. Unfortunately, the reaction barrier for the above-mentioned step and subsequent concerted water-assisted oxo pathway were not explored by the ONIOM calculations [56d]. Additionally, a recent experiment ruled out the possibility of compound I in HO [57].

Very recently, Shaik and coworkers performed QM/MM(B3LYP/CHARMM) calculations to study the reaction mechanism in the protein [56e]. Again, the stepwise mechanism initiated with the homolytic O–O bond cleavage is more favorable, and has a barrier height of roughly 18–20 kcal mol⁻¹, depending on the snapshot chosen. Interestingly, the resultant hydroxyl radical species, which forms a very strong hydrogen bond with the water cluster in the distal pocket (not shown in Scheme 3.2), is not stable as “minimum” and can further undergo radical addition to the α -meso carbon without barrier in the protein. The barrierless radical addition of the transient hydroxyl radical and hydrogen bonding with the water cluster were shown to be essential for the stereospecific oxidation suppressing other side-reactions. The overall reaction mechanism calculated in the proteins can be regarded as a concerted but asynchronous mechanism, which may be more consistent with the experimental deduction based on the solvent kinetic isotope effect studies.

3.3.1.3 Indoleamines Dioxygenase (IDO) and Tryptophan Dioxygenase (TDO)

Indoleamine 2,3-dioxygenase (IDO) and tryptophan 2,3-dioxygenase (TDO) catalyze oxidative cleavage of the pyrrole ring of L-tryptophan (L-Trp) and insert both oxygen atoms of a dioxygen molecule into the organic substrate to afford *N*-formylkynurenine (VI in Scheme 3.3) [1c, 58]. This is the first and rate-limiting step in L-Trp metabolism through the kynurenine pathway. IDO and TDO are associated with a few important physiological roles, such as suppression of T-cell proliferation [1c, 59]. In sharp contrast with other heme-containing monooxygenases, and non-heme



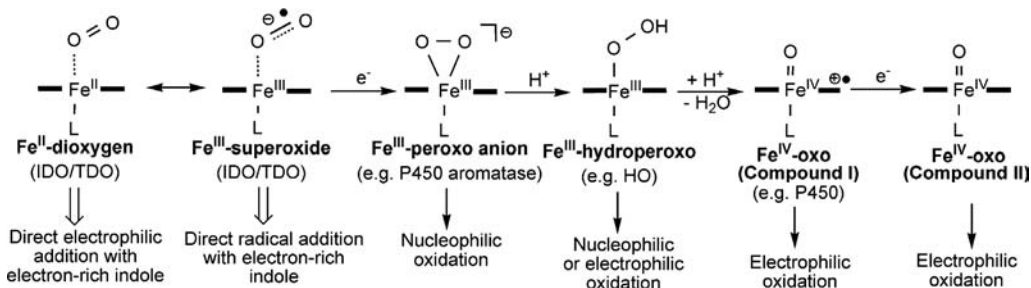
Scheme 3.3 Recently proposed reaction mechanism for IDO and TDO, c.f. [60–61].

catechol dioxygenases and naphthalene dioxygenase, the reaction mechanism in IDO and TDO has remained vague, even though they were discovered more than 40 years ago. This is mainly due to the lack of structural determinations for IDO or TDO and of observation of proposed intermediates, and that no similar reaction has been observed in other heme-containing oxygenases. Therefore, IDO and TDO were regarded as a missing piece in our understanding of the heme chemistry.

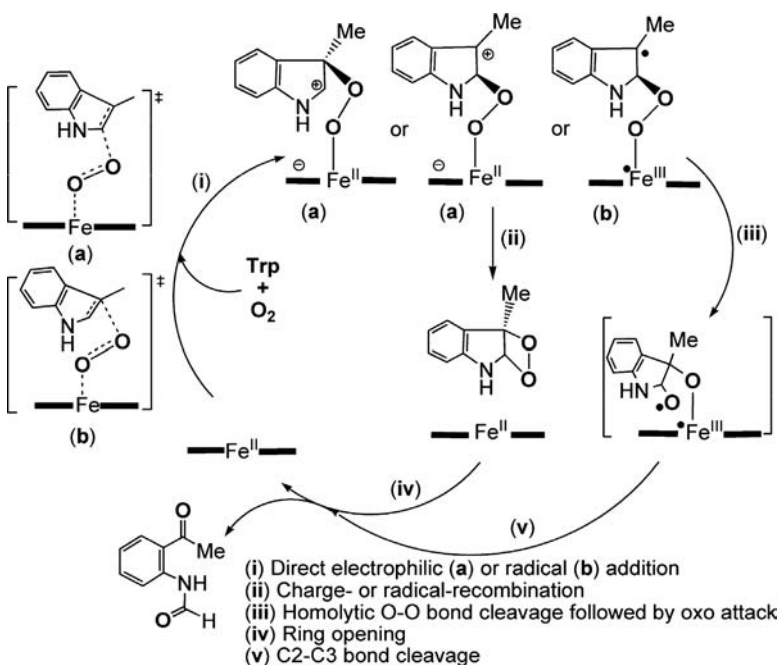
Recently, though, three groups obtained X-ray crystal structures of IDO and TDO [60]. Structural analysis and site-directed mutagenesis studies showed that polar residues in the active site are not important in the enzymatic reaction, and that a hydrogen-bond network is absent in the distal pocket. These features are distinct from other heme-containing monooxygenases or non-heme dioxygenases. Two new mechanistic pathways initiating with electrophilic addition of Fe-(II) bound dioxygen coupled with proton transfer (concerted oxygen ene-type reaction) followed by either formation of a dioxetane intermediate (III in Scheme 3.3) or Criegee-type rearrangement (IV in Scheme 3.3) were accordingly proposed.

However, we were suspicious of the proposed distorted concerted transition state for IDO and TDO. Indeed, our B3LYP calculations showed a very high reaction barrier (27–40 kcal mol⁻¹, relative to the precursor complexes) for the proposed concerted oxygen ene-type transition state [61]. Instead, we found two new energetically favorable dioxygen activation mechanisms in heme systems as the first step in IDO and TDO: (i) direct electrophilic addition of the Fe (ii) -dioxygen to the C2 or C3

position of the indole and (iii) direct radical addition of the Fe(III)-superoxide to the C2 position of the indole (Figure 3.5 and Schemes 3.4 and 3.5).



Scheme 3.4 Dioxygen activation process for IDO/TDO and the other heme systems.



Scheme 3.5 Proposed alternative mechanisms for IDO and TDO in the gas phase.

On the other hand, attempts to locate transition states for the proposed Criegee-type rearrangement pathway from the neutral hydroperoxy intermediate of IDO and TDO were unsuccessful. However, the resultant rearrangement products were calculated to be highly endothermic ($>70 \text{ kcal mol}^{-1}$), due to the unfavorable charge separation (IV in Scheme 3.3). Notably, such rearrangement proceeds from the formally anionic intermediates in non-heme catechol dioxygenases, in which the presence of the anionic or radical oxy is the key to lowering the barrier [61, 62]. Therefore,

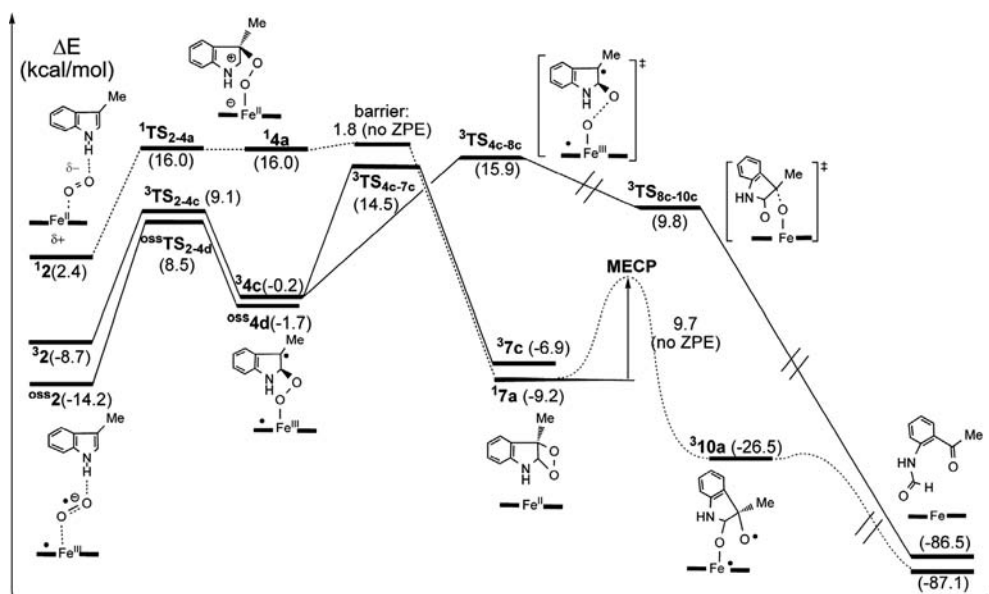
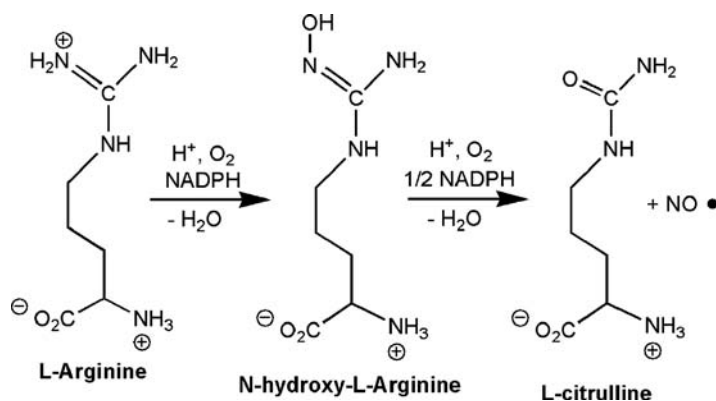


Figure 3.5 Potential energy surface for the proposed alternative mechanistic pathway for IDO and TDO in the gas phase.

these calculations deny the Criegee-type rearrangement pathway from the neutral ferric-hydroperoxy intermediate in IDO and TDO. Instead, the gas-phase calculations suggested that the resultant diradical or zwitterionic intermediates undergo radical recombination or nearly-barrierless charge recombination to give metastable dioxetane intermediates, followed by ring-opening to give the product (Figure 3.5 and Scheme 3.5). Interestingly, the formation of dioxetane intermediates, which were questioned due to the ring strain, was calculated as unstable as the neutral ferric-hydroperoxy intermediate (II in Scheme 3.3). Instead, homolytic O–O bond cleavage from the diradical intermediate followed by oxo attack and facile C2–C3 bond cleavage was found to compete with the dioxetane formation pathway (Figure 3.5 and Scheme 3.5). Notably, we do not rule out the possibility for formation of the corresponding epoxide along with ferryl-oxo intermediates, which has not been considered before [1c, 60, 63], after our proposed direct pathways in the proteins, although this pathway was calculated to be slightly higher in energy than the above two pathways [61]. In addition, electron transfer from the indole moiety to the heme during the oxidation is also possible. Our ONIOM(B3LYP:AMBER) calculations suggested that our proposed direct radical addition of the Fe(III)-superoxide to the C2 position of the Trp is the favorable pathway in the bacterial TDO, while the addition to the C3 position was blocked by the steric hindrance. The mechanism we found is sharply different from the proposed mechanisms in Scheme 3.3 [1c, 60, 63]. Our ONIOM calculations on the reaction mechanism in bacterial TDO and its protein effect will be reported soon.

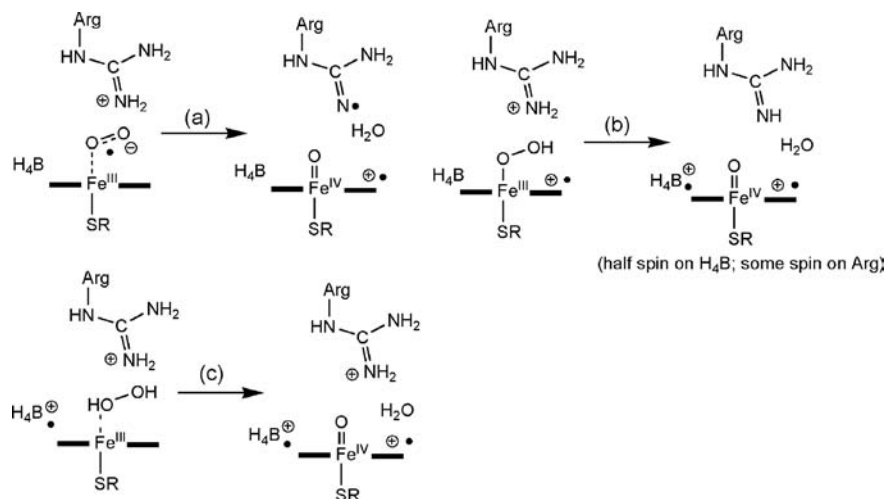
3.3.1.4 Nitric Oxide Synthase (NOS)

Nitric oxide synthase (NOS) catalyzes the two-step oxidation of L-arginine to form L-citrulline and nitric oxide via N^ω-hydroxy-L-arginine (Scheme 3.6) [1c,d,g]. Reactive NO can act as neurotransmitter and also influence blood pressure as well as heart rate [64]. The oxidation of L-arginine requires one oxygen molecule and two electrons in the first half-reaction, which is regarded as a P450-like reaction. Oxidation of N^ω-hydroxy-L-arginine to generate nitric oxide demands one oxygen molecule and one electron in the second half-reaction, which has no precedence. Both reactions take place in the presence of a tetrahydrobiopterin (H₄B) cofactor, which can act as electron donor and is then converted into the cationic radical. It is kinetically coupled with the decay of a ferric-superoxo (Fe^{III}-O₂⁻) intermediate [65]. Electron transfer from H₄B was proposed to give a ferric-peroxo anion (Fe^{III}-O₂²⁻), but Fe^{III}-O₂²⁻ was detected only under cryogenic conditions [66]. Since the reactions are very fast and no other intermediate has been observed, the detailed reaction mechanisms are unclear. Several QM and QM/MM calculations have been carried out to understand these important and unique oxidations steps [67].



Scheme 3.6 Formation of nitric oxide in nitric oxide synthase (NOS).

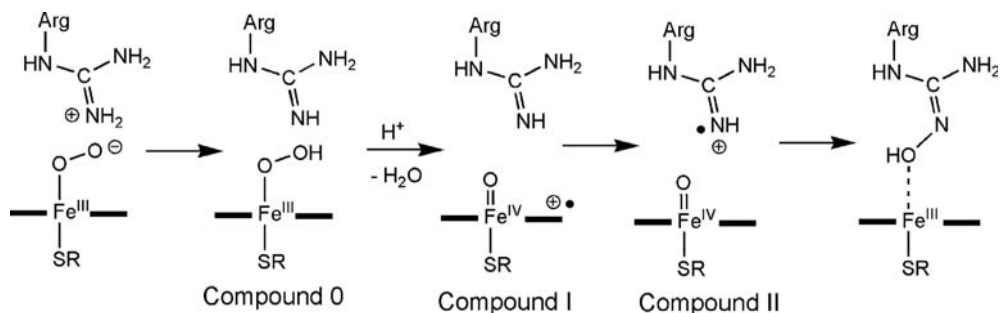
For the first half-reaction, Shaik and coworkers have performed QM/MM(B3LYP/CHARMM) calculations to study the mechanism to afford the proposed oxidant, compound I (Scheme 3.7) [67d]. First, the formation of the ferryl-oxo intermediate, compound I, via route (a) in Scheme 3.7 was calculated to be endothermic by 23–32 kcal mol⁻¹, in which one proton and one hydrogen atom are transferred from the protonated arginine substrate to give one water molecule. Electron transfer from H₄B was also not found in the reactant and product. Therefore, this reaction pathway was ruled out. Second, when one external proton is transferred to the oxy-heme intermediate, the computed reaction energy to give compound I was endothermic by about 12–14 kcal mol⁻¹ for the singlet, triplet and quintet states [route (b) in Scheme 3.7]. About half spin from H₄B was found to be transferred to compound I. Third, when two external protons are transferred to give the ferric hydrogen peroxide intermediate, electron transfer from H₄B to the heme was observed [route (c) in



Scheme 3.7 Three pathways leading to compound I in the first half-reaction of NOS studied by QM/MM calculations [67d].

Scheme 3.7]. Only this pathway was found to be exothermic. One new mechanism via homolytic O–O bond cleavage coupled with internal electron transfer followed by abstraction of the proximal hydrogen was found in this energetically favorable pathway, with a barrier of about 8 kcal mol^{−1}, and was considered as the most possible pathway. These QM/MM calculations also showed that, different from the heme systems, a ferric-hydroperoxide intermediate (so-called compound 0) was not present in NOS.

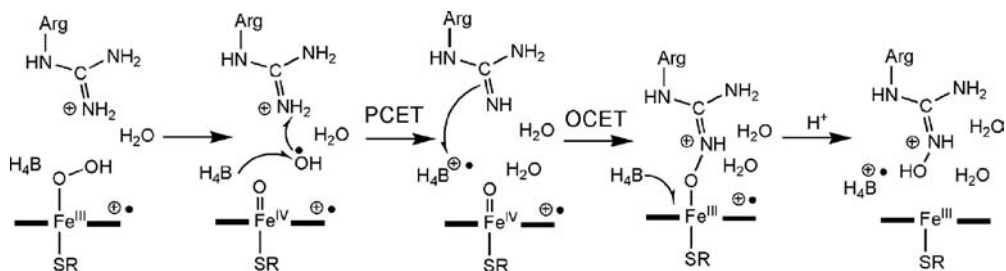
Recently, the reaction mechanism of the first half-reaction, hydroxylation of arginine, of NOS were studied by B3LYP and QM/MM(B3LYP:CHARMM) by two groups [67e,f]. The B3LYP calculations showed that hydrogen atom transfer from the protonated arginine to the oxo ligand of compound I has a barrier of about 17–18 kcal mol^{−1} [67f]. This relatively large barrier for hydrogen atom transfer of electron-deficient protonated arginine in NOS is in contrast with the small barrier in P450 with neutral substrates. It is followed by the rate-determining radical rebound transition state with an overall barrier of about 21–24 kcal mol^{−1}. The overall reaction was calculated to be slightly exothermic. After all, this mechanism was ruled out, due to relatively high barriers and small reaction exothermicity. Alternatively, another mechanism via hydroxylation of the neutral arginine was investigated. Electron transfer from the neutral arginine to the heme was found to take place to give compound II and the arginine cationic radical (Scheme 3.8). Hydrogen atom transfer from the arginine cationic radical to compound II followed by a barrierless concerted rebound process in the doublet state (Scheme 3.8) has a much lower barrier (~5 kcal mol^{−1}). In addition, the reaction was computed to be highly exothermic, by about 23 kcal mol^{−1}. Therefore, the DFT calculations suggest compound II as the actual oxidant in NOS, which is sharply different from P450 [67f]. Based on these results, the overall reaction mechanism was also proposed (Scheme 3.8): the



Scheme 3.8 Proposed mechanism for the first half-reaction of NOS based on B3LYP calculations [67f].

protonated arginine binds in the active site and donates one proton to give compound 0, then a proton is externally transferred from the bulk phase to afford compound I. Electron transfer proceeds from the neutral arginine to compound I to generate compound II, followed by hydroxylation steps.

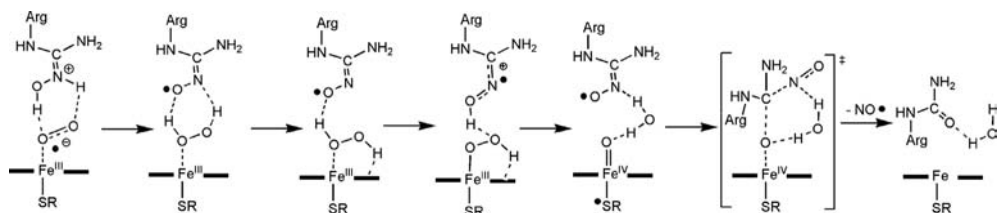
Meanwhile, Shaik and coworkers have also examined the reaction mechanism for the first half-reaction of NOS by the QM/MM(B3LYP/CHARMM) method, including the NOS protein as well as larger QM models with the key H₄B cofactor [67e]. Although the reaction starting from the ferric hydrogen peroxide intermediate, Fe^{III}(HOOH), was calculated to be thermodynamically more favorable to give the proposed oxidant compound I [route (c) in Scheme 3.7] in their previous calculations [67d], the reaction barrier of the subsequent proton transfer from the protonated arginine to compound I was more than 30 kcal mol⁻¹ [67e]. In addition, the resultant “product,” protonated compound I, was found to be unstable and converted into compound I. Moreover, a direct N-oxygenation step of the protonated arginine by compound I was calculated to be a very high-energy process (~44 kcal mol⁻¹). Therefore, an oxidation reaction involving the protonated arginine and compound I was concluded to be infeasible [67e]. Another mechanism was then considered. Instead of compound 0 (Fe^{III}-OOH) employed by the active-site model calculations [67f], the reaction mechanism starting from a Por⁺•Fe^{III}-OOH intermediate was considered [67e]. The O–O bond cleavage yielding compound I and hydroxyl radical species has a barrier of about 20 kcal mol⁻¹ in the alternative pathway (Scheme 3.9).



Scheme 3.9 Proposed mechanism for the first half-reaction of NOS based on QM/MM calculations [67e].

Spin density showed that about half an electron was transferred from H₄B to give a “half-radical/half an ion” OH species, which promotes the subsequent proton-coupled electron-transfer (PCET, or so-called hydrogen atom transfer) driven by full electron transfer from H₄B. An estimated barrier height of the PCET step is roughly 19 kcal mol⁻¹. When the H₄B cofactor was not treated by the QM method to suppress electron transfer, the reaction barrier increased to 27 kcal mol⁻¹ and the resultant compound I and hydroxyl species were destabilized by about 8 kcal mol⁻¹. The third-step oxygenation of the neutral arginine by compound I occurs via a so-called oxygen-coupled electron-transfer (OCET), in which some amount of electron of the neutral arginine was found to be transferred to the H₄B cation radical. The reaction barrier of this oxygenation step is rather high (23 kcal mol⁻¹; 16 kcal mol⁻¹ with the approximate thermal effects). On the other hand, suppressing the electron transfer from the H₄B cofactor slightly raised the barrier by about 1 kcal mol⁻¹ [67e], and electron transfer from the neutral arginine to compound I to give compound II occurs [67f]. Therefore, the QM/MM calculations demonstrated that the H₄B cofactor plays an important role as electron shutter, especially prompting PCET. However, the QM/MM calculations did not support the formation of compound II after PCET step in NOS [67e], which was suggested by the active-site model calculations [67f]. The alternative mechanism via compound II may be operative when electron transfer is given from some other electron donor, rather than the H₄B cofactor. Both the DFT and QM/MM calculations supported the view that one proton transferred from the bulk phase and one proton from the protonated arginine are required in the first half-reaction of NOS.

The reaction mechanism of the unique second half-reaction of NOS was theoretically investigated by the active-site model calculations and B3LYP method [67a,c]. First, Gauld and coworker studied several possible binding models of the protonated N^ω-hydroxy-L-arginine (NHA) with and without dioxygen in NOS [67a]. The favorable binding mode involves two hydrogen bonds of NHA with the bound dioxygen (Scheme 3.10). However, the proposed tetrahedral intermediate (Fe–O–O–C_{guan}) derived from the attack of distal oxygen to the guanidinium carbon was calculated to be highly unstable (29–45 kcal mol⁻¹) in the subsequent DFT study [67b]. Instead, one proton and one hydrogen transfer from NHA to give a slightly less stable ferric hydrogen peroxide intermediate (Scheme 3.10) was found to take place with a modest barrier of about 15 kcal mol⁻¹. A low-barrier of O–O bond rotation occurs to give another form of ferric hydrogen peroxide intermediate, with one hydrogen bond



Scheme 3.10 Proposed mechanism for the second half-reaction of NOS based on B3LYP calculations.

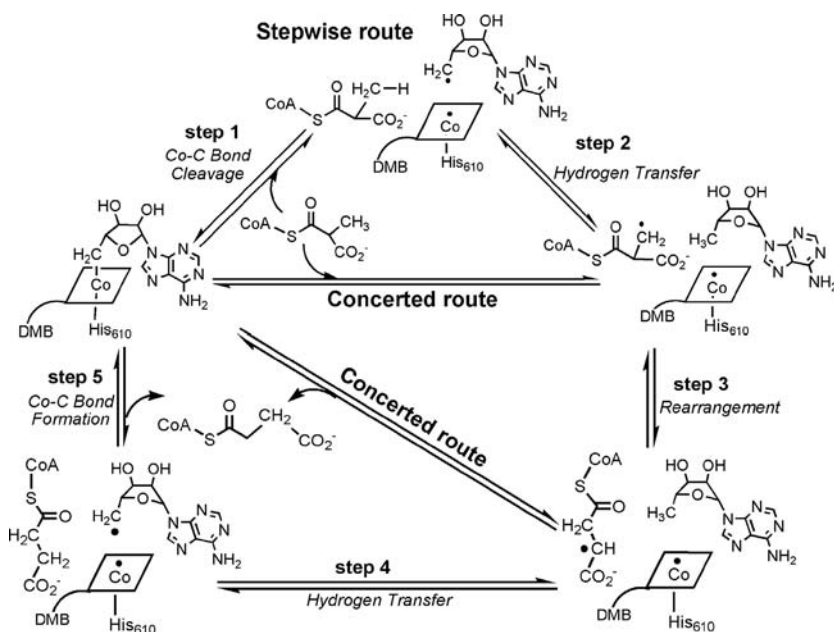
between the distal OH and a heme nitrogen [67c]. Then, similar to “ping-pong” mechanism in heme peroxidases [1c], the NO group of the substrate acts as a base to facilitate proton migration from the proximal oxygen to the distal oxygen to give stable compound I and a water molecule (i.e., heterolytic O—O bond cleavage). The overall barrier of this step is about 20 kcal mol^{-1} . Finally, the oxo ligand of compound I was shown to attack the guanidinium carbon via the concerted tetrahedral transition state to generate NO radical and citrulline with a barrier of about 18 kcal mol^{-1} , instead of the proposed stepwise pathway via the tetrahedral intermediate [67c].

3.3.2

Cobalamin-Dependent Enzymes

3.3.2.1 Methylmalonyl-CoA Mutase

Methylmalonyl-CoA mutase (MMCM) is an important member of B_{12} -dependent enzymes that catalyzes a radical-based transformation of methylmalonyl-CoA into succinyl-CoA (Scheme 3.11) [2,68–70]. A unique feature of this enzyme is the formation of a carbon-centered 5'-deoxyadenosyl (Ado) radical and a five-coordinated cob(II)alamin radical generated from homolytic cleavage of the Co—C bond of the adenosylcobalamin (AdoCbl) coenzyme. The Ado radical then abstracts a hydrogen atom from the methylmalonyl-CoA substrate to generate a key intermediate, substrate radical; a 1,2-rearrangement step can proceed only if the substrate radical is formed. Finally, the hydrogen transfer from the Ado moiety to the product radical



Scheme 3.11 Proposed reaction pathways in the rearrangement of methylmalonyl-CoA to succinyl-CoA in MMCM.

followed by reformation of the Co–C bond of the AdoCbl completes the full catalytic cycle (Scheme 3.11).

Remarkably, the first-step Co–C bond homolytic cleavage in B₁₂-dependent enzymes can be accelerated by a factor of about 10¹², compared to that of the coenzyme in the aqueous solution (BDE = ~30 kcal mol⁻¹, $K = \sim 8 \times 10^{-18}$ M, $k = \sim 10^{-9}$ s⁻¹) [68, 69, 71]. The origin of this considerable catalytic enhancement is of significant interest, and has been extensively investigated. However, it is difficult to quantify experimentally the energetic contributions from different factors and, consequently, the catalytic effect in B₁₂-dependent enzymes was under debate. In contrast, homolytic bond cleavage is found to be kinetically coupled to hydrogen transfer from the substrate in these enzymes [68, 69, 72]. In addition, the exceptionally large deuterium KIE effect (~50 at 5 °C) was measured in MMCM. The hydrogen transfer process is still poorly understood, although the stepwise route is generally believed (Scheme 3.11). However, the concerted route has been recently proposed [2e,f, 72b]. Several theoretical groups have performed DFT and QM/MM calculations to investigate the origin of large catalytic enhancements in the initial Co–C bond cleavage (steps 1 or/and 2) [73, 74]. The reaction mechanism for the 1,2-rearrangement of the substrate (step 3) was also studied by DFT and QM/MM calculations [75]. Morokuma, Paneth and coworkers studied the large protein effect on the Co–C bond dissociation by using the active-site model as well as including key parts of both the unreactive form and reactive form of MMCM [74]. The ONIOM (UBP86:MM) calculations showed that the Co–C bond dissociation energy is significantly reduced to 2.5 kcal mol⁻¹ in the reactive form (substrate-bound state) of the MMCM, compared to the coenzyme in the gas phase and in the unreactive form (substrate-free state). In addition, the Co–C bond dissociation transition state was located in the reactive form for the first time by the new optimization algorithm [34h]. The reaction barrier for the dissociation is largely reduced to 10.1 kcal mol⁻¹ in the protein. Moreover, the conformational change of the Ado was also found to take place in the transition state in the protein.

Very recently, we performed systematic calculations and analysis on the Co–C bond cleavage, and studied the feasibility of a newly proposed concerted mechanism for the bond cleavage with hydrogen transfer by using nearly the entire MMCM protein (the reactive form) with several larger QM models [76]. Significant protein effects on the reaction barrier (~15.3–17.3 kcal mol⁻¹) and reaction energy (~9.8–12.1 kcal mol⁻¹) for the Co–C bond cleavage were again found, compared to the gas phase. The protein effects were decomposed into several factors: the cage effect, the reactant destabilization and the protein MM effect. First, for the cage effect, the Ado radical does not completely dissociate from the active-site cob(II)alamin, and thus still has interaction with the cob(II)alamin in the dissociated state of MMCM (~5 kcal mol⁻¹). Second, reactant destabilization comes from the strain destabilization of the coenzyme, principally from the Ado moiety, in the bound state R more than the bond cleaved state I2 (~6–8 kcal mol⁻¹). Third, the protein MM effect consists mainly of the interaction of the active (QM) part of the coenzyme with the remainder (MM part) of the coenzyme as well as with its surrounding residues via Coulomb, van der Waals and torsion terms. Several residues, particularly Glu370 and

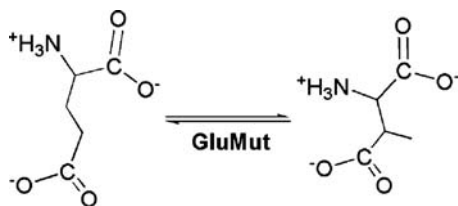
also Gln330, play essential roles in stabilizing the Co–C bond cleaved state, since the dissociated Ado radical approaches these nearby residues and forms stronger hydrogen bonds. The most critical residue, Glu370, promotes the Co–C bond cleavage by about 9–11 kcal mol^{−1} via very strong hydrogen bonding. The importance of Glu370 is in line with the recent empirical valence bond (EVB) simulations by Warshel and coworkers [73c], which, however, considered the strain effect unimportant. Our calculations indicate that the three above-mentioned factors cooperatively facilitate the Co–C bond cleavage in the protein. Ryde and coworker have also performed QM/MM calculations and found similar factors responsible for the considerable protein effects on the Co–C bond cleavage in another B₁₂-dependent glutamine mutase (GluMut) [73d]. Conformation of the flexible Ado in MMCM was also found to be finely tuned during the Co–C bond cleavage process in our study [76]. We also demonstrated that a larger QM part of the coenzyme, including the adenine base as well as the inclusion of almost all MMCM protein, in the calculation is quite important to describe reliable geometry and energy for the Co–C bond cleaved step, especially the highly strained bound state.

On the other hand, the newly proposed concerted mechanism for the Co–C bond cleavage and hydrogen transfer was examined and supported by gas-phase calculations for GluMut and the EVB simulations for MMCM (Scheme 3.11) [73b,c]. We also studied the feasibility of the concerted mechanism in the MMCM protein using the ONIOM(DFT:MM) method [76]. Similar to the DFT study on GluMut by Kozłowski and Yoshizawa [73b], our active site calculations showed that the concerted pathway for the truncated MMCM substrate without protein was calculated to be more favorably than the stepwise pathway by about 8 kcal mol^{−1} [76]. The concerted transition state is stabilized by its interaction with the cob(II)alamin radical, the sole source of extra stabilization in the gas phase. Such interaction is absent in the stepwise transition state in the gas phase. On the other hand, our ONIOM study with explicit consideration of protein demonstrated a large protein effect on this hydrogen transfer step [76]. For the stepwise hydrogen transfer in MMCM, the computed reaction barrier for hydrogen transfer is significantly reduced from 55 kcal mol^{−1} in the gas phase to about 19 kcal mol^{−1} due to very high stability of the bond cleaved state I2 and the substrate preorganization in the protein. In contrast, optimization of the assumed concerted transition state in the protein converged to the stepwise transition states in MMCM; the concerted pathway disappeared and the low-energy stepwise pathway is the only mechanism available in the protein. The proposed concerted transition state is not stabilized as much as the stepwise transition state, as in the former the Ado moiety loses some interaction with nearby residues (i.e., the protein MM effect) and interaction between Co and proximal histidine is weakened, when shortening the Co–C bond. Moreover, Ado and the substrate become more distorted in the assumed concerted transition state. Notably, sharply different from the gas phase, the reaction barrier for the first-step Co–C bond cleavage is considerably reduced and the bond cleaved state I2 is significantly stabilized by the protein. The concerted pathway might be possible, when the protein and substrate are flexible or/and the hydrogen transfer process is endowed with some significant driving force in the protein.

In this regards, Paneth and coworkers have performed QM/MM calculations (AM1 as the QM method and CHARMM22 as the MM method) combined with ensemble-averaged variational transition-state theory for the stepwise hydrogen transfer step [73a]. Multidimensional tunneling increases the magnitude of the computed intrinsic hydrogen KIE by a factor of 3.6, from 14 to 51, which is consistent with the exceptionally large measured KIE effect. Sharp corner-cutting, rather than an unusually thin barrier, was shown to shorten the tunneling path and thus increase tunneling probability of hydrogen transfer.

3.3.2.2 Glutamine Mutase

Another important B_{12} -dependent mutase is glutamate mutase (GluMut), which can catalyze the reversible and stereospecific equilibration of (*S*)-glutamate with (2*S*,3*S*)-3-methylaspartate through a radical mechanism (Scheme 3.12). Similar to MMCM, GluMut also tremendously accelerates the rate of homolysis of the C–C bond. Unlike MMCM, two bond-dissociated intermediates with the Co–C bond of 3.17 and 4.19 Å were observed to adopt different ribose conformations in X-ray crystal structures of GluMut [77]. Ryde's group employed a QM/MM method to show a very large catalytic origin in GluMut ($\sim 24 \text{ kcal mol}^{-1}$) [73d]. The calculated catalytic effect has been divided into four terms: the cage effect ($\sim 4.8 \text{ kcal mol}^{-1}$), the strain effect of the coenzyme ($\sim 14.6 \text{ kcal mol}^{-1}$), the protein MM effect ($\sim 10.0 \text{ kcal mol}^{-1}$, by electrostatic and van der Waals interactions) and the stabilized protein itself ($\sim 2.6 \text{ kcal mol}^{-1}$) in the dissociated state responsible for the catalytic effect in GluMut. Therefore, the catalytic effect was suggested not to be predominantly determined by electrostatic interactions, but by van der Waals interaction with the surrounding amino acids, a side-chain of the coenzyme, the substrate and two water molecules. In contrast with our calculations on MMCM, the conformational change of the Ado during the Co–C bond cleavage was not reported in GluMut.



Scheme 3.12 Conversion of glutamate into methylaspartate catalyzed by GluMut.

Recently, the concerted and stepwise pathways for the Co–C bond homolysis and hydrogen transfer were theoretically examined for the first time by using a truncated model of the GluMut substrate in the gas phase by Kozłowski and Yoshizawa [73b]. They showed that the concerted route is lower in energy than the stepwise route by about $7.0 \text{ kcal mol}^{-1}$, in which the cob(II)alamin acts as a conductor to stabilize the concerted route. However, as shown in our calculations in MMCM [76], the proposed concerted pathway might be less favorable if the effects of GluMut protein are explicitly taken into account.

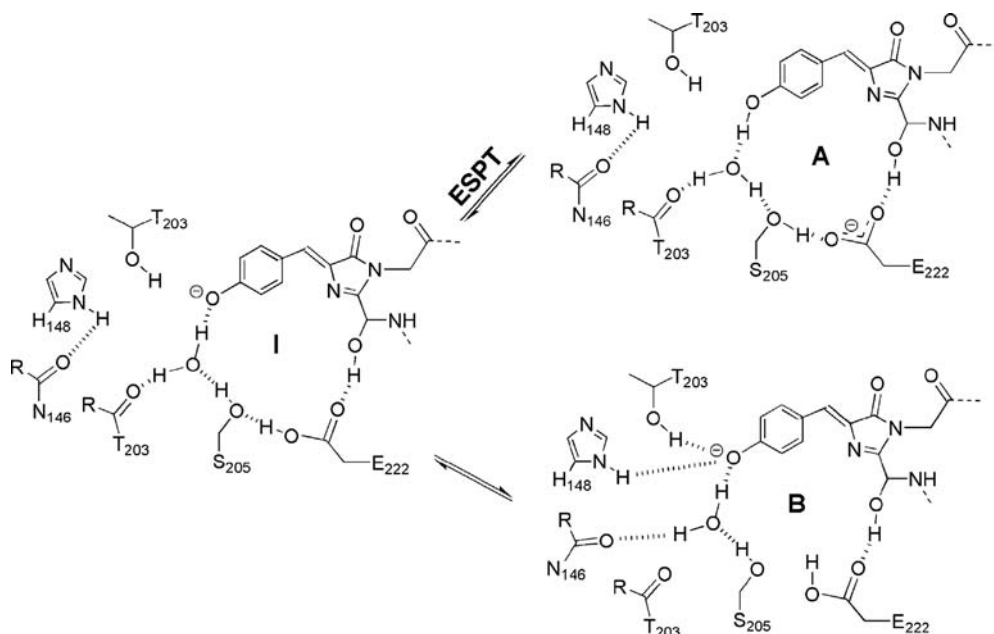
3.4

Photobiology

3.4.1

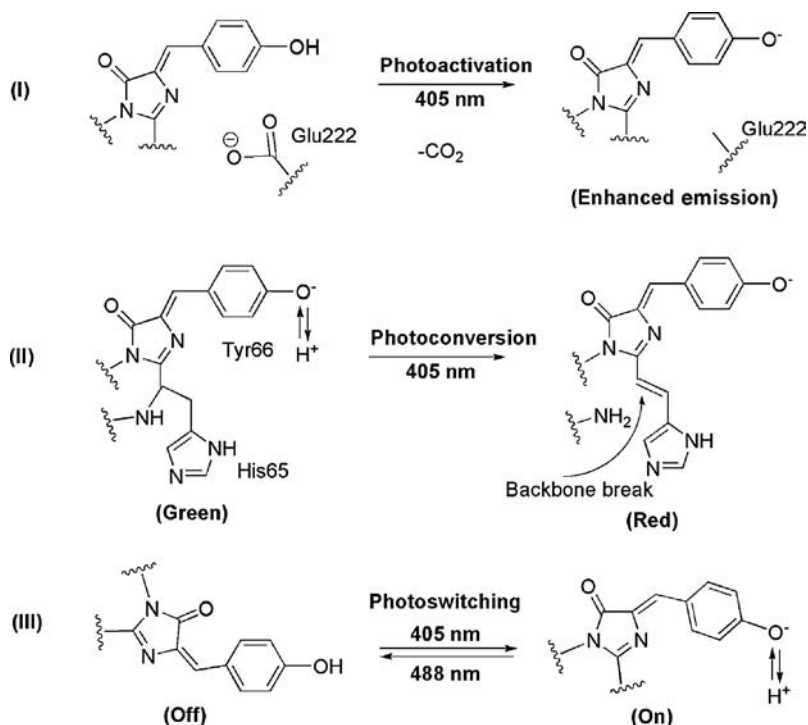
Fluorescent Proteins (FPs)

Green fluorescent protein (GFP), which was discovered from the jellyfish *Aequorea victoria* [78], and its variants have become one of the most widely studied and exploited proteins in biochemistry and cell biology, particularly for biological imaging and analysis [6, 79]. The three-state model has been proposed for photoisomerization of wild-type GFP (Scheme 3.13). The GFP proton wire operating upon photoexcitation of the internally caged chromophore, that is, excited-state proton transfer (ESPT), has been experimentally shown to be essential for the fluorescence emission [80].



Scheme 3.13 Three-state model proposed for the photoisomerization of GFP.

Furthermore, a new class of fluorescent proteins, photoactivatable fluorescent proteins (PAFPs), have recently been developed, in which photophysical properties of the chromophore can be dramatically altered by illumination [81–85]. In general, PAFPs can be categorized into three classes, based on respective photoactivation mechanisms (Scheme 3.14). The first class is irreversible photoactivation (such as PA-GFP and PS-GFP), in which photoactivation of the chromophore significantly enhances intensity of the fluorescence by excitation with ultraviolet (UV) to violet light. The second class is irreversible photoconversion (such as Kaede, EosFP and



Scheme 3.14 Three types of photoactivation mechanisms.

IrisFP), in which the color of the fluorescence can be irreversibly switched from green to red on illumination with UV to violet light. The third class is reversible photo-switching between fluorescent and non-fluorescent states by using appropriate illumination wavelengths (such as Dronpa, Padron, asFP595, IrisFP and mTFP0.7). Such rapid progress and development in PAFP advances fluorescent protein technologies and could be potentially used for nano applications, such as molecular switch and data storage [86].

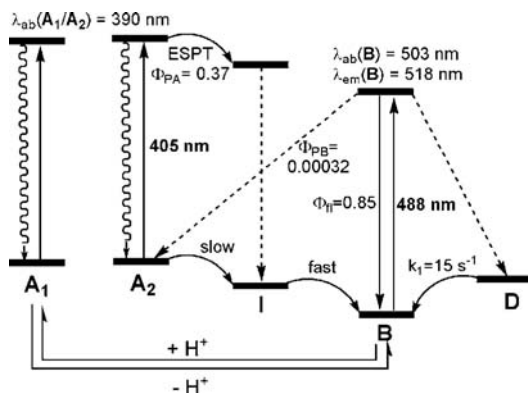
3.4.1.1 Green Fluorescent Proteins (GFP)

Based on the X-ray crystal structures [80a] and ultrafast transient infrared spectroscopy [80e], the neutral chromophore in GFP (so-called A form) was proposed to be excited and undergo ESPT along the hydrogen-bond wire to Glu222 via one water molecule and Ser205 (Scheme 3.13). A few theoretical studies have also been devoted to understanding the complex photoactivation process in GFP [87]. CASPT2 calculations, performed recently by adopting a simplified QM model for GFP, suggested the importance of the photoactive π, π^* state for ESPT [87a], rather than the charge-transfer state π, σ^* . The π, σ^* state with some Rydberg character in the σ^* orbital was found to be important for excited-state hydrogen-transfer in other systems [88]. The most energetically favorable proton-transfer pathway via the π, π^* state was suggested to initiate with proton transfer from serine to glutamate residue, then a proton transfer from water to serine, and finally a proton from the chromophore to the water

(i.e., stepwise pathway) in a static CASPT2 study [87a]. However, a roughly concerted model was calculated to be only 4 kcal mol^{-1} higher in energy than the stepwise pathway. An essentially concerted, synchronous and fast excited-state proton transfer and at least two dynamical regimes were observed in the subsequent quantum dynamics simulation [87b].

3.4.1.2 Reversible Photoswitching Fluorescent Proteins (RFPs)

Reversibly photoswitching fluorescent proteins (RFPs) are a new class of fluorescent proteins, in which fluorescent on-state and non-fluorescent off-state of RFPs can be reversibly switched by irradiation of two different radiations (Scheme 3.15) [81, 84]. Miyawaki and coworkers discovered one of the most promising RFPs, Dronpa, which was engineered from a coral Pectiniidae [81a]. Dronpa successfully tracked the protein dynamics *in vivo* (nucleocytoplasmic shuttling of signaling proteins). Two mutants of Dronpa (Dronpa-2 and Dronpa-3) endowed with faster response to light and faster thermal relaxation from the “off” state to the “on” state have also been developed [89, 90].

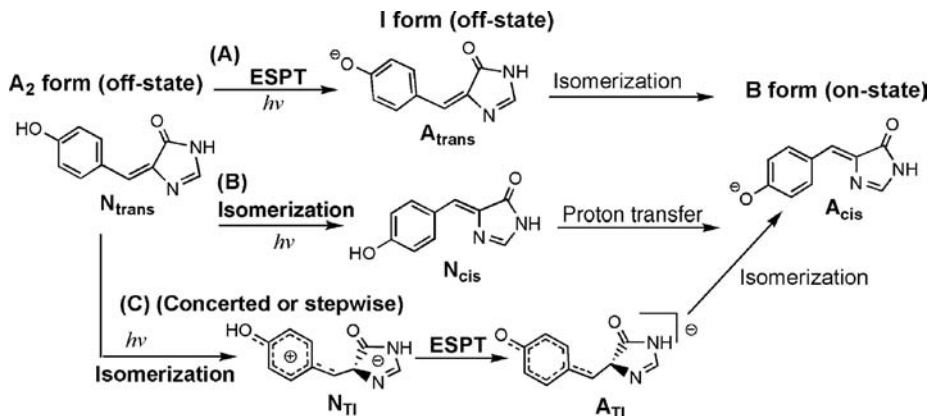


Scheme 3.15 Proposed mechanism for the reversible photoswitching in Dronpa.

On the basis of photophysical properties of Dronpa determined by single-molecule spectroscopy, the reaction mechanism of reversible photoswitching of Dronpa was proposed as shown in Scheme 3.15 [81]. ESPT was proposed to proceed from the neutral non-fluorescent A_2 form to give an anionic non-fluorescent intermediate I, assumedly in an unrelaxed protein environment, and eventually giving the anionic fluorescent B form (Scheme 3.15). This reaction mechanism is analogous to the three-state photoisomerization model for wild-type GFP [6b, 80]. The ESPT in Dronpa was supported by the kinetic deuterium isotope effect (KIE ~ 2) [81e]. The unknown non-fluorescent metastable D form was also proposed to account for the dynamic behavior of Dronpa [81].

Recently, X-ray crystal structures of the on- and off-states of Dronpa were obtained by different groups [91]. The chromophore is formed by posttranslational modification from the Cys62-Tyr63-Gly64 (CYG) tripeptides. The chromophore adopts a cis, coplanar conformation in the on-state crystal structure, while it was suggested to be in a trans, non-planar conformation in the off-state crystal structure [91]. In addition, the

local immediate environment around the chromophore was suggested to influence protonation states of the chromophore [91d]. As a result, the reaction mechanism involving cis–trans isomerization of the chromophore was proposed to dictate the protonation state and, in turn, the on/off-states, rather than the reaction mechanism initiated with ESPT (mechanisms A and B in Scheme 3.16).



Scheme 3.16 Proposed reaction mechanisms in Dronpa.

However, the detailed reaction mechanism of reversibly photoswitching in Dronpa at the atomic level remains vague. Characterizing the nature of experimentally observed on- and off-states is of great importance in understanding the reaction mechanism of the photoswitching process and designing better molecular photo-switches. Assignment of the correct protonation state of the chromophore is challenging, as different protonation states have been proposed to be responsible for wild-type GFP. To understand the mechanism of the reversible photoswitching process (Scheme 3.15), we performed QM and ONIOM(QM:MM) calculations to study the nature of the proposed “on” and “off” states [92]. Several high-level QM methods (TD-B3LYP, CASSCF, CASPT2 and SAC-CI) were employed to compute the vertical absorption and emission energies in four different protonation states [i.e., anionic (A), zwitterionic (Z), neutral (N) and cationic (C)] in two conformations (i.e., cis and trans) in the gas phase. The vertical absorption and emission energies of the on- and off-states in proteins were further studied by the ONIOM(QM:MM) calculations (Table 3.1).

Table 3.1 Calculated vertical absorption energies (eV) of the different forms of the Dronpa chromophore calculated by the ONIOM(SAC-CI(Level2)/D95 (d):AMBER)-EE method at ONIOM(B3LYP/6-31 + G(d,p):AMBER)-EE optimized ground state structures.

Absorption	Anionic (A)	Zwitterionic (Z)	Neutral (N)	Cationic (C)	Exptl
Cis (on-state)	2.36	2.21	3.09	2.42	2.46
Trans (off-state)	2.24	2.73	3.01	2.36	3.18

The calculated absorption energy of the chromophore by the TD-B3YP method was found to have a large error for the anionic form (with some character of charge-transfer state) and be very similar in energy among different protonation states. Therefore, the TD-B3LYP method was showed to be unreliable to determine the protonation states in Dronpa. In contrast, as shown in Tables 3.1 and 3.2, more elaborate ONIOM(SAC-CI:MM) calculations on vertical absorption and emission energies in proteins supported A_{cis} and N_{trans} forms as the dominant protonation states of the chromophore in the on- and off-states of Dronpa, respectively. A_{cis} was further supported by the ONIOM(CASPT2:MM) calculations (Table 3.2). Unless including four states, the ONIOM(CASPT2:MM) calculations gives a large error for the neutral forms (e.g., N_{trans}). The feasibility of multiconfigurational Z_{cis} , which was suggested to be involved in another RFPF asFP595, is not supported by multiconfiguration CASPT2 calculations. Although the absorption energy for C_{cis} is similar to the experimental value for the on-state Dronpa, it is inconsistent with the experimental pH conditions. These calculations also support the cis–trans isomerization of the chromophore along with the change of the protonation state as one of the feasible pathways (mechanism B in Scheme 3.16).

The protonation state in the on- and off-states Dronpa was further examined by solving linear Poisson–Boltzmann equations (PBEs) via the program MEAD [42a,b] and by sampling the ensemble of protonation patterns by a Monte Carlo (MC) algorithm via the program Karlsberg [42i]. As shown in Table 3.3, the trans chromophore was calculated to be neutral (i.e., N_{trans}) in the off-state Dronpa, while the cis chromophore was found to be essentially anionic (A_{cis}), but not zwitterionic (Z_{cis}), in the on-state Dronpa. As a result, these PBE calculations supported that the local protein environment and trans–cis isomerization of the chromophore modulate the protonation state of the chromophore (mechanism B in Scheme 3.16).

To qualitatively estimate the effect of electrostatic interaction with nearby key residues (Glu144, Ser142, His193 and Glu211) on the protonation state in the on-state Dronpa, we further performed the Poisson–Boltzmann electrostatic calcula-

Table 3.2 Absorption and emission energies of A_{cis} , Z_{cis} and N_{trans} in Dronpa proteins calculated by the ONIOM(CASPT2(14e,13o)/6-31G(d):AMBER)-EE and ONIOM(SAC-CI(Level2)/D95 (d):AMBER)-EE methods at the ONIOM (CASSCF (14e,13o)/6-31G(d):AMBER)-ME optimized ground state (absorption) and excited state (emission) structures (cf. ref. 92).

Method	CASPT2	SAC-CI	Exptl.
Absorption energy (eV)			
A_{cis}	2.71	2.42	2.46
Z_{cis}	3.29	2.54	
N_{trans}	3.24 (4 MS-CASPT2)	3.18	3.18
Emission energy (eV)			
A_{cis}	2.42	2.08	2.39
Z_{cis}	1.63	0.67	
N_{trans}	2.84 (4 MS-CASPT2)	2.84	2.76

Table 3.3 Calculated population (%) of different protonation states at pH 7.^{a)}

Geometry	Off-state A_{trans} vs. N_{trans}		On-state A_{cis} vs. N_{cis}		On-state A_{cis} vs. Z_{cis}		
	ONIOM// N_{trans} ^{b)}	X-ray	ONIOM// A_{cis} ^{c)}	X-ray	ONIOM// A_{cis} ^{c)}	X-ray	ONIOM// Z_{cis} ^{d)}
Chromophore	N: 100	N: 100	A: 100	A: 94	A: 98	A: 100	A: 100
Glu144	A: 100	A: 100	A: 100	A: 100	A: 100	A: 100	A: 100
His193	HIE: 100	HIE: 100	HIP: 99	HIP: 49	HIP: 98	HIP: 43	HIP: 92
Glu211	A: 100	A: 100	A: 99	A: 52	A: 100	A: 57	A: 88

a) Neutral and anionic forms of chromophore are denoted by N and A, respectively. Neutral and cationic His193 are denoted by HIE and HIP, respectively. The anionic Glu144 and Glu211 are denoted by A.

b) ONIOM-EE optimized structure for N_{trans} in the off-state protein is used.

c) ONIOM-EE optimized structure for A_{cis} in the on-state protein is used.

d) ONIOM-EE optimized structure for Z_{cis} in the on-state protein is used.

tions by turning off the charge on the side chain of one of these residues. The population of A_{cis} is hardly changed by eliminating the charges of Glu144, Glu211 or His193 (~93–100%). However, population of A_{cis} was predicted to be reduced from 100% to about 46–66% by excluding the charges of Ser142, which is qualitatively supported by a recent mutation study, where mutation of Ser142 by Ala, Asp, Cys or Gly affords the neutral chromophore [91e].

The reaction mechanism of reversibly photochromic properties in Dronpa is not clear, although it has been explained to be affected by several factors (protonation states, conformations, non-planarity or flexibility of the chromophore, as well as intersystem crossing). Miyawaki and coworkers first proposed that the reaction is initiated by excited-state proton transfer (ESPT) from the neutral chromophore in the off-state Dronpa and gives the anionic chromophore in the on-state Dronpa (analogous to mechanism A in Scheme 3.16) [81]. The mechanism involving isomerization of the chromophore leading to changes of the protonation state was calculated to be one of the feasible pathways in our study (mechanism B in Scheme 3.16). However, the reaction barrier of the trans–cis isomerization (the first-step in the mechanism B), probably via a hula-twisted pathway [93], is still unknown. Also, this proposed pathway may not explain the observed KIE in Dronpa [81e]. To account for the observed KIE and conformational change of the chromophore in the crystal structures, we proposed an alternative mechanism involving photoisomerization followed by ESPT in a concerted or stepwise manner and, finally, isomerization (mechanism C in Scheme 3.16). Our CASSCF calculations suggested a lower photoisomerization barrier along the R5 bond, to give a stable twisted minimum N_{TI} , than photoisomerization along the R6 bond via a twisted transition state. The imidazolinone and phenol rings in the excited-state N_{TI} bear about -0.89 and $+0.49$ e [i.e., twisted intramolecular charge transfer (TICT) state], respectively. Therefore, the acidity of phenol moiety is enhanced by more favorable photoisomerization along the R5 bond, which should promote ESPT to afford an anionic twisted

intermediate (A_{TI}). Isomerization of A_{TI} eventually gives A_{cis} . Our theoretical studies on the reaction mechanism of Dronpa via photoisomerization and excited-state proton transfer are in progress.

asFP595, which was obtained from the sea anemone *Anemonia sulcata*, is another photoswitchable protein for the emission of red light [83b, 84a]. asFP595 can be switched from the non-fluorescent “off” state to the fluorescent “on” state by irradiation with green light. It has been investigated by QM and QM/MM calculations, as well as QM/MM excited-state molecular dynamics (MD) simulations [94]. Excited-state dynamics of the three possible protonated states (neutral, anionic and zwitterionic) of the chromophore were explored by QM/MM(CASSCF(6e,6o)/3-21G:OPLS) MD simulations [94a]. The trans neutral chromophore was found to undergo trans-to-cis photoisomerization in one of the five trajectories. Upon excitation to S_1 , rotation of the imidazolinone part of the chromophore takes place to give a twisted excited-state minimum, followed by accessing a conical intersection and hopping to the ground state. Rotation of the phenol part of the chromophore becomes important after decay from the excited state to the ground state. In contrast, only two of the five trajectories starting from the cis neutral chromophore were found to decay to the ground state within 10 ps, in which the cis-to-trans photoisomerization was observed in the one trajectory via a pathway similar to the neutral trans chromophore. These simulations suggested a higher probability of the cis-to-trans isomerization than the trans-to-cis isomerization. In addition, the protein matrix was found to stabilize S_1 more than S_0 (on average by 32 and 47 kJ mol⁻¹ for the trans and cis forms, respectively) in the isomerization process, particularly at the surface crossing seam.

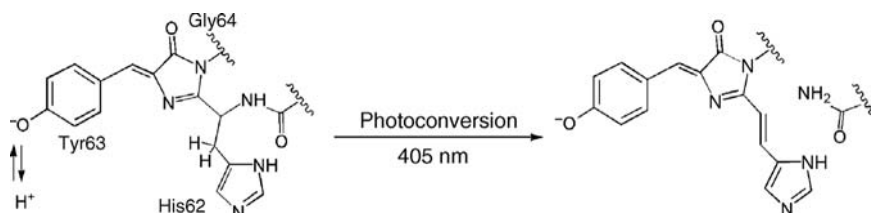
Moreover, both the trans and cis anionic chromophores suffer from a rapid radiationless decay in 20 trajectories. No photoisomerization process was observed. In addition, the lifetime of the excited-state cis chromophore was calculated to be roughly four times longer than the trans anionic chromophore. Unlike with the neutral chromophore, the decay was driven by rotation of the phenoxide part only and thus no isomerization was rendered. Therefore, the anionic forms were concluded to be responsible for ultrafast radiationless deactivation, especially for the non-planarity of the trans chromophore. Moreover, the protein environment strongly stabilized S_1 more than S_0 , by 26 and 20 kJ mol⁻¹ at the conical intersection.

The trans and cis zwitterionic chromophores were found to be stable as the planar excited-state minimum and not to proceed radiativeless decay within 10 ps from 10 trajectories. It was found that the minimum energy conical intersection was higher in energy than the planar S_1 minimum and Frank–Condon point by 70 and 23 kJ mol⁻¹, respectively, in the gas phase. Also, the protein environment did not stabilize the excited state in the isomerization process. These factors were attributed to suppressing radiationless relaxation for the zwitterionic chromophores. Therefore, the zwitterionic forms were proposed as the putative fluorescent state for asFP595.

3.4.1.3 Photoconversion of Fluorescent Proteins

The irreversible photoconversion fluorescent proteins, Kaede by Miyawaki's group and EosFP and IrisFP by Nienhaus's groups, respectively [82a–c,f, 84d, 95]. The color of fluorescence in these fluorescent proteins was found to be irreversibly

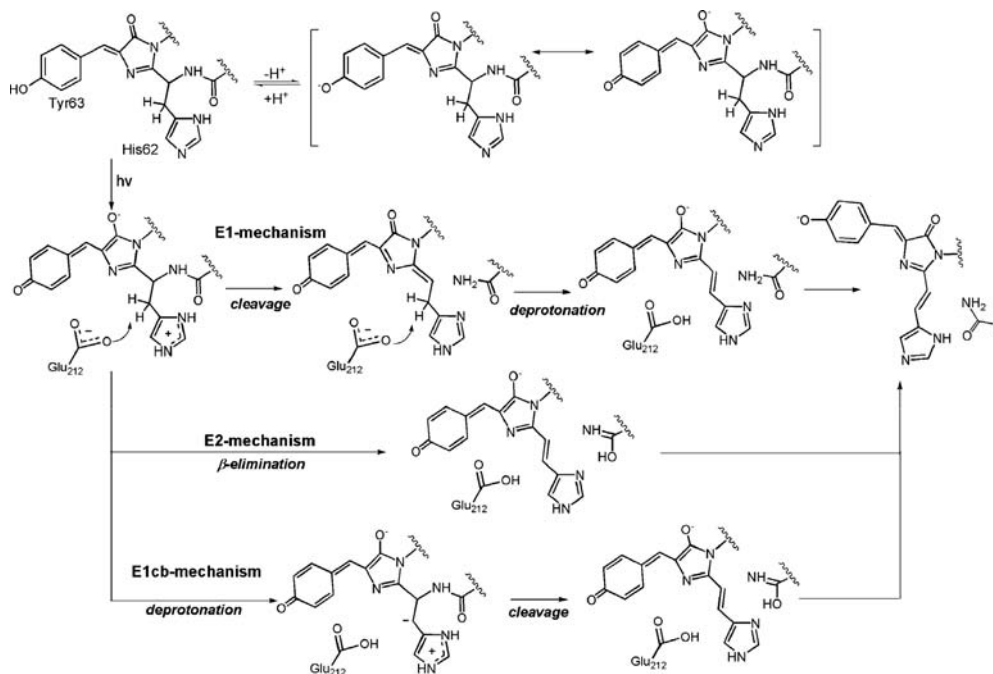
changed from green to red by irradiation with UV/VIS light. X-ray crystal structures of the red- and green-forms for Kaede, EosFP and IrisFP recently became available [82c, 84d, 95]. The structures of Kaede and EosFP are very similar, with a root mean square deviation of 0.78 Å for 876 C_α [82c, 95]. These three fluorescent proteins have a chromophore formed from the His62-Tyr63-Gly64 tripeptide. The UV/VIS irradiation was found to unusually break the peptide backbone and extend the π -conjugation, and thus results in the change of emission from green to red (Scheme 3.17). Irradiation at 350–410 nm was shown to be important in photo-induced conversion, which presumably excites the neutral green-form chromophore. The observed redshift of absorption (green and red forms: 508 and 572 nm, respectively) and fluorescence (green and red forms: 518 and 582 nm, respectively) in Kaede is because of an extension of π -conjugation of the chromophore [82a]. Photoinduced cleavage of such a peptide backbone between the α -nitrogen and the α -carbon at His62 is unique, since proteases catalyze cleavage of more reactive peptide amide bonds [82c,f, 84d, 95].



Scheme 3.17 UV-induced protein cleavage and green-to-red conversion of fluorescent proteins Kaede and EosFP.

Mutation studies showed that His62 and Glu212 are critical for irreversible photoconversion [82b,c,f, 95]. Accordingly, two mechanisms, E1 and E2, were proposed by Miyawaki's and Nienhaus's groups, respectively (Scheme 3.18). The stepwise E1 mechanism starts with the C–N bond cleavage followed by deprotonation via the basic carboxylic group of Glu212. Alternatively, the concerted E2 mechanism for the C–N bond cleavage and deprotonation followed by tautomerization can also give the final product.

We have performed ONIOM(B3LYP:MM) calculations with a very large QM model to study the feasibility of the proposed mechanisms for photoinduced peptide cleavage and green-to-red conversion of fluorescent protein Kaede [96]. Three pathways, namely, E1, E1cb, and E2, involving cleavage of the peptide backbone as well as deprotonation from the β -carbon by Glu212 were investigated (Scheme 3.18). Interestingly, the stepwise E1 mechanism (i.e., C–N bond cleavage followed by deprotonation process) was calculated to be comparable in energy to an alternative E1cb mechanism (deprotonation step prior to C–N bond cleavage) in our ONIOM (DFT:MM) calculations. The reaction barriers in these two pathways are about 20 kcal mol^{−1}. However, the E2-elimination transition state, that is, the concerted deprotonation and C–N bond cleavage, can not be found and the calculations instead led to the lowest-energy E1-type transition state. A two-dimensional PES scan by the



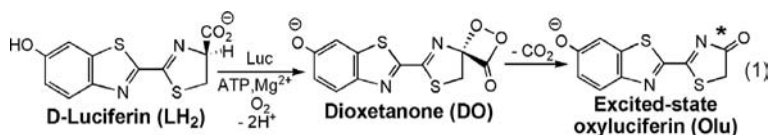
Scheme 3.18 Proposed mechanisms (E1, E2 and E1cb) of the green-to-red photoconversion reaction in Kaede and EosFP.

ONIOM calculations showed that the E2-type pathway was much higher in energy (about 34 kcal mol^{-1}).

3.4.2

Luciferases

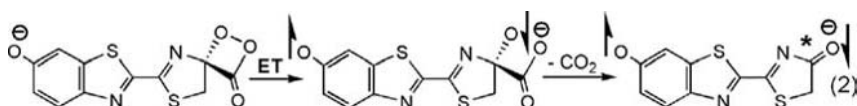
Firefly emission is a well-known efficient bioluminescence [5]. The recently revised quantum yield (Z_{bi}) is about 0.41 [97]. The widely accepted reaction mechanism of firefly bioluminescence involves reaction of D-luciferin, ATP and O_2 to give oxyluciferin in a singlet excited state, the assumed emitter, via formation of D-luciferyl adenylate (not shown) and high-energy dioxetanone intermediate (DO) in the firefly luciferase (Luc) (Scheme 3.19). Recently, Kato and coworkers obtained X-ray crystal structures of Japanese firefly luciferase containing a high-energy intermediate analogue or oxyluciferin (Olu) product [98] that clearly showed the catalytic and emission center. In contrast, decomposition of simple dioxetanes or dioxetanones



Scheme 3.19 Schematic diagram for firefly bioluminescence.

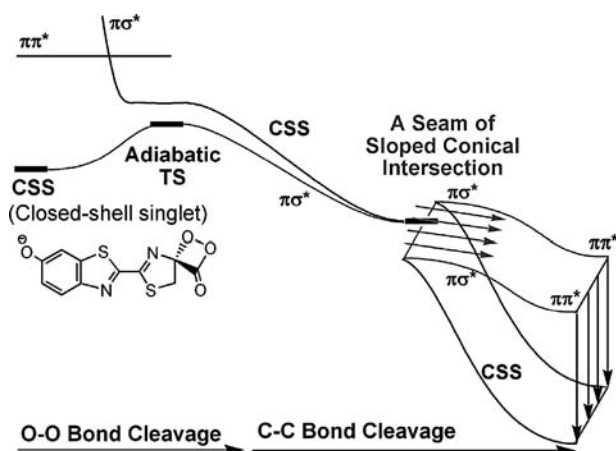
(without the strong electron donor) gives rise to the respective carbonyl compounds predominantly in a triplet excited state, rather than the singlet excited state [99]. For these simpler systems, a diradical mechanism or, later, “merged” mechanism initiating with homolytic O–O bond cleavage was proposed and supported by theoretical studies [100].

The mystery of efficient thermal generation of electronic excited states in firefly remains unclear, particularly at the atomic and molecular levels. Intramolecular chemically initiated electron-exchange luminescence (CIEEL), in which one electron transfer from the phenoxide anionic moiety to the dioxetanone part, has been proposed to account for chemiexcitation process in firefly (Scheme 3.20) [101a–c]. However, the reaction mechanism via electron transfer was questioned, and a modified mechanism via charge transfer was suggested [101d]. Although a sloped conical intersection (CI) was suggested to be a key to access the excited state from the ground state [102], the nature of mysterious channel for efficient thermal generation of the singlet excited state in firefly is still unclear.



Scheme 3.20 Proposed intramolecular chemically initiated electron-exchange luminescence.

Recently, we carried out SA-CASSCF(12,12)/6-31G* and CASPT2(12,12)/6-31G*//SA-CASSCF(12,12)/6-31G* calculations to elucidate the reaction mechanism of bioluminescence from the firefly dioxetanone in the gas phase [103]. Our calculations showed that the decomposition reaction of the high-energy anionic intermediate DO starts with the O–O bond cleavage via an adiabatic transition state (TS). When the O–O bond is significantly elongated, the ground- and excited-state surfaces were calculated to become close in energy and thus the O–O bond cleavage transition state was found to be mixed with the two electronic configurations, closed-shell singlet (CSS) and π, σ^* state (Scheme 3.21), suggesting the occurrence of an avoided



Scheme 3.21 Schematic reaction mechanism of firefly bioluminescence based on our CASSCF calculations.

crossing. The π, σ^* state can be regarded formally as CIEEL [101a], but charge transfer, rather than electron transfer, was found to occur in our CASSCF calculations [103].

When the subsequent C–C bond cleavage occurs, the two surfaces for the closed-shell configuration and intramolecular charge-transfer π, σ^* state re-cross at the minimum energy conical intersection (MECI), which was discovered for the first time. Interestingly, gradients on S_0 and S_1 , and the two branching space coordinates of MECI [i.e., gradient difference vector (GDV) and derivative coupling vector (DCV)] essentially follow along the intrinsic reaction pathway of the second step (the C–C bond stretching and O–C–O bending). Since the final reaction and surface crossing are qualitatively in the same coordinate space, the molecule should have high probability to encounter MECI. In addition, the computed pathway from the adiabatic transition state to access MECI is barrierless. Moreover, MECI is a *sloped* conical intersection, not a *peaked* conical intersection, and looks like a $n-1$ -dimensional seam (rather than the usual $n-2$ -dimensional point, where n is the number of the vibrational modes) [102], with two surfaces along one of the branching space coordinates (GDV) very close in energy (Scheme 3.21 and Figure 3.6). Such unique topology provides a widely extended channel to diabatically access the excited state from the ground state and thereby large transition probability, attained partly by large velocity along the reaction coordinate ($\Delta E_{TS} - E_{MECI} = 20.7$ and $26.4-27.8$ kcal mol⁻¹ by CASSCF and CASPT2//CASSCF methods, respectively).

In comparison, the C–C cleavage process in the ground-state surface has a very larger energy gap between the ground- and excited-states, and is also much lower in energy than MECI. Moreover, conformation change occurs, in which the oxyluciferin part becomes planar and is dominated by the closed-shell singlet state. Therefore, it should be a potential non-radiative decay channel in the gas phase, and should be partly suppressed by the protein. Studies of the effects of protein and finite temperature are in progress. In contrast with the firefly dioxetanone, the C–C bond cleavage has to go through transition states in S_1 and T_1 for the case of the simple dioxetanes or dioxetanone [100, 103]. The higher preference for the formation of triplet carbonyl compound over singlet excited-state product was attributed to a higher energy of the C–C cleavage transition state for S_1 than for T_1 [100, 103].

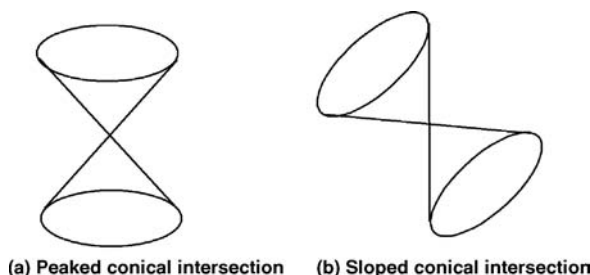


Figure 3.6 Schematic topology of two different conical intersections.

3.5

Conclusion

New mechanisms and insight into metalloenzyme and photobiological reactions have been gained from QM and QM/MM calculations, leading to the advancement of our understanding of these complex systems. Recent insights into the reaction mechanisms for critical heme-containing oxygenases, B_{12} -dependent enzymes, fluorescent proteins and firefly luciferase have been reviewed and discussed in this chapter. It is anticipated that the understanding of enzymatic reactions is being significantly improved by several recent methodological developments in multi-scale simulations, such as faster and more reliable QM and MM methods and algorithms [20, 104–107], efficient sampling and MD algorithms [108], free energy profiles via QM/MM free energy perturbation [109] and minimum free-energy paths (MFEPs) [33k, 110], QM/MM combining with continuum solvent models [Poisson-Boltzmann (PB)] [111], generalized solvent boundary potential (GSBP) [112], generalized-Born surface-area (GBSA) [113]), or even possibly combining with coarse grain (CG) model [114], especially for membranes. In addition, de novo computational design of artificial metalloenzymes and photobiology is another challenging and important application for the theoreticians and experimentalists to collaboratively realize [115]. We are confident that a revised chapter will have to be written before long to accommodate the anticipated great and rapid progress in this field.

Acknowledgments

L.W.C. acknowledges the Fukui Institute Fellowship. This work is in part supported by the Japan Science and Technology Agency (JST) with a Core Research for Evolutional Science and Technology (CREST) grant in the Area of High Performance Computing for Multiscale and Multiphysics Phenomena.

References

- 1 (a) Heme: Springer, B.A., Sligar, S.G., Olson, J.S., and Phillips, G.N. Jr. (1994) *Chem. Rev.*, **94**, 699; (b) Stenkamp, R.M. (1994) *Chem. Rev.*, **94**, 715; (c) Sono, M., Roach, M.P., Coulter, E.D., and Dawson, J.H. (1996) *Chem. Rev.*, **96**, 2841; (d) Rosen, G.M., Tsai, P., and Pou, S. (2002) *Chem. Rev.*, **102**, 1191; (e) Colas, C. and Ortiz de Montellano, P.R. (2003) *Chem. Rev.*, **103**, 2305; (f) Lukin, J.A. and Ho, C. (2004) *Chem. Rev.*, **104**, 1219; (g) Meunier, B., de Visser, S.P., and Shaik, S. (2004) *Chem. Rev.*, **104**, 3947; (h) Denisov, I.G., Makris, T.M., Sligar, S.G., and Schlichting, I. (2005) *Chem. Rev.*, **105**, 2253; (i) Collman, J.P., Boulatov, R., Sunderland, C.J., and Fu, L. (2004) *Chem. Rev.*, **104**, 561; (j) Poulos, T.L. (2005) *Biochem. Biophys. Res. Commun.*, **338**, 337; (k) Nam, W. (2007) *Acc. Chem. Res.*, **40**, 522; (l) Watanabe, Y., Nakajima, H., and Ueno, T. (2007) *Acc. Chem. Res.*, **40**, 554; (m) Sigel, A., Sigel, H., and Sigel, R.K.O. (2007) *Ubiquitous Roles of Cytochrome P450 Proteins*, John Wiley & Sons, Ltd., Chichester, UK.
- 2 (a) B_{12} : Banerjee, R. (2003) *Chem. Rev.*, **103**, 2083; (b) Banerjee, R. and Ragsdale,

- S.W. (2003) *Annu. Rev. Biochem.*, **72**, 209;
- (c) Toraya, T. (2003) *Chem. Rev.*, **103**, 2095;
- (d) Banerjee, R. (1999) *Chemistry and Biochemistry of B12*, John Wiley & Sons, Inc., New York; (e) Kräutler, B., Arigoni, D., and Golding, B.T. (1998) *Vitamin B12 and the B12 proteins*, Wiley-VCH Verlag GmbH, Weinheim; (f) Brown, K.L. (2005) *Chem. Rev.*, **105**, 2075; (g) Ludwig, M.L. and Matthews, R.G. (1997) *Annu. Rev. Biochem.*, **66**, 269.
- 3 (a) Que, L. and Tolman, W.B. (2004) *Comprehensive Coordination Chemistry II*, vol. 8, Elsevier, Oxford; (b) Bertini, I., Sigel, A., and Sigel, H. (2001) *Handbook on Metalloproteins*, Marcel Dekker, New York; (c) Messerschmidt, A. (2001) *Handbok of Metalloproteins*, John Wiley & Sons, Inc., New York; (d) Holm, R.H., Kennepohl, P., and Solomon, E.I. (1996) *Chem. Rev.*, **96**, 2239.
- 4 (a) Theoretical: Siegbahn, P.E.M. and Blomberg, M.R. (1999) *Annu. Rev. Phys. Chem.*, **50**, 221; (b) Siegbahn, P.E.M. and Blomberg, M.R.A. (2000) *Chem. Rev.*, **100**, 421; (c) Loew, G.H. and Harris, D.L. (2000) *Chem. Rev.*, **100**, 407; (d) Spiro, T.G., Zgierski, M.Z., and Kozlowski, P.M. (2001) *Coord. Chem. Rev.*, **219–221**, 923; (e) Himo, F. and Siegbahn, P.E.M. (2003) *Chem. Rev.*, **103**, 2421; (f) Baik, M.-H., Newcomb, M., Friesner, R.A., and Lippard, S.J. (2003) *Chem. Rev.*, **103**, 2385; (g) Friesner, R.A., Baik, M.-H., Guallar, V., Gherman, B.F., Wirstam, M., Murphy, R.B., and Lippard, S.J. (2003) *Coord. Chem. Rev.*, **238–239**, 267; (h) Lovell, T., Himo, F., Han, W.-G., and Noodleman, L. (2003) *Coord. Chem. Rev.*, **238–239**, 211; (i) Noodleman, L., Lovell, T., Han, W.-G., Li, J., and Himo, F. (2004) *Chem. Rev.*, **104**, 459; (j) Shaik, S., Kumar, D., de Visser, S.P., Altun, A., and Thiel, W. (2005) *Chem. Rev.*, **105**, 2279; (k) Ghosh, A. (2005) *Acc. Chem. Res.*, **38**, 943; (l) Yoshizawa, K. (2006) *Acc. Chem. Res.*, **39**, 375; (m) Siegbahn, P.E.M. and Borowski, T. (2006) *Acc. Chem. Res.*, **39**, 729; (n) Shaik, S., Hirao, H., and Kumar, D. (2007) *Acc. Chem. Res.*, **40**, 532; (o) Dudev, T. and Lim, C. (2007) *Acc. Chem. Res.*, **40**, 85; (p) Sproviero, E.M., Gascon, J.A., McEvoy, J.P., Brudvig, G.W., and Batista, V.S. (2008) *Coord. Chem. Rev.*, **252**, 395; (q) Jensena, K.P. and Ryde, U. (2009) *Coord. Chem. Rev.*, **253**, 769; (r) Morokuma, K., Musaev, D.G., Vreven, T., Basch, H., Torrent, M., and Khoroshun, D.V. (2001) *IBM J. Res. & Dev.*, **45**, 367.
- 5 (a) Shimomura, O. (2006) *Bioluminescence: Chemical Principles and Methods*, World Scientific, New Jersey; (b) McCapra, F. (2000) *Methods Enzymol.*, **305**, 3; (c) Wilson, T. and Hastings, J.W. (1998) *Annu. Rev. Cell Dev. Biol.*, **14**, 197.
- 6 (a) Conn, P.M. (1999) *Methods in Enzymology*, vol. 302, Academic Press, San Diego; (b) Zimmer, M. (2002) *Chem. Rev.*, **102**, 759; (c) Remington, S.J. (2006) *Curr. Opin. Struct. Biol.*, **16**, 714; (d) Tsien, R.Y. (1998) *Annu. Rev. Biochem.*, **67**, 509; (e) Chalfie, M. and Kain, S.R. (2006) *Green Fluorescent Protein: Properties, Applications and Protocols*, Wiley-Interscience, Hoboken; (f) Sullivan, K.F. (2008) *Methods in Cell Biology*, vol. 85, Academic Press, London; (g) Lippincott-Schwartz, J., Altan-Bonnet, N., and Patterson, G.H. (2003) *Nat. Cell Biol.*, **5**, S7; (h) Miyawaki, A., Sawano, A., and Kogure, T. (2003) *Nat. Cell Biol.*, **5**, S1; (i) Chudakov, D.M., Lukyanov, S., and Lukyanov, K.A. (2005) *Trends Biotechnol.*, **23**, 605; (j) Zhang, J., Campbell, R.E., Ting, A.Y., and Tsien, R.Y. (2002) *Nat. Rev. Mol. Cell Biol.*, **3**, 906; (k) Shaner, N.C., Patterson, G.H., and Davidson, M.W. (2007) *J. Cell Sci.*, **120**, 4247; (l) Henderson, J.N. and Remington, S.J. (2006) *Physiology*, **21**, 161; (m) Lukyanov, K.A., Chudakov, D.M., Lukyanov, S., and Verkhusha, V.V. (2005) *Nat. Rev. Mol. Cell Biol.*, **6**, 885; (n) Lippincott-Schwartz, J. and Patterson, G.H. (2003) *Science*, **300**, 87.
- 7 (a) Roos, B.O. (1999) *Acc. Chem. Res.*, **32**, 137; (b) Helms, V. (2002) *Curr. Opin. Struct. Biol.*, **12**, 169; (c) Dreuw, A. and Head-Gordon, M. (2005) *Chem. Rev.*, **105**, 4009; (d) Olsen, S., Toniolo, A., Ko, C., Manohar, L., Lamothe, K., and Martinez, T.J. (2005) *Theoretical and Computational Chemistry*, vol. 16 (ed. M. Olivucci), Elsevier, pp. 225–254; (e) Martínez, T.J. (2006) *Acc. Chem. Res.*, **39**, 119; (f) Gascón, A., Sproviero, E.M., and Batista, V.S. (2006) *Acc. Chem. Res.*, **39**, 184; (g) Dreuw, A. (2006) *ChemPhysChem*,

- 7, 2259; (h) Garavelli, M. (2006) *Theor. Chem. Acc.*, **116**, 87; (i) Levine, B.G. and Martínez, T.J. (2007) *Annu. Phys. Chem.*, **58**, 613.
- 8 (a) Kohn, W. and Sham, L.J. (1965) *Phys. Rev.*, **140**, A1133; (b) Hohenberg, P. and Kohn, W. (1964) *Phys. Rev.*, **136**, B864; (c) Kohn, W. and Holthausen, M.C.A. (2000) *Chemist's Guide to Density Functional Theory*, Wiley-VCH, Verlag, GmbH, Weinheim; (d) Kohn, W., Becke, A.D., and Parr, R.G. (1996) *J. Phys. Chem.*, **100**, 12974; (e) Parr, R.G. and Yang, W. (1989) *Density Functional Theory of the Electronic Theoretical Methods for Structure of Molecules*, Oxford University Press, New York; (f) Cramer, C.J. (2004) *Essentials of Computational Chemistry*, John Wiley & Sons Ltd., Chichester, pp. 249–303.
- 9 (a) Becke, A.D. (1993) *J. Chem. Phys.*, **98**, 5648; (b) Lee, C., Yang, W., and Parr, R.G. (1988) *Phys. Rev. B*, **37**, 785.
- 10 Curtiss, L.A., Redfern, P.C., and Raghavachari, K. (2005) *J. Chem. Phys.*, **123**, 124107.
- 11 (a) Schwabe, T. and Grimme, S. (2008) *Acc. Chem. Res.*, **41**, 569; (b) Zhao, Y. and Truhlar, D.G. (2008) *Acc. Chem. Res.*, **41**, 15 and references therein.
- 12 (a) Niu, S. and Hall, M.B. (2000) *Chem. Rev.*, **100**, 353; (b) Schultz, N.E., Zhao, Y., and Truhlar, D.G. (2005) *J. Phys. Chem. A*, **109**, 11127; (c) Furche, F. and Perdew, J.P. (2006) *J. Chem. Phys.*, **124**, 44103.
- 13 (a) Siegbahn, P.E.M. (2006) *J. Biol. Inorg. Chem.*, **11**, 695; (b) Neese, F. (2006) *J. Biol. Inorg. Chem.*, **11**, 702.
- 14 Raghavachari, K., Trucks, G.W., Pople, J.A., and Head-Gordon, M. (1989) *Chem. Phys. Lett.*, **87**, 5968.
- 15 (a) Reiher, M., Salomon, O., and Hess, B.A. (2001) *Theor. Chem. Acc.*, **107**, 48; (b) Salomon, O., Reiher, M., and Hess, B.A. (2002) *J. Chem. Phys.*, **117**, 4729.
- 16 (a) Roos, B.O. (1987) *Adv. Chem. Phys.*, **69**, 399; (b) Cramer, C.J. (2004) *Essentials of Computational Chemistry*, John Wiley & Sons, Ltd., Chichester, pp. 203–223; (c) Jensen, F. (2007) *Introduction to Computational Chemistry*, 2nd edn, John Wiley & Sons, Ltd., Chichester, pp. 153–159.
- 17 (a) Roos, B.O. and Andersson, K. (1995) *Chem. Phys. Lett.*, **245**, 215; (b) Andersson, K., Malmqvist, P.-Å., and Roos, B.O. (1992) *J. Chem. Phys.*, **96**, 1218.
- 18 Ghosh, A. (2007) *J. Biol. Inorg. Chem.*, **11**, 712.
- 19 Pierloot, K. and Vancoillie, S. (2008) *J. Chem. Phys.*, **128**, 34104.
- 20 Aquilante, F., Malmqvist, P.-Å., Pedersen, T.B., Ghosh, A., and Roos, B.O. (2008) *J. Chem. Theory Comput.*, **4**, 694.
- 21 (a) Olivucci, M. (2005) *Theoretical and Computational Chemistry*, vol. 16, Elsevier, (b) Kutateladze, A.G. (2005) *Molecular and Supramolecular Photochemistry*, vol. 13, CRC Press; (c) Robb, M.A., Garavelli, M., Olivucci, M., and Bernardi, F. (2000) *Rev. Comput. Chem.*, **15**, 87; (d) Grimme, S. (2004) *Rev. Comput. Chem.*, **20**, 153; (e) Serrano-Andrés, L. and Merchán, M. (2005) *J. Mol. Struct. Theorchem*, **729**, 99.
- 22 (a) Runge, E. and Gross, E.K.U. (1984) *Phys. Rev. Lett.*, **52**, 997; (b) Stratmann, R.E., Scuseria, G.E., and Frisch, M.J. (1998) *J. Chem. Phys.*, **109**, 8218; (c) Bauernschmitt, R. and Ahlrichs, R. (1996) *Chem. Phys. Lett.*, **256**, 454; (d) Furche, F. and Ahlrichs, R. (2002) *J. Chem. Phys.*, **117**, 7433; (e) Elliot, P., Furche, F., and Burke, K. (2009) *Rev. Comput. Chem.*, **26**, 91.
- 23 Christiansen, O., Koch, H., and Jørgensen, P. (1995) *Chem. Phys. Lett.*, **243**, 409.
- 24 (a) Nakatsuji, H. (1978) *Chem. Phys. Lett.*, **59**, 362; (b) Nakatsuji, H. (1989) *Chem. Phys. Lett.*, **67**, 329; (c) Nakatsuji, H. (1989) *Chem. Phys. Lett.*, **67**, 334.
- 25 (a) Tozer, D.J. and Handy, N.C. (1998) *J. Chem. Phys.*, **109**, 10180; (b) Allen, M.J. and Tozer, D.J. (2000) *J. Chem. Phys.*, **113**, 5185; (c) Dreuw, A. and Head-Gordon, M. (2004) *J. Am. Chem. Soc.*, **126**, 4007; (d) Grimme, S. and Parac, M. (2003) *ChemPhysChem*, **3**, 292.
- 26 (a) Silva-Junior, M.R., Schreiber, M., Sauer, S.P.A., and Thiel, W. (2008) *J. Chem. Phys.*, **129**, 104103; (b) Schreiber, M., Silva-Junior, M.R., Sauer, S.P.A., and Thiel, W. (2008) *J. Chem. Phys.*, **129**, 134110.
- 27 (a) Blomberg, M.R.A. and Siegbahn, P.E.M. (1997) *Theor. Chem. Acc.*, **97**, 72;

- (b) Blomberg, M.R.A. and Siegbahn, P.E.M. (1999) *Mol. Phys.*, **96**, 571.
- 28 (a) Harvey, S.C. (1989) *Proteins: Struct., Funct., Genet.*, **5**, 78; (b) Sharp, K.A. and Honig, B. (1990) *Annu. Rev. Biophys. Biophys. Chem.*, **19**, 301; (c) Richards, F.M. (1977) *Annu. Rev. Biophys. Bioeng.*, **6**, 151.
- 29 Koshland, D.E. (1958) *Proc. Natl. Acad. Sci. U.S.A.*, **44**, 98.
- 30 Chen, S.-L., Fang, W.-H., and Himio, F. (2008) *Theor. Chem. Acc.*, **120**, 515.
- 31 Warshel, A. and Levitt, M. (1976) *J. Mol. Biol.*, **103**, 227.
- 32 (a) Singh, U.C. and Kollman, P.A. (1986) *J. Comput. Chem.*, **7**, 718; (b) Field, M.J., Bash, P.A., and Karplus, M. (1990) *J. Comput. Chem.*, **11**, 700.
- 33 (a) Gao, J. (1996) *Rev. Comput. Chem.*, **7**, 119; (b) Monard, G. and Merz, K.M. Jr. (1999) *Acc. Chem. Res.*, **32**, 904; (c) Gao, J. and Truhlar, D.G. (2002) *Annu. Rev. Phys. Chem.*, **53**, 467; (d) Field, M.J. (2002) *J. Comput. Chem.*, **23**, 48; (e) Garcia-Viloca, M., Gao, J., Karplus, M., and Truhlar, D.G. (2004) *Science*, **303**, 186; (f) Friesner, R.A. and Guallar, V. (2005) *Annu. Rev. Phys. Chem.*, **56**, 389; (g) Mulholland, A.J. (2005) *Drug Discov. Today*, **10**, 1393; (h) Warshel, A., Sharma, P.K., Kato, M., Xiang, Y., Liu, H.B., and Olsson, M.H.M. (2006) *Chem. Rev.*, **106**, 3210; (i) Lin, H. and Truhlar, D.G. (2007) *Theor. Chem. Acc.*, **117**, 185; (j) Senn, H.M. and Thiel, W. (2007) *Top. Curr. Chem.*, **268**, 173; (k) Hu, H. and Yang, W. (2008) *Annu. Rev. Phys. Chem.*, **59**, 573; (l) Senn, H.M. and Thiel, W. (2008) *Angew. Chem., Int. Ed.*, **47**, 1198.
- 34 (a) Maseras, F. and Morokuma, K. (1995) *J. Comput. Chem.*, **16**, 1170; (b) Humbel, S., Sieber, S., and Morokuma, K. (1996) *J. Chem. Phys.*, **105**, 1959; (c) Matsubara, T., Sieber, S., and Morokuma, K. (1996) *Int. J. Quantum Chem.*, **60**, 1101; (d) Svensson, M., Humbel, S., Froese, R.D.J., Matsubara, T., Sieber, S., and Morokuma, K. (1996) *J. Phys. Chem.*, **100**, 19357; (e) Svensson, M., Humbel, S., and Morokuma, K. (1996) *J. Chem. Phys.*, **105**, 3654; (f) Dapprich, S., Komáromi, I., Byun, S., Morokuma, K., and Frisch, M.J. (1999) *J. Mol. Struct. (THEOCHEM)*, **461**, 1; (g) Vreven, T. and Morokuma, K. (2000) *J. Comput. Chem.*, **21**, 1419; (h) Vreven, T., Frisch, M.J., Kudin, K.N., Schlegel, H.B., and Morokuma, K. (2006) *Mol. Phys.*, **104**, 701; (i) Vreven, T., Byun, K.S., Komáromi, I., Dapprich, S., Montgomery, J.A. Jr., Morokuma, K., and Frisch, M.J. (2006) *J. Chem. Theory Comput.*, **2**, 815; (j) Vreven, T. and Morokuma, K. (2006) *Annu. Rep. Comput. Chem.*, **2**, 35.
- 35 Bakowies, D. and Thiel, W. (1996) *J. Phys. Chem.*, **100**, 10580.
- 36 Hratchian, H.P., Parandekar, P.V., Raghavachari, K., Frisch, M.J., and Vreven, T. (2008) *J. Chem. Phys.*, **128**, 34107.
- 37 (a) Zhang, Y., Lin, H., and Truhlar, D.G. (2007) *J. Chem. Theory Comput.*, **3**, 1378; (b) Zhang, Y. and Lin, H. (2008) *J. Chem. Theory Comput.*, **4**, 414; (c) Geerke, D.P., Thiel, S., Thiel, W., and van Gunsteren, W.F. (2007) *J. Chem. Theory Comput.*, **3**, 1499; (d) Lu, Z. and Zhang, Y. (2008) *J. Chem. Theory Comput.*, **4**, 12.
- 38 (a) Eurenium, K.P., Chatfield, D.C., Brooks, B.R., and Hodoscek, M. (1996) *Int. J. Quantum Chem.*, **60**, 1189; (b) Koenig, P., Hoffman, M., Frauenheim, T., and Cui, Q. (2005) *J. Phys. Chem. B*, **109**, 9082; (c) Altun, A., Shaik, S., and Thiel, W. (2006) *J. Comput. Chem.*, **27**, 1324; (d) Zheng, J., Altun, A., Shaik, S., and Thiel, W. (2007) *J. Comput. Chem.*, **28**, 2147.
- 39 (a) Hooft, R.W.W., Vriend, G., Sander, C., and Abola, E.E. (1996) *Nature*, **381**, 272; (b) <http://swift.cmbi.ru.nl/gv/pdbreport/>, last accessed on 12.10.2009.
- 40 (a) Lovell, S.C., Davis, I. W., Arendall, W.B. III, de Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S., and Richardson, D.C. (2003) *Proteins: Struct., Funct., Genet.*, **50**, 437; (b) <http://molprobity.biochem.duke.edu/>, last accessed on 12.10.2009.
- 41 (a) Dolinsky, T.J., Nielsen, J.E., McCammon, J.A., and Baker, N.A. (2004) *Nucleic Acids Res.*, **32**, W665; (b) Brünger, A.T. and Karplus, M. (1988) *Proteins*, **4**, 148; (c) Hooft, R.W., Sander, C., and Vriend, G. (1996) *Proteins*, **26**, 363.
- 42 (a) Bashford, D. and Karplus, M. (1990) *Biochemistry*, **29**, 10219; (b) Bashford, D. and Gerwert, K.J. (1992) *Mol. Biol.*, **224**, 473; (c) Gordon, J.C., Myers, J.B., Foltz, T., Shoja, V., Heath, L.S., and Onufriev, A.

- (2005) *Nucleic Acids Res.*, **33**, W368; (d) <http://biophysics.cs.vt.edu/>, last accessed on 12.10.2009.; (e) <http://bioserv.rpbs.jussieu.fr/cgi-bin/PCE-pKa>; (f) Nicholls, A. and Honig, B. (1991) *J. Comput. Chem.*, **12**, 435; (g) Yang, A.-S., Gunner, M.R., Sampogna, R., Sharp, K., and Honig, B. (1993) *Proteins*, **15**, 252; (h) Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001) *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 10037; (i) Rabenstein, B. (1999) Karlsberg online manual <http://agknapp.chemie.fu-berlin.de/karlsberg>, last accessed on 12.10.2009; (j) Kieseritzky, G., and Knapp, E.W. (2008) *Proteins: Struct. Funct. Bioinf.*, **71**, 1335.
- 43 PROPKA: (a) Li, H., Robertson, A.D., and Jensen, J.H. (2005) *Proteins: Struct., Funct., Bioinf.*, **61**, 704; (b) <http://propka.ki.ku.dk/>, last accessed on 12.10.2009.
- 44 Ghosh, A. (2008) *The Smallest Biomolecules: Diatomics and their Interactions with Heme Proteins*, Elsevier.
- 45 (a) Ghosh, A. and Bocian, D.F. (1996) *J. Phys. Chem.*, **100**, 6363; (b) Spiro, T.G. and Kozlowski, P.M. (1998) *J. Am. Chem. Soc.*, **120**, 4524; (c) Rovira, C., Kunc, K., Hutter, J., Ballone, P., and Parrinello, M. (1997) *J. Phys. Chem. A*, **101**, 8914; (d) Rovira, C., Kunc, K., Hutter, J., Ballone, P., and Parrinello, M. (1998) *Int. J. Quantum Chem.*, **69**, 31; (e) Sigfridson, E. and Ryde, U. (1999) *J. Biol. Inorg. Chem.*, **4**, 99; (f) Kaupp, M., Rovira, C., and Parrinello, M. (2000) *J. Phys. Chem. B*, **104**, 5200; (g) Harvey, J.N. (2000) *J. Am. Chem. Soc.*, **122**, 12401; (h) Zhang, Y., Gossman, W., and Oldfield, E. (2003) *J. Am. Chem. Soc.*, **125**, 16387; (i) Jensen, K.P. and Ryde, U. (2004) *J. Biol. Chem.*, **279**, 14561; (j) Jensen, K.P., Roos, B.O., and Ryde, U. (2005) *J. Inorg. Biochem.*, **99**, 45; (k) Jensen, K.P., Roos, B.O., and Ryde, U. (2005) *J. Inorg. Biochem.*, **99**, 978; (l) Blomberg, L.M., Blomberg, M.R., and Siegbahn, P.E.M. (2005) *J. Inorg. Biochem.*, **99**, 949; (m) Ibrahim, M., Xu, C., and Spiro, T.G. (2006) *J. Am. Chem. Soc.*, **128**, 16834; (n) Degtyarenko, I., Nieminen, R.M., and Rovira, C. (2006) *Biophys. J.*, **91**, 2024; (o) Maréchal, J.-D., Maseras, F., Lledós, A., Mouawad, L., and Perahia, D. (2006) *J. Comput. Chem.*, **27**, 1446; (p) Rutkowska-zbik, D., Witko, M., and Stochel, G. (2007) *J. Comput. Chem.*, **28**, 825; (q) Strickland, N. and Harvey, J.N. (2007) *J. Phys. Chem. B*, **111**, 841; (r) Ribas-Ariño, J. and Novoa, J.J. (2007) *Chem. Commun.*, 3160. (s) Radoń, M. and Pierloot, K. (2008) *J. Phys. Chem. A*, **112**, 11824; (t) Nakashima, H., Hasegawa, J.-Y., and Nakatsuji, H. (2006) *J. Comput. Chem.*, **27**, 426; (u) Franzen, S. (2002) *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 16754.
- 46 (a) Dreuw, A., Dunietz, B.D., and Head-Gordon, M. (2002) *J. Am. Chem. Soc.*, **124**, 12070; (b) Dunietz, B.D., Dreuw, A., and Head-Gordon, M. (2003) *J. Phys. Chem. B*, **107**, 5623; (c) Angelis, F.D., Car, R., and Spiro, T.G. (2003) *J. Am. Chem. Soc.*, **125**, 15710; (d) Ohta, T., Pal, B., and Kitagawa, T. (2005) *J. Phys. Chem. B*, **109**, 21110.
- 47 Capece, L., Marti, M.A., Crespo, A., Doctorovich, F., and Estrin, D.A. (2006) *J. Am. Chem. Soc.*, **128**, 12455.
- 48 (a) Sigfridson, E. and Ryde, U. (2002) *J. Inorg. Biochem.*, **91**, 101; (b) Rovira, C. (2003) *J. Mol. Struct. (THEOCHEM)*, **632**, 309; (c) Angelis, F.D., Jarzecki, A.A., Car, R., and Spiro, T.G. (2005) *J. Phys. Chem. B*, **109**, 3065; (d) Alcantara, R.E., Xu, C., Spiro, T.G., and Guallar, V. (2007) *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 18451; (e) Chen, H., Ikeda-Saito, M., and Shaik, S. (2008) *J. Am. Chem. Soc.*, **130**, 14778.
- 49 (a) Guallar, V., Jarzecki, A.A., Friesner, R.A., and Spiro, T.G. (2006) *J. Am. Chem. Soc.*, **128**, 5427; (b) Ming, X. and Fang, W.-H. (2008) *J. Phys. Chem. B*, **112**, 990; (d) Marti, M.A., Capece, L., Crespo, A., Doctorovich, F., and Estrin, D.A. (2005) *J. Am. Chem. Soc.*, **127**, 77218.
- 50 Hoffmann, R., Chen, M.M.-L., and Thorn, D.L. (1977) *Inorg. Chem.*, **16**, 503.
- 51 (a) Pauling, L. and Coryell, C.D. (1936) *Proc. Natl. Acad. Sci. U.S.A.*, **22**, 210; (b) Weiss, J.J. (1964) *Nature*, **202**, 83; (c) Pauling, L. (1964) *Nature*, **203**, 182; (d) Weiss, J.J. (1964) *Nature*, **203**, 183; (e) McClure, D.S. (1960) *Radiat. Res. Suppl.*, **2**, 218; (g) Goddard, W.A. III and Olafson, B.D. (1975) *Proc. Natl. Acad. Sci. U.S.A.*, **72**, 2335.
- 52 (a) Bytheway, I. and Hall, M.B. (1994) *Chem. Rev.*, **94**, 639; (b) Newton, J.E. and Hall, M.B. (1984) *Inorg. Chem.*, **23**, 4627.

- 53 Phillips, G.N. Jr., Teodoro, M.L., Li, T., Smith, B., and Olson, J.S. (1999) *J. Phys. Chem. B*, **103**, 8817.
- 54 Perutz, M.F., Wilkinson, A.J., Paoli, M., and Dodson, G.G. (1998) *Annu. Rev. Biophys. Biomol. Struct.*, **137**, 1.
- 55 (a) Dunn, R.C., Xie, X., and Simon, J.D. (1993) *Methods Enzymol.*, **226**, 177; (b) Franzen, S., Kiger, L., Poyart, C., and Martin, J.-L. (2001) *Biophys. J.*, **80**, 2372.
- 56 (a) Kamachi, T., Shestakov, A.F., and Yoshizawa, K. (2004) *J. Am. Chem. Soc.*, **126**, 3672; (b) Sharma, P.K., Kevorkiants, R., de Visser, S.P., Kumar, D., and Shaik, S. (2004) *Angew. Chem., Int. Ed.*, **43**, 1129; (c) Kumar, D., de Visser, S.P., and Shaik, S. (2005) *J. Am. Chem. Soc.*, **127**, 8204; (d) Kamachi, T. and Yoshizawa, K. (2005) *J. Am. Chem. Soc.*, **127**, 10686; (e) Chen, H., Moreau, Y., Derat, E., and Shaik, S. (2008) *J. Am. Chem. Soc.*, **130**, 1953.
- 57 Matsui, T., Kim, S.H., Jin, H., Hoffman, B.M., and Ikeda-Saito, M. (2006) *J. Am. Chem. Soc.*, **128**, 1090.
- 58 (a) Kotake, Y. and Masayama, I.Z. (1936) *Physiol. Chem.*, **243**, 237; (b) Hayaishi, O., Rothberg, S., Mehler, A.H., and Saito, Y. (1957) *J. Biol. Chem.*, **229**, 889; (c) Yamamoto, S. and Hayaishi, O. (1967) *J. Biol. Chem.*, **242**, 5260; (d) Yoshida, R. and Hayaishi, O. (1987) *Methods Enzymol.*, **142**, 188.
- 59 Muller, A.J. and Scherle, P.A. (2006) *Nat. Rev. Cancer*, **6**, 613.
- 60 (a) Sugimoto, H., Oda, S.-I., Otsuki, T., Hino, T., Yoshida, T., and Shiro, Y. (2006) *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 2611; (b) Forouhar, E., Anderson, J.L.R., Mowat, C.G., Vorobiev, S.M., Hussain, A., Abashidze, M., Bruckmann, C., Thackray, S.J., Seetharaman, J., Tucker, T., Xiao, R., Ma, L.-C., Zhao, L., Acton, T.M., Montelione, G.T., Chapman, S.K., and Tong, L. (2007) *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 473; (c) Zhang, Y., Kang, S.A., Mukherjee, T., Bale, S., Crane, B.R., Begley, T.P., and Ealick, S.E. (2007) *Biochemistry*, **46**, 145.
- 61 Chung, L.W., Li, X., Sugimoto, H., Shiro, Y., and Morokuma, K. (2008) *J. Am. Chem. Soc.*, **130**, 12298.
- 62 (a) Siegbahn, P.E.M. and Haefner, F. (2004) *J. Am. Chem. Soc.*, **126**, 8919; (b) Borowski, T. and Siegbahn, P.E.M. (2006) *J. Am. Chem. Soc.*, **128**, 12941.
- 63 (a) Terentis, A.C., Thomas, S.R., Takikawa, O., Littlejohn, T.K., Truscott, R.J.W., Armstrong, R.S., Yeh, S.-R., and Stocker, R. (2002) *J. Biol. Chem.*, **277**, 15788; (b) Batabyal, D. and Yeh, S.-R. (2007) *J. Am. Chem. Soc.*, **129**, 15690.
- 64 (a) Murad, F. (1999) *Angew. Chem., Int. Ed.*, **38**, 1856; (b) Furchgott, R.F. (1999) *Angew. Chem., Int. Ed.*, **38**, 1870; (c) Ignarro, L.J. (1999) *Angew. Chem., Int. Ed.*, **38**, 1882.
- 65 (a) Hurshman, A.R., Krebs, C., Edmondson, D.E., Huynh, B.H., and Marletta, M.A. (1999) *Biochemistry*, **38**, 15689; (b) Wei, C.C., Wang, Z.Q., Hemann, C., Hille, R., and Stuehr, D.J. (2003) *J. Biol. Chem.*, **278**, 46668.
- 66 Davydov, R., Ledbetter-Rogers, A., Martásek, P., Larukhin, M., Sono, M., Dawson, J.H., Masters, B.S.S., and Hoffman, B.M. (2002) *Biochemistry*, **41**, 10375.
- 67 (a) Cho, K.-B. and Gauld, J.W. (2004) *J. Am. Chem. Soc.*, **126**, 10267; (b) Cho, K.-B. and Gauld, J.W. (2005) *J. Phys. Chem. B*, **109**, 23706; (c) Robinet, J.J., Cho, K.-B., and Gauld, J.W. (2008) *J. Am. Chem. Soc.*, **130**, 3328; (d) Cho, K.-B., Derat, E., and Shaik, S. (2007) *J. Am. Chem. Soc.*, **129**, 3182; (e) Cho, K.-B., Carvajal, M.A., and Shaik, S. (2009) *J. Phys. Chem. B*, **113**, 336; (f) De Visser, S.P. and Tan, L.S. (2008) *J. Am. Chem. Soc.*, **130**, 12961; (g) Morao, I., Periyasamy, G., Hillier, I.H., and Joule, J.A. (2006) *Chem. Commun.*, 3525. (h) Tantillo, D.J., Fukuto, J.M., Hoffman, B.M., Silverman, R.B., and Houk, K.N. (2000) *J. Am. Chem. Soc.*, **122**, 536.
- 68 (a) Banerjee, R. (2006) *ACS Chem. Biol.*, **1**, 149; (b) Reed, G.H. (2004) *Curr. Opin. Chem. Biol.*, **8**, 477; (c) Randaccio, L., Geremia, S., and Wuerger, J. (2007) *J. Organomet. Chem.*, **692**, 1198; (d) Vlasie, M.D. and Banerjee, R. (2003) *J. Am. Chem. Soc.*, **125**, 5431; (e) Marsh, E.N. and Drennan, C.L. (2001) *Curr. Opin. Chem. Biol.*, **5**, 499; (f) Babior, B.M. (1975) *Acc. Chem. Res.*, **8**, 376; (g) Toraya, T. (2000) *Cell. Mol. Life Sci.*, **57**, 106.
- 69 (a) Pratt, J.M. (2001) *Handbook on Metalloproteins* (ed. I. Bertini, A. Sigel, and H. Sigel), Marcel Dekker, New York,

- p. 603; (b) Brown, K.L. (2006) *Dalton Trans.*, 1123; (c) Kratky, C. and Gruber, C. (2001) *Handbok of Metalloproteins* (ed. A. Messerschmidt), John Wiley & Sons, Inc., New York, pp. 983–994, and references therein.
- 70 (a) Riordan, C.G. (2004) *Comprehensive Coordination Chemistry II*, vol. 8 (eds L. Que Jr. and W.B. Tolman), Elsevier, Oxford, p. 677; (b) Marzilli, L.G. (1999) *Bioinorganic Catalysis* (eds J. Reedijk and E. Bouwman), Marcel Dekker, New York, p. 423 and references therein.
- 71 (a) Hay, B.P. and Finke, R.G. (1986) *J. Am. Chem. Soc.*, **108**, 4820; (b) Finke, R.G. and Hay, B.P. (1988) *Polyhedron*, **7**, 1469; (c) Brown, K.L. and Zou, X. (1999) *J. Inorg. Biochem.*, **77**, 185; (d) Finke, R.G. (1998) *Vitamin B12 and the B12 Proteins* (eds B. Kräutler, D. Arigoni, and B.T. Golding), Wiley-VCH Verlag GmbH, Weinheim, pp. 383–402.
- 72 (a) Padmakumar, R., Padmakumar, R., and Banerjee, R. (1997) *Biochemistry*, **36**, 3713; (b) Buckel, W., Golding, B.T., and Kratky, C. (2006) *Chem.-Eur. J.*, **12**, 352; (c) Pratt, J.M. (1999) *Chemistry and Biochemistry of B12* (ed. R. Banerjee), John Wiley & Sons, Inc., New York, pp. 113–164; (d) Licht, S.S., Booker, S., and Stubbe, J. (1999) *Biochemistry*, **38**, 1221; (e) Chih, H.-W. and Marsh, E.N.G. (1999) *Biochemistry*, **38**, 13684; (f) Meier, T.W., Thoma, N.H., and Leadlay, P.F. (1996) *Biochemistry*, **35**, 11791; (g) Brown, K.L. and Li, J. (1998) *J. Am. Chem. Soc.*, **120**, 9466; (h) Licht, S.S., Lawrence, C.C., and Stubbe, J. (1999) *Biochemistry*, **38**, 1234; (i) Marsh, E.N.G. and Ballou, D.P. (1998) *Biochemistry*, **37**, 11864.
- 73 (a) Dybala-Defratyka, A., Paneth, P., Banerjee, R., and Truhlar, D.G. (2007) *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 10774; (b) Kozłowski, P.M., Kamachi, T., Toraya, T., and Yoshizawa, K. (2007) *Angew. Chem., Int. Ed.*, **46**, 980; (c) Sharma, P.K., Chu, Z.T., Olsson, M.H.M., and Warshel, A. (2007) *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 9661; (d) Jensen, K.P. and Ryde, U. (2005) *J. Am. Chem. Soc.*, **127**, 9117; (e) Dölker, N., Maseras, F., and Siegbahn, P.E.M. (2004) *Chem. Phys. Lett.*, **386**, 174; (f) Freindorf, M. and Kozłowski, P.M. (2004) *J. Am. Chem. Soc.*, **126**, 1928; (g) Kozłowski, P.M. and Zgierski, M.Z. (2004) *J. Phys. Chem. B*, **108**, 14163; (h) Kuta, J., Patchkovskii, S., Zgierski, M.Z., and Kozłowski, P.M. (2006) *J. Comput. Chem.*, **27**, 1429; (i) Banerjee, R., Truhlar, D.G., Dybala-Defratyka, A., and Paneth, P. (2007) *Hydrogen Transfer Reactions* (eds J.T. Hynes, J.P. Klinman, H.-H. Limbach, and R.L. Schowen), Wiley-VCH Verlag GmbH, Weinheim, pp. 1473–1495; (j) Banerjee, R., Dybala-Defratyka, A., and Paneth, P. (2006) *Philos. Trans. R. Soc. London, B: Biol. Sci.*, **361**, 1333; (k) Dybala-Defratyka, A. and Paneth, P. (2001) *J. Inorg. Biochem.*, **86**, 681; (l) Rovira, C. and Kozłowski, P.M. (2007) *J. Phys. Chem. B*, **111**, 3251; (m) Kozłowski, P.M., Andruniow, T., Jarzecki, A.A., Zgierski, M.Z., and Spiro, T.G. (2006) *Inorg. Chem.*, **45**, 5585; (n) Jensen, K.P. and Ryde, U. (2003) *J. Phys. Chem. A*, **107**, 7539; (o) Andruniow, T., Zgierski, M.Z., and Kozłowski, P.M. (2001) *J. Am. Chem. Soc.*, **123**, 2679; (p) Brown, K.L. and Marques, H.M. (2005) *J. Mol. Struct. (THEOCHEM)*, **714**, 209; (q) Brown, K.L. and Marques, H.M. (2001) *J. Inorg. Biochem.*, **83**, 121 and references therein.
- 74 Kwiecien, R.A., Khavrutskii, I.V., Musaev, D.G., Morokuma, K., Banerjee, R., and Paneth, P. (2006) *J. Am. Chem. Soc.*, **128**, 1287.
- 75 (a) Sandala, G.M., Smith, D.M., Marsh, E.N.G., and Radom, L. (2007) *J. Am. Chem. Soc.*, **129**, 1623; (b) Sandala, G.M., Smith, D.M., and Radom, L. (2006) *J. Am. Chem. Soc.*, **128**, 16004; (c) Sandala, G.M., Smith, D.M., Coote, M.L., Golding, B.T., and Radom, L. (2006) *J. Am. Chem. Soc.*, **128**, 3433; (d) Wetmore, S.D., Smith, D.M., Bennett, J.T., and Radom, L. (2004) *J. Am. Chem. Soc.*, **126**, 14054; (e) Wetmore, S.D., Smith, D.M., Bennett, J.T., and Radom, L. (2001) *ChemBioChem*, **2**, 919; (f) Wetmore, S.D., Smith, D.M., Golding, B.T., and Radom, L. (2001) *J. Am. Chem. Soc.*, **123**, 7963; (g) Kamachi, T., Toraya, T., and Yoshizawa, K. (2007) *Chem.-Eur. J.*, **13**, 7864; (h) Kamachi, T., Toraya, T., and Yoshizawa, K. (2004) *J. Am. Chem. Soc.*, **126**, 16207; (i) Loferer, M.J.,

- Webb, B.M., Grant, G.H., and Liedl, K.R. (2003) *J. Am. Chem. Soc.*, **125**, 1072 and references therein.
- 76 Li, X., Chung, L.W., Paneth, P., and Morokuma, K. (2009) *J. Am. Chem. Soc.*, **131**, 5115.
- 77 Gruber, K., Reitzer, R., and Kratky, C. (2001) *Angew. Chem., Int. Ed.*, **40**, 3377.
- 78 Shimomura, O., Johnson, F.H., and Saiga, Y. (1962) *J. Cell. Comp. Physiol.*, **59**, 223.
- 79 Selected theoretical works: (a) Martin, M.E., Negri, F., and Olivucci, M. (2004) *J. Am. Chem. Soc.*, **126**, 5452; (b) Altoe, P., Bernardi, F., Garavelli, M., Orlandi, G., and Negri, F. (2005) *J. Am. Chem. Soc.*, **127**, 3952; (c) Sinicropi, A., Andruniow, T., Ferre, N., Basosi, R., and Olivucci, M. (2005) *J. Am. Chem. Soc.*, **127**, 11534; (d) Weber, W., Helms, V., Mcammon, J.A., and Langhoff, P.W. (1999) *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 6177; (e) Olsen, S. and Smith, S.C. (2007) *J. Am. Chem. Soc.*, **129**, 2054; (f) Olsen, S. and Smith, S.C. (2008) *J. Am. Chem. Soc.*, **130**, 8677; (h) Toniolo, A., Olsen, S., Manohar, L., and Martinez, T.J. (2004) *Faraday Discuss.*, **127**, 149; (j) Das, A.K., Hasegawa, J., Miyahara, T., Ehara, M., and Nakatsuji, H. (2003) *J. Comput. Chem.*, **24**, 1421; (k) Demachy, I., Ridard, J., Laguitton-Pasquier, H., Durnerin, E., Vallverdu, G., Archirel, P., and Levy, B. (2005) *J. Phys. Chem. B*, **109**, 24121; (l) Nifosi, R., Amat, P., and Tozzini, V. (2007) *J. Comput. Chem.*, **28**, 2366; (m) Bravaya, K.B., Bochenkova, A.V., Granovsky, A.A., Savitsky, A.P., and Nemukhin, A.V. (2008) *J. Phys. Chem. A*, **112**, 8804; (n) Nemukhin, A.V., Topol, I.A., and Burt, S.K. (2006) *J. Chem. Theory Comput.*, **2**, 292; (o) Amat, P., Granucci, G., Buda, F., Persico, M., and Tozzini, V. (2006) *J. Phys. Chem. B*, **110**, 9348; (p) Voityuk, A.A., Kummer, A.D., Michel-Beyerle, M.-E., and Rosch, N. (2001) *Chem. Phys.*, **269**, 83.
- 80 (a) Brejc, K., Sixma, T.K., Kitts, P.A., Kain, S.R., Tsien, R.Y., Ormoe, M., and Remington, S.J. (1997) *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 2306; (b) Heim, R., Prasher, D.C., and Tsien, R.Y. (1994) *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 12501;
- (c) Palm, G.J., Zdanov, A., Gaitanaris, G.A., Stauber, R., Pavlakis, G.N., and Wlodawer, A. (1997) *Nat. Struct. Biol.*, **4**, 361; (d) Chattoraj, M., King, B.A., Bubltz, G.U., and Boxer, S.G. (1996) *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 8362; (e) Stoner-Ma, D., Jaye, A.A., Matousek, P., Towrie, M., Meech, S.R., and Tonge, P.J. (2005) *J. Am. Chem. Soc.*, **127**, 2864.
- 81 Dronpa: (a) Ando, R., Mizuno, H., and Miyawaki, A. (2004) *Science*, **306**, 1370; (b) Habuchi, S., Ando, R., Dedecker, P., Verheijen, W., Mizuno, H., Miyawaki, A., and Hofkens, J. (2005) *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 9511; (c) Dedecker, P., Hotta, J., Ando, R., Miyawaki, A., Engelborghs, Y., and Hofkens, J. (2006) *Biophys. J.*, **91**, L45; (d) Habuchi, S., Dedecker, P., Hotta, J.I., Flors, C., Ando, R., Mizuno, H., Miyawaki, A., and Hofkens, J. (2006) *Photochem. Photobiol. Sci.*, **5**, 567; (e) Fron, E., Flors, C., Schweitzer, G., Habuchi, S., Ando, R., De Schryver, F.C., Miyawaki, A., and Hofkens, J. (2007) *J. Am. Chem. Soc.*, **129**, 4870.
- 82 (a) Ando, R., Hama, H., Yamamoto-Hino, M., Mizuno, H., and Miyawaki, A. (2002) *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 12651; (b) Wiedenmann, J., Ivanchenko, S., Oswald, F., Schmitt, F., Röcker, C., Salih, A., Spindler, K.D., and Nienhaus, G.U. (2004) *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 15905; (c) Nienhaus, G.U., Wiedenmann, J., and Nar, H. (2005) *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 9156; (d) Gurskaya, N.G., Verkhusha, V.V., Shcheglov, A.S., Staroverov, D.B., Chepurnykh, T.V., Fradkov, A.F., Lukyanov, S., and Lukyanov, K.A. (2006) *Nat. Biotechnol.*, **24**, 461; (e) Patterson, G.H. and Lippincott-Schwartz, J. (2002) *Science*, **297**, 1873; (f) Mizuno, H., Mal, T.K., Tong, K.I., Ando, R., Furuta, T., Ikura, M., and Miyawaki, A. (2003) *Mol. Cell*, **12**, 1051.
- 83 (a) Chudakov, D.M., Belousov, V.V., Zarsky, A.G., Novoselov, V.V., Staroverov, D.B., Zorov, D.B., Lukyanov, S., and Lukyanov, K.A. (2003) *Nat. Biotechnol.*, **21**, 191; (b) Chudakov, D.M., Feofanov, A.V., Mudrik, N.N., Lukyanov, S., and Lukyanov, K.A. (2003) *J. Biol.*

- Chem.*, **278**, 7215; (c) Chudakov, D.M., Verkhusha, V.V., Staroverov, D.B., Souslova, E.A., Lukyanov, S., and Lukyanov, K.A. (2004) *Nat. Biotechnol.*, **22**, 1435; (d) Miyawaki, A. (2004) *Nat. Biotechnol.*, **22**, 1374; (e) Verkhusha, V.V. and Lukyanov, K.A. (2004) *Nat. Biotechnol.*, **22**, 289; (f) Lukyanov, K.A., Fradkov, A.F., Gurskaya, N.G., Matz, M.V., Labas, Y.A., Savitskyi, A.P., Markelov, M.L., Zaraisky, A.G., Zhao, X.-N., Fang, Y., Tan, W.-Y., and Lukyanov, S.A. (2000) *J. Biol. Chem.*, **275**, 25879.
- 84 (a) Andresen, M., Wahl, M.C., Stiel, A.C., Gräter, F., Schäfer, L.V., Trowitzsch, S., Weber, G., Eggeling, C., Grubmüller, H., Hell, S.W., and Jakobs, S. (2005) *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 13070; (b) Henderson, N.J., Ai, H.-W., Campbell, R.E., and Remington, S.J. (2007) *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 6672; (c) Hofmann, M., Eggeling, C., Jakobs, S., and Hell, S.W. (2005) *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 17565; (d) Adam, V., Lelimosin, M., Boehme, S., Desfonds, G., Nienhaus, K., Field, M.J., Wiedenmann, J., McSweeney, S., Nienhaus, G.U., and Bourgeois, D. (2008) *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 18343.
- 85 Wilmann, P.G., Petersen, J., Devenish, R.J., Prescott, M., and Rossjohn, J. (2005) *J. Biol. Chem.*, **280**, 2401.
- 86 Sauer, M. (2005) *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 9433.
- 87 (a) Vendrell, O., Gelabert, R., Moreno, M., and Lluch, J.M. (2006) *J. Am. Chem. Soc.*, **128**, 3564; (b) Vendrell, O., Gelabert, R., Moreno, M., and Lluch, J.M. (2008) *J. Phys. Chem. B*, **112**, 5500; (c) Vendrell, O., Gelabert, R., Moreno, M., and Lluch, J.M. (2008) *J. Chem. Theory Comput.*, **4**, 1138; (d) Lill, M.A. and Helms, V. (2002) *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 2778.
- 88 (a) Tanner, C., Manca, C., and Leutwyler, S. (2003) *Science*, **302**, 1736; (b) Ashfold, M.N.R., Cronin, B., Devine, A.L., Dixon, R.N., and Nix, M.G.D. (2006) *Science*, **312**, 1637; (c) Sobolewski, A.L. and Domcke, W. (1999) *J. Phys. Chem. A*, **103**, 4494.
- 89 (a) Ando, R., Flors, C., Mizuno, H., Hofkens, J., and Miyawaki, A. (2007) *Biophys. J.*, **92**, L97; (b) Flors, C., Hotta, J.-I., Uji-i, H., Dedeker, P., Ando, R., Mizuno, H., Miyawaki, A., and Hofkens, J. (2007) *J. Am. Chem. Soc.*, **129**, 13970.
- 90 bsDronpa and Padron, mutants of Dronpa: Andresen, M., Stiel, A.C., Fölling, J., Wenzel, D., Schönle, A., Egner, A., Eggeling, C., Hell, S.W., and Jakobs, S., (2008) *Nat. Biotechnol.*, **26**, 1035.
- 91 (a) Wilmann, P.G., Turcic, K., Battad, J.M., Wilce, M.C.J., Devenish, R.J., Prescott, M., and Rossjohn, J. (2006) *J. Mol. Biol.*, **364**, 213; (b) Stiel, A.C., Trowitzsch, S., Weber, G., Andresen, M., Eggeling, C., Hell, S.W., Jakobs, S., and Wahl, M.C. (2007) *Biochem. J.*, **402**, 35; (c) Nam, K.-H., Kwon, O.Y., Sugiyama, K., Lee, W.-H., Ki, Y.K., Song, H.K., Kim, E.E., Park, S.-Y., Jeon, H., and Hwang, K.S. (2007) *Biochem. Biophys. Res. Commun.*, **354**, 962; (d) Andresen, M., Stiel, A.C., Trowitzsch, S., Weber, G., Eggeling, C., Wahl, M.C., Hell, S.W., and Jakobs, S. (2007) *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 13005; (e) Mizuno, H., Kumar, M.T., Wälchli, M., Kikuchi, A., Fukano, T., Ando, R., Jeyakanthan, J., Taka, J., Shiro, Y., Ikura, M., and Miyawaki, A. (2008) *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 9927.
- 92 Li, X., Chung, L.W., Mizuno, H., Miyawaki, A., and Morokuma, K. submitted.
- 93 (a) Liu, R.S.H. (2001) *Acc. Chem. Res.*, **34**, 555; (b) Liu, R.S.H. and Hammond, G.S. (2000) *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 11153; (c) Liu, R.S.H. and Asato, A.E. (1985) *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 259.
- 94 (a) Schäfer, L.V., Groenhof, G., Boggio-Pasqua, M., Robb, M.A., and Grubmüller, H. (2008) *PLoS Comput. Biol.*, **4**, e1000034; (b) Schfer, L.V., Groenhof, G., Klingen, A.R., Ullmann, G.M., Boggio-Pasqua, M., Robb, M.A., and Grubmüller, H. (2007) *Angew. Chem. Int. Ed.*, **46**, 530; (c) Bravaya, K.B., Bochenkova, A.V., Granovsky, A.A., Savitsky, A.P., and Nemukhin, A.V. (2008) *J. Phys. Chem. A*, **112**, 8804; (d) Nemukhin, A.V., Topol, I.A., and Burt, S.K. (2006) *J. Chem. Theory Comput.*, **2**, 292.
- 95 (a) Hayashi, I., Mizuno, H., Tong, K.I., Furuta, T., Tanaka, F., Yoshimura, M.,

- Miyawaki, A., and Ikura, M. (2007) *J. Mol. Biol.*, **372**, 918 and references therein.
- 96 Li, X., Chung, L.W., Miyawaki, A., and Morokuma, K., manuscript in preparation.
- 97 Ando, Y., Niwa, K., Yamada, N., Enomoto, T., Irie, T., Kubota, H., Ohmiya, Y., and Akiyama, H. (2008) *Nat. Photonics*, **2**, 44.
- 98 Nakatsu, T., Ichiyama, S., Hiratake, J., Saldanha, A., Kobashi, N., Sakata, K., and Kato, H. (2006) *Nature*, **440**, 372.
- 99 (a) Turro, N.J. (1978) *Modern Molecular Photochemistry*, Benjamin/Cummings Publishing Co, Menlo Park, CA, pp. 597–611; (b) Adam, W. (1983) *The Chemistry of Functional Groups, Peroxides* (ed. S. Patai), John, Wiley & Sons, Inc., New York, pp. 830–920; (c) Adam, W. and Cilento, G. (1982) *Chemical and Biological Generation of Excited States*, Academic Press, New York; (d) Adam, W. and Baader, W.J. (1985) *J. Am. Chem. Soc.*, **107**, 410; (e) Adam, W. and Trofimov, A.V. (2006) *The Chemistry of Peroxides*, vol. 2 (ed. Z. Rappoport), John Wiley & Sons, Inc., Hoboken, pp. 1171–1209; (f) Matsumoto, M. (2004) *J. Photochem. Photobiol. C-Photochem. Rev.*, **5**, 27.
- 100 (a) Wilsey, S., Bernardi, F., Olivucci, M., Robb, M.A., Murphy, S., and Adam, W. (1999) *J. Phys. Chem. A*, **103**, 1669; (b) Tanaka, C. and Tanaka, J. (2000) *J. Phys. Chem. A*, **104**, 2078; (c) Rodríguez, E. and Reguero, M. (2002) *J. Phys. Chem. A*, **106**, 504; (d) De Vico, L., Liu, Y.-J., Krogh, J.W., and Lindh, R. (2007) *J. Phys. Chem. A*, **111**, 8013.
- 101 (a) Koo, J.-Y., Schmidt, S.P., and Schuster, G.B. (1978) *Proc. Natl. Acad. Sci. U.S.A.*, **75**, 30; (b) Zaklika, K.A., Thayer, A.L., and Schaap, A.P. (1978) *J. Am. Chem. Soc.*, **100**, 4916; (c) Baader, W.J., Stevani, C.V., and Bastos, E.L. (2006) *The Chemistry of Peroxides*, vol. 2 (ed. Z. Rappoport), John Wiley & Sons, Inc., Hoboken, pp. 1211–1278; (d) Catalani, L.H. and Wilson, T. (1989) *J. Am. Chem. Soc.*, **111**, 2633.
- 102 (a) Isobe, H., Takano, Y., Okumura, M., Kuramitsu, S., and Yamaguchi, K. (2005) *J. Am. Chem. Soc.*, **127**, 8667; (b) Carpenter, B.K. (2006) *Chem. Soc. Rev.*, **35**, 736; (c) Blancafort, L., Jolibois, F., Olivucci, M., and Robb, M.A. (2001) *J. Am. Chem. Soc.*, **123**, 722; (d) Atchity, G.J., Xantheas, S.S., and Ruedenberg, K. (1991) *J. Chem. Phys.*, **95**, 1862.
- 103 Chung, L.W., Hayashi, S., Lundberg, M., Nakatsu, T., Kato, H., and Morokuma, K. (2008) *J. Am. Chem. Soc.*, **130**, 12880.
- 104 (a) White, S.R. (1992) *Phys. Rev. Lett.*, **69**, 2863; (b) Chan, G.K.-L. and Head-Gordon, M. (2002) *J. Chem. Phys.*, **116**, 4462; (c) Ghosh, D., Hachmann, J., Yanai, T., and Chan, G.K.-L. (2008) *J. Chem. Phys.*, **128**, 144117.
- 105 (a) Semi-empirical methods. DFTB, Riccardi, D., Schaefer, P., Yang, Y., Yu, H., Ghosh, N., Prat-Resina, X., König, P., Li, G., Xu, D., Guo, H., Elstner, M., and Cui, Q. (2006) *J. Phys. Chem. B*, **110**, 6458, (b) PM6, Stewart, J.J.P. (2007) *J. Mol. Modeling*, **13**, 1173; (c) PDDG/PM3 and PDDG/MNDO, Repasky, M.P., Chandrasekhar, J., and Jorgensen, W.L. (2002) *J. Comput. Chem.*, **23**, 1601, (d) OM2, Weber, W. and Thiel, W. (2000) *Theor. Chem. Acc.*, **103**, 495, (e) FOMO–CI, Toniolo, A., Granucci, G., and Martinez, T.J. (2003) *J. Phys. Chem. A*, **107**, 3822, (f) MNDOC-CI, Klessinger, M., Pötter, T., and van Wüllen, C. (1991) *J. Theor. Chim. Acta*, **80**, 1.
- 106 Empirical methods. Molecular mechanics-valence bond (MMVB), (a) Bernardi, F., Olivucci, M., and Robb, M.A. (1992) *J. Am. Chem. Soc.*, **114**, 1606, (b) electron force field (eFF), Su, J.T. and Goddard, W.A. (2007) *Phys. Rev. Lett.*, **99**, 185003, (c) ReaxFF, van Duin, A.C.T., Dasgupta, S., Lorant, F., and Goddard, W.A. (2001) *J. Phys. Chem. A*, **105**, 9396.
- 107 Ufimtsev, I.S. and Martínez, T.J. (2008) *J. Chem. Theory Comp.*, **4**, 222; (b) Ufimtsev, I.S. and Martínez, T.J. (2009) *J. Chem. Theory Comp.*, **5**, 1004; (c) Friedrichs, M.S., Eastman, P., Vaidyanathan, V., Houston, M., LeGrand, S., Beberg, A.L., Ensign, D.L., Bruns, C.M., and Pande, V.S. (2009) *J. Comput. Chem.*, **30**, 864.
- 108 (a) Kühne, T.D., Krack, M., Mohamed, F.R., and Parrinello, M. (2007) *Phys. Rev. Lett.*, **98**, 66401; (b) Ensing, B., De Vivo, M., Liu, Z., Moore, P., and Klein, M.L. (2006) *Acc. Chem. Res.*, **39**, 73; (c) Laio, A. and Parrinello, M. (2002) *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 12562; (d) Car, R. and

- Parrinello, M. (1985) *Phys. Rev. Lett.*, **55**, 2471; (e) Yuji, S. and Okamoto, Y. (1999) *Chem. Phys. Lett.*, **314**, 141.
- 109 Zhang, Y., Liu, H., and Yang, W. (2000) *J. Chem. Phys.*, **112**, 3483.
- 110 Hu, H., Lu, Z., and Yang, W. (2007) *J. Chem. Theory Comput.*, **3**, 390.
- 111 (a) Hayik, S.A., Liao, N., and Merz, K.M. Jr. (2008) *J. Chem. Theory Comput.*, **4**, 1200; (b) Kaukonen, M., Söderhjelm, P., Heimdal, J., and Ryde, U. (2008) *J. Phys. Chem. B*, **112**, 12537.
- 112 (a) Schaefer, P., Riccardi, D., and Cui, Q. (2005) *J. Chem. Phys.*, **123**, 14905; (b) Benighaus, T. and Thiel, W. (2008) *J. Chem. Theory Comput.*, **4**, 1600.
- 113 Pellegrini, E. and Field, M.J. (2002) *J. Phys. Chem. A*, **106**, 1316.
- 114 Voth, G.A. (2008) *Coarse-Graining of Condensed Phase and Biomolecular Systems*, CRC Press, Boca Raton, FL.
- 115 (a) Rothlisberger, D., Khersonsky, O., Wollacott, A.M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J.L., Althoff, E.A., Zanghellini, A., Dym, O., Albeck, S., Houk, K.N., Tawfik, D.S., and Baker, D. (2008) *Nature*, **453**, 190; (b) Jiang, L., Althoff, E.A., Clemente, F.R., Doyle, L., Rothlisberger, D., Zanghellini, A., Gallaher, J.L., Betker, J.L., Tanaka, F., Barbas, C.F. III, Hilvert, D., Houk, K.N., Stoddard, B.L., and Baker, D. (2008) *Science*, **319**, 1387; (c) Bender, G.M., Lehmann, A., Zou, H., Cheng, H., Fry, H.C., Engel, D., Therien, M.J., Kent Blasie, J., Roder, H., Saven, J.G., and DeGrado, W.F. (2007) *J. Am. Chem. Soc.*, **129**, 10732; (d) Wade, H., Stayrook, S.E., and DeGrado, W.F. (2006) *Angew Chem. Int. Ed.*, **45**, 4951; (e) Cristian, L., Piotrowiak, P., and Farid, R.S. (2003) *J. Am. Chem. Soc.*, **125**, 11814.

4

From Molecular Electrostatic Potentials to Solvation Models and Ending with Biomolecular Photophysical Processes

Jacopo Tomasi, Chiara Cappelli, Benedetta Mennucci, and Roberto Cammi

4.1

Introduction

The interest of our group in quantum biology dates back, for the older author of this chapter (J. T.), to the early 1960s.

Our group, located in Pisa, was developing, among the first in the world, computational codes for the *ab initio* calculation of molecular wavefunctions using electronic computers.

Among the topics selected to explore the possibilities of this new instrument were the noncovalent interactions among organic molecules, having in view the fascinating, and at that time almost completely unexplored, world of molecular interactions within living bodies.

The interesting results obtained in these first years of activity prompted our small group, of varying composition, to explore other themes, all starting from the original one, some of which have connections with quantum biology.

We present here a short synopsis of a selection of some these themes having a logic sequential correlation and a connection with quantum biology. For brevity we have discarded other topics of our activity related to those here presented. We cite among them the counterpoise corrections, static description of some dynamical solvent effects in chemical reactions (fluctuations, solvent delays), and response properties of composite systems (metal nanoparticle/organic chromophore).

We hope that what is presented here is sufficient to illustrate one among the many research lines thus far developed to gain a better understanding of chemical and biological systems.

4.2

The Molecular Electrostatic Potential and Noncovalent Interactions among Molecules

4.2.1

Molecular Electrostatic Potential

The starting point of our exposition is the molecular electrostatic potential (MEP), which was introduced in the late 1960s [1].

In standard quantum mechanical (QM) calculations, a time-independent description of the electron charge distribution is easily obtained from the wavefunction Ψ , which is the solution of the Schrödinger equation for a given electronic state of the molecule. In fact in the QM theory the information is carried by the product $\Psi^*\Psi$, a function defined in the N -dimensional space (N is the number of electrons). This information is, however, redundant, and it can be reduced by integrating $\Psi^*\Psi$ over a number of coordinates. When the integration is extended to $3N - 3$ coordinates we obtain a function only depending on the three coordinates (the vector position $\mathbf{r}$) of an electron.

This quantity, which we shall indicate with $\varrho^{\text{el}}(\mathbf{r})$, is the molecular electron density. In QM theory $\varrho^{\text{el}}(\mathbf{r})$ is a weighted probability function. The value of this function at a given value of the variable $\mathbf{r}$ gives the probability of finding electrons within an infinitesimal volume centered at $\mathbf{r}$ weighted by the number of electrons.

The electron density function has many interesting properties, of which we recall just one because it will be used in the following discussion: when the wavefunction is expressed in terms of a single determinant function (Hartree–Fock, HF, or Kohn–Sham, KS) the electron density can be exactly partitioned into a sum of single molecular orbital (MO) contributions:

$$\varrho^{\text{el}}(\mathbf{r}) = \sum_{i=1}^n \phi_i^2(\mathbf{r}) \quad (4.1)$$

For a full description of the MEP we have to introduce the nuclear charge distribution. In standard QM calculations the nuclei are kept fixed at given positions in the 3D space.

It is formally convenient to define the nuclear distribution as an analytical expression with explicit dependence on the position vector $\mathbf{r}$, similar to its electronic counterpart. In this case:

$$\varrho^{\text{tot}}(\mathbf{r}) = -\varrho^{\text{el}}(\mathbf{r}) + \varrho^{\text{nuc}}(\mathbf{r}) = -\varrho^{\text{el}}(\mathbf{r}) + \sum_{\alpha} Z_{\alpha} \delta(\mathbf{r} - \mathbf{R}_{\alpha}) \quad (4.2)$$

The vector $\mathbf{R}_{\alpha}$ defines the position of nucleus α with charge Z_{α} (in au); $\delta(\mathbf{r} - \mathbf{R}_{\alpha})$ is the Dirac function.

Once the total density has been defined, the MEP has the following expression:

$$V(\mathbf{r}) = \int \frac{\varrho^{\text{tot}}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \quad (4.3)$$

The integration shows that in all points of the space the MEP is given by a sum of contributions from all the local charges components, each scaled by a factor equal to

the reciprocal of the distance and point of application. The linear scaling of the Coulomb interaction has to be compared with the scaling law of the other components of interactions in molecules. All scale with a higher negative power of the distance and have all high local character. The MEP is strongly non-local.

The first calculations of MEP maps, performed in Pisa, showed a clear and detailed description of chemical groups: lone pairs, C=C bonds and aromatic delocalization – all concepts empirically well known but never observed before. Later, it was observed that these local characteristics were to a fairly good extent due to contributions of distant groups [2]. In exceptional cases distant contributions were able to reverse the usual characteristics of a group, as in the case of organic acids bearing the C=C group [3].

4.2.1.1 Use of MEP

The applications of MEP in molecular sciences are numerous [4]. It is not possible to give here a survey, and so we limit ourselves to the applications pioneered by our group and having relevance within quantum biochemistry (QB). The MEP has been also acknowledged as the more important intellectual concept, with the most important impact on the development of QM in its first 30 years [5].

There were, essentially, two motivations that led us to introduce the concept of MEP into theoretical chemistry: to acquire tools able to analyze and understand interactions within a molecule, and to obtain more accurate descriptions of the interaction between molecules.

We shall give only limited space to the first motivation, which has led to the formulation of the so-called semiclassical approximation, and dedicate instead more attention to the second, which, *inter alia*, has paved the way to the formulation of methods for the description of solvent effects.

4.2.1.2 Semiclassical Approximation

The MEP is a quantity of quantum origin, which behaves as its classical counterpart. Other factors playing a role in the theory of molecular structure, for example, polarization and forces, have the same correspondence between quantum and classical behavior.

We originally formulated and then tested the hypothesis that a description based on classical concepts would have given a first-order description of the phenomena of interest in the molecular domain. The scope of using this approximation was the interpretation of QM results, the recognition of effects not easily reduced to a classical counterpart, and, in the case of positive results, the formulation of computational methods applicable to molecules too large to be studied with a full QM procedure.

The first step in this planned work was the definition of the starting electron densities. It was found useful to resort to a quantum description of the electron density of the molecules selected as prototypes, and to decompose them into their MO components according to Equation 4.1.

The MO description of ρ^{el} has the noticeable feature of being invariant under a linear transformation of the MOs. Thus, localized orbitals (LOs) were selected as

starting point for the analysis; the Boys localization [6] was used, but this choice is not compelling, at present the localization accompanying the natural bond orbital (NBO) method [7] is more extensively used.

The appreciably high degree of transferability of LOs between molecules was extensively measured and considered to be sufficient to define a first-order procedure to describe unknown molecular charge distributions by a simple juxtaposition of molecular fragments taken from a library. Refinements to this description are given by polarization effects (through bonds and through space effects) and by forces inducing changes in the conformation and bond lengths.

4.2.1.3 MEP as a Component of the Intermolecular Interaction

The Coulomb term plays a role in intermolecular interactions in biological systems. This was well known even before the beginning of our activity, but we were not satisfied by the almost general consensus that a more rigorous study of these interactions had to remain in the use of perturbation theory (PT) approaches. The original PT was conceived when QM calculations on molecules were not possible and, to circumvent this obstacle, PT was expressed in a form not utilizing wavefunctions and energies of the two interaction partners A and B; only the interaction energy ΔE_{AB} was the quantity to be determined, by means of an expansion series with elements separately computed in terms of experimentally determined properties of the monomers.

We addressed first the first-order term of this expansion, the Coulomb interaction energy, E_{coul} . There were insufficient experimental values to determine this quantity via the suggested polar expansion and, even with a quite improbable sudden flow of experimental high order polar terms, there was little hope of reaching convergence for molecules with an irregular shape. It surely was more effective to use the information coming from the computer simulation of electron densities directly, without any expansion. This was a good reason for examining the performance of a MEP-based description of this term.

A little later it was argued that, because the *ab initio* approach of molecular calculations was showing very good perspectives of accurate calculations, the time was right to devise an alternative strategy to compute weak molecular interactions and to give at the same time a detailed decomposition, in perspective more accurate than that given by PT.

Let us start with the Coulomb term, by also including some simplifications of it. The complete decomposition of ΔE_{AB} will be discussed in the following.

Before starting the discussion on this point, we return to PT. This method has been greatly changed in more recent years, by introducing QM calculations on the monomers (also using the complete basis set of the dimer for each monomer) to compute the elements of the PT expansion, so that a simplification present in the original version has been abandoned [the neglecting of consideration of a complete (anti)symmetry description of the electron distribution].

The new versions, known as SAPT (symmetry-adapted PT) [8] are in use with satisfactory results, especially for molecules not too large.

4.2.1.4 Definition of the Coulomb Interaction Term

The term E_{coul} is calculated in complete analogy with the classical description:

$$E_{\text{coul}} = \int \int_{r_1 r_2} \frac{Q_A^{\text{tot}}(\mathbf{r}_1) Q_B^{\text{tot}}(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \quad (4.4)$$

This integral can be rewritten in terms of the MEP of one of the partners and the charge distribution of the second:

$$E_{\text{coul}} = \int_{r_2} V_A(\mathbf{r}_2) Q_B^{\text{tot}}(\mathbf{r}_2) d\mathbf{r}_2 \quad (4.5)$$

Expression (4.5), which is exact, can be simplified with various approximations of decreasing accuracy and in parallel with decreasing computational cost.

4.2.1.5 Simplifications in the Expression of E_{es} : Point Charge Descriptions

The charge distribution of B can be very well represented by a set of point charges q_{kB} placed at suitably selected r_K points. The expression of E_{coul} then assumes the simple form of a limited sum of MEP values, each multiplied by a weight factor:

$$E_{\text{coul}} = \sum_k V_A(k) q_{kB} \quad (4.6)$$

The selection of the point charges q_{kB} is more easily done starting from the MEP of B. Also, the MEP of the partner, V_A , can be simplified in a similar manner. The electrostatic interaction is so reduced to a finite sum of products with each involving the calculation of a distance.

Probably the first elaboration of point models addressing the calculation of E_{es} was carried out in our laboratory [9, 10]. For a molecule with n valence electrons the number of negative point charges is a little less than $2n$, being the couple of electrons involved in a bond described by two charges and the couple describing a lone pair by a charge alone. These charges, bearing a negative value, are to be supplemented by m positive charges placed on the nuclei, with value equal to the formal nuclear charge screened by the core electrons. The procedure for obtaining the charges is based on a description of the electronic charge distribution of the molecule expressed in terms of LOs. The position of the couple of charges for a bond is based on the location of the charge center of this LO. Actually LOs, when the delocalized tails are cancelled, have a great degree of transferability from molecule to molecule and within the same molecule. This approach permits the building of libraries of LOs and of point charges exhibiting an appreciably low deviation with respect to full QM calculations (see the preceding section on the semiclassical approach).

In the above short description a way to obtain the parameters for simplified non-QM procedures is reported, and this is an alternative to extensive fittings on training sets of molecules. In current opinion, the larger the training set, and more refined the fitting, the more reliable are the parameters. The fitting of training sets may be considered a brute force approach in which no attention is paid to some features of the systems of interest, which can be exploited in a different way.

4.2.1.6 Simplifications in the Expression of E_{es} : Atomic Charges

The simplest formulation of point charge models is based on atomic charges. Atomic charges in molecules, which are considered not to have a well-defined quantum mechanical status, have had, and continue to have, great popularity in theoretical and computational chemistry. The non-observable nature of atomic charges in molecules is the traditional view in quantum chemistry, but there are now good reasons to reconsider this belief [11a,b]. Atomic charges were the first molecular index drawn from QM calculations (in Mulliken's definition [12]) and continue to-day to be redefined and used in a large variety of ways. Atomic charges in quantum biochemistry deserve a dedicated chapter of sizeable dimension; here we limit ourselves to consider atomic charges defined by numerical fitting of the MEP in the periphery of the molecule. The first to exploit this idea was Momany in 1978 [13]. The atomic charges defined in this way are often called potential derived atomic charges (PDAC). They rapidly gained wide popularity, with many proposals for refinements in their definition. The material was considered sufficient, ten years later, for a detailed review [14] and other reviews followed. We cite one among them, by our group, where several key topics treated in this chapter are collected with comments complementing what is here briefly reported [15].

4.2.1.7 Simplifications in the Expression of E_{es} : Multipolar Expansions

Expansion of the potential (and of the charge distribution) into spherical harmonics dates back to the first years of quantum mechanics and it has been a basic tool in the physical interpretation of weak noncovalent interactions. The expansions used in these basic studies were of one-center type, in other words all the expansion functions were all centered at the same point. The expression of this expansion is compact:

$$V_A^{\text{oc}}(r) = \sum_{l=1}^{\infty} \sum_{m=-l}^l Q_l^m r^{l+1} Y_l^m(\theta, \phi) \quad (4.7)$$

The superscript oc indicates that the expression corresponds to a one-center expansion; $Y_l^m(\theta, \phi)$ s are the harmonic functions (expressed in polar coordinates) and Q_l^m s are the elements of the sequence of the multipolar coefficients. The multipolar coefficients are each the QM expectation value of a corresponding operator (dipole, quadrupole, etc., in increasing order). The expansion does not depend on the chosen position of the expansion center.

The one-center expansion is still in use for small molecules, but it has been abandoned in favor of multi-center expansions for larger molecules. In fact, the expansion theorem holds for points r lying outside a sphere containing all the elements of the charge distribution of the molecule. $V_A^{\text{oc}}(r)$ has a correct asymptotic behavior; when the number of elements in a truncated expression is kept fixed, the description improves at larger distances from the expansion origin. $V_A^{\text{oc}}(r)$ is also convergent at large distances. Asymptoticity and convergence are not sufficient to ensure good results at the short distances of chemical interest; there, the active region of the molecule is often surrounded by peripheral groups, lying at r distances larger than that of the reactive position.

The adoption of a multi-center expansion means that the molecular charge is divided into local fragments, each expanded separately. The sphere for each fragment is smaller than that of the whole molecule and it is relatively easy to find a definition of fragments giving a reasonable description of the potential in the molecular crevices of interest. This said, the number of local expansions, their location and the number of truncated expansion terms for each center have to be defined. The definition of the fragmentation is an arbitrary act, because in the QM description of the electron distribution there are no spatially separated groups. There are of course methods giving approximated partitions (localized orbitals, atomic basins in the AIM method [11c,d]) with different performances with respect to the convergence of truncated expansions. The decay with the distance of the single contributions is faster for higher values of the angular momentum, l , and thus the contribution of distant fragments can be described with shorter expansions.

Clearly the definition of the most convenient multi-center expansion requires attention and checks, but the results may be rewarding.

4.2.2

Interaction Energy between Two Molecules

Interactions among molecules are a basic element in the whole field of chemistry and related fields. Only a part of this extremely varied category of phenomena is interesting in biochemistry, and within this specific field we limit ourselves to the noncovalent interactions between a couple of molecules. The interactions we shall consider have a wide range of stabilization energies, about two orders of magnitude (the only exception considered here being the proton affinity, which is larger). In other sections of this chapter other interactions, leading to the formation of covalent bonds, are considered, and in other chapters of this book this topic is considered again, from other viewpoints.

The theory of weak molecular interactions is relatively well understood also in fine detail (topics still obscure are not of interest for biochemistry), but actually it is not so simple as to be applied in its completeness.

It is widely accepted that accurate descriptions of weak noncovalent interactions require variational calculations performed with QM levels and large basis sets – a computational level not accessible yet for the large molecules of interest in biochemistry. In this field, acceptable quality is given by MP2 or DFT procedures with appropriate functionals (HF calculations are still in use). We shall limit ourselves to consider applications of these methods to weak interactions. MP2 and DFT calculations, compared to the HF ones, add a portion of the dynamical correlation and of dispersion effects. DFT calculations on single molecules are in general a little better than the MP2 ones, but only a portion of the DFT functionals describes passably well weak interactions, and appropriate functional and empirical corrections of dispersive type are often recommended [16].

This short summary of the state of the art for the calculations of weak interactions for biomolecules shows that very accurate interaction energies are not expected. Even

if not perfect, HF, DFT and MP2 calculations, accompanied by a proper decomposition, are useful, giving a reliable appraisal of the specificities of the interaction and giving elements for the definition of force field parameters.

The decomposition is obtained as follows; the method has been devised by Morokuma and elaborated by Kitaura and Morokuma [17]. A similar decomposition was also elaborated by our group in the same years, see Reference [18] for a comparison between the two approaches and for an account of the friendly collaboration between Morokuma's and Pisa's groups.

The Coulomb, polarization, repulsion and charge-transfer components are separately computed (a dispersion contribution can also be computed). The variational method gives the whole value of ΔE_{AB} , and so the four (or five) components give a residuum of the whole interaction not subjected to decomposition. The whole formula reads:

$$\Delta E_{AB} = E_{AB} - (E_A + E_B) = E_{\text{coul}} + E_{\text{pol}} + E_{\text{exc}} + E_{\text{ctr}} + E_{\text{dis}} + E_{\text{res}} \quad (4.8)$$

Let us look at the meaning of these terms:

- 1) E_{coul} has been already examined. In this dissection of ΔE_{AB} , it is computed by collecting in the supermolecule calculation the interactions among the occupied orbitals of the two partners, without allowing for deformation.
- 2) E_{pol} gives a second electrostatic contribution corresponding to the mutual polarization of the two monomer's charge distribution under the effect of the electric field of the partner. Its value (always negative, i.e., attractive) is obtained by making an additional calculation on the two monomers, each with its own basis set, but feeling the effect of the electrostatic field of the partner.

E_{coul} and E_{pol} are of semiclassical nature and with this procedure are computed exactly (within the limitations due to the basis set and to the QM level adopted for the calculation). In other words, no use is made of truncated expansions both in the property and in the PT orders.

- 3) E_{exc} takes into account the repulsive effects due to the overlap of the charge distributions of the two monomers. It is a non-classical component having its origin in the Pauli exclusion principle, which does not permit multiple occupation of fermionic (i.e., electronic) functions. It is calculated by repeating the calculation of the expectation energy of the dimer using the occupied molecular orbitals of the monomers after the introduction of a complete antisymmetrization among them. This calculation is quite easy with one-determinant (HF or KS) wavefunctions.
- 4) E_{ctr} is computed by repeating the calculation on both monomers, thus allowing for a mixing of the occupied orbital of A with the empty ones of B, and symmetrically for the second partner. This contribution is interpreted as charge transfer from a monomer to the other.
- 5) E_{dis} is an attractive term interpreted as the interaction among the mutual fluctuations of the two charge distributions. It is formally described as temporary excitations of each monomer to its virtual space. An approximation to it can be obtained in the MP2 framework. Single determinant descriptions do not give it.

PT descriptions only separately give the first term of the dispersion expansion series, the following terms being present but mixed with contributions of different nature.

- 6) E_{res} is the difference between ΔE_{AB} and the sum of the preceding contributions (the dispersion contribution, not present in almost all decomposition schemes, is empirically described, and not included in the calculation of E_{res}). Actually this residuum, which is generally small, can be further partitioned, with two pairs of terms describing additional couplings between orbitals and leaving a smaller residual term. This additional partition has been rarely used.

Formulas, graphic descriptions, and so on of the several orbital couplings used in the method, as well as additional features, are available in numerous papers and monographs.

4.2.3

Examples of Energy Decomposition Analyses

The variety of molecular interactions occurring in biochemistry is remarkably large and for this reason we are compelled to limit ourselves to the consideration of a few classes. Limitations of space also suggest adopting a verbal presentation, avoiding numerical tables and graphs.

4.2.3.1 Interactions with a Proton

This is formally the simplest example of molecular interaction. We have already remarked the MEP gives information about the energetically preferred positions of a proton and about the relative strength of the various protonation sites of a molecule. However, this simple model neglects other important contributions to the energy of the process.

The proton produces a strong electric field, which polarizes the molecule; E_{pol} will be large. At the same time the proton, bare of any electronic charge, will attract electrons from the molecule: also the E_{ctr} will be large; this interaction, which is active at short distances, leads to the formation of a new covalent bond. The other components of ΔE , see Equation 4.8, are zero or negligible.

Equation 4.8 regards, however, molecular interactions at fixed internal geometry; the field produced by the proton is strong enough to produce changes in the internal geometry of the proton acceptor, which influence the proton affinity value.

4.2.3.2 Interactions with Other Cations

A contribution not playing a role in proton interactions has here an important role. The nucleus in atomic cations is encircled by a shell of electrons, which activate the repulsion contribution. The interaction at large distances is still ruled out by the coulomb interaction (the MEP continues to be a good indicator), but the equilibrium position is reached at a larger distance than in the case of a proton. There is no formation of covalent bonds and the effect on the internal geometry is modest.

4.2.3.3 Hydrogen Bonding

Hydrogen bonding interactions in neutral systems are ruled out in the approaching path by an electrostatic interaction fairly well represented by a dipolar term; this indicator is easily derived from MEP. At shorter distances, E_{coul} continues to dominate, but the dipolar term is no longer sufficient, so that the whole electrostatic potential (or a faithful point description of it) is necessary. Near to the equilibrium distance, E_{pol} is accompanied by rapidly increasing repulsion and dispersion contributions. At equilibrium distance, the short distance repulsion and dispersion terms (the first positive, i.e., repulsive, the second negative), when combined with the polarization term (which has an intermediate decay and is everywhere negative), cancel out each other at a good extent, leading the $\Delta E(R_{\text{eq}})$ value to almost $E_{\text{coul}}(R_{\text{eq}})$. This is why hydrogen bonding energies have been calculated for many years by using the Coulomb term alone, as if it would be a purely electrostatic interaction. The interaction is not purely electrostatic, as the success of these applications prompted researchers to consider. Actually, all the terms play a role in defining ΔE all over the space.

Many other types of interactions could be defined and characterized in such a simple way (some are more complex, however), but we have limited ourselves to the most important cases, to which we have contributed in the past to establishing properties and methods of use. References [19, 20] detail some of our past work.

The primary scope of the analyses we have briefly touched on was the advance in understanding of molecular interactions; notably, the remarkable improvement of computational methods have not changed the basic aspects of the initial analyses. More refined QM calculations have been, however, instrumental in increasing the accuracy necessary for reliable applications (the so-called chemical accuracy). Among the applications of biochemical interest we will treat the definition of force fields for computer simulations. The next section also acts as an introduction to the analysis of computational models to describe solvent effects on energies, properties and reactivity of molecular systems. This analysis is presented in Section 4.3.

4.2.4

Interaction Potentials (Force Fields) for Computer Simulations of Liquid Systems

The description of the properties of liquids via computer simulations is among the fields in molecular sciences requiring the largest use of computer power. The reason is well known and is due to the combination of two factors: (i) the versatility of the procedures able to give in many cases information hardly obtainable by other means and (ii) the complexity and high computational demand of even a simple application. The computational demand is related to the necessity of computing many times (millions of “moves” or of time steps) the interaction energy in a system composed of hundreds or thousands separate molecules. Much effort has been spent to gain accessible computational times. To reach this goal every aspect of these complex procedures has been scrutinized with care.

An important goal is an efficient method to compute these interactions. The solution was sought by replacing QM calculations with classical estimates of the

interaction energies. However, even now, after more than forty years of effort, work in this field has not terminated yet.

The first sets of interaction potentials were completely empirical, based on experimental parameters. This approach was abandoned owing to a lack of reliable experimental data.

The use of quantum calculations to determine the interaction potential was pioneered by Clementi in the late-1970s in a series of landmark studies and was progressively adopted by others. At present, all new interaction potentials for liquids derive from *ab initio* calculations. Progression in these studies has at present reached the stage of searching force fields of chemical accuracy, with the use of large basis sets and QM methods of high level (at this level of elaboration of methods the computational costs are not a problem: the problem is to have low-cost interaction potentials).

It is convenient to introduce a broad classification of the existing models taken from a study by Burnham and Xantheas [21]. The potential may be of two-body type (pairwise additive) or of many-body type (generally limited to include mutual polarization effects, the many-body dispersion effects are in general neglected). In parallel, the model can be rigid or flexible, a classification based on allowing single solvent molecules the possibility of internal geometric deformation as a result of the interactions felt. Thus, four classes of interaction potentials can be defined:

- 1) rigid/pairwise additive
- 2) rigid/polarizable
- 3) flexible/pairwise additive
- 4) flexible/polarizable.

The computational cost of using such potentials in a simulation notably increases from class 1 to 4; good computational potentials of class 4 have at present a computational cost more than 100 times larger than a good model of class 1.

The interactions we have examined in the preceding sections can be directly used for models of class 1 alone, while for the other classes QM calculations on larger molecular aggregates are necessary: clusters with geometry optimization as well as derived from simulations are used, notably using first principles molecular dynamics derived from the seminal work of Car and Parrinello [22]. In all cases the computational data are used to define site potentials of appropriate and simple analytical forms [23–26].

This classification can be applied to water, for which there are models of the four classes, but it is potentially only defined for other solvents, their elaboration being at a more primitive stage.

In quantum biochemistry, water dominates as solvent and so we may dispense with further consideration of other solvents, but this limitation does not eliminate all the open issues in the definition of force fields for biochemical simulations. The other bodies to be considered in simulations of biochemical material models are of different nature (molecules of small and large size, dissociated salts, organized biological systems, etc.), which are to be treated differently.

Material models with a single molecule can be treated with QM/MM procedures, that is, with a QM description of the solute molecule and a molecular mechanics (MM) description of the solvent using the force fields we have considered above; the interaction potential between the solute and the solvent molecules must be defined in a congruent way with respect to the adopted force field.

When the solute is too large to be described with a QM procedure, a MM/MM description is used, where the solute is also described with appropriate force fields. The line of separation between “small” and “large” solutes is continuously shifted by improvements in the computational setup.

In other material models the biological component of the system exhibits a large degree of aggregation (consider for example biological membranes) and full QM descriptions of the biological component are not possible yet.

4.3

Solvation: the “Continuum Model”

4.3.1

Basic Formulation of PCM

This section presents an aspect of computational chemistry that is important in biochemical problems, namely the modeling of solvent effects. Reference is made to one of the possible computational strategies proposed, that is, the method we developed in Pisa and currently called PCM (polarizable continuum model) [27, 28].

PCM was born as an *ab initio* method, and we shall explore first its basic *ab initio* version.

The Hamiltonian of the system (a solution at infinite dilution) can be written without approximations in the following form:

$$\hat{H}_{MS}(m, s) = \hat{H}_M(m) + \hat{H}_S(s) + \hat{V}_{int}(m, s) \quad (4.9)$$

where m and s are the degrees of freedom of the solute (M) and of the solvent (S), respectively. The first two Hamiltonians on the right-hand side represent the solute and the solvent separately and the third operator represents the interactions among the components of the two subsystems.

The approximation we introduce consists in the definition of the following effective Hamiltonian:

$$\hat{H}_{MS}^{eff} = \hat{H}_M(m) + \hat{V}_{int}^{eff}[m, Q_{int}(\mathbf{r}, \mathbf{r}')] \quad (4.10)$$

The Hamiltonian of the solute is left unchanged, but that of the solvent is eliminated and replaced with a two-body interaction operator that depends on the solvent response function Q_{int} ; in its expression $\mathbf{r}$ and $\mathbf{r}'$ are position vectors. This enormous simplification of the problem is one of the most important characteristics of continuum methods.

The effective interaction potential of Equation 4.10 is composed by several terms, which in PCM are defined as:

$$\hat{V}_{\text{int}}^{\text{eff}}[Q_{\text{int}}(\mathbf{r}, \mathbf{r}')] = \hat{V}_{\text{cav}}^{\text{eff}}(Q_{\text{cav}}(\mathbf{r}, \mathbf{r}')) + \hat{V}_{\text{el}}^{\text{eff}}[Q_{\text{el}}(\mathbf{r}, \mathbf{r}')] + \hat{V}_{\text{rep}}^{\text{eff}}[Q_{\text{rep}}(\mathbf{r}, \mathbf{r}')] + \hat{V}_{\text{dis}}^{\text{eff}}[Q_{\text{dis}}(\mathbf{r}, \mathbf{r}')] \quad (4.11)$$

where we have eliminated the argument m , as it is redundant.

The choice of components is based on analyses of the interaction potential in dimers. Note that, in contrast to the decomposition of the dimeric interaction, Equation 4.11 is not a decomposition of an already known quantity but an expression giving the contributions that compose the interaction potential. The terms have the following meaning:

- 1) V_{cav} regards the formation within the solvent of a cavity where the solute will be accommodated. The reason for the presence of this term (not present in the dimeric decomposition) is explained below.
- 2) V_{el} collects the electrostatic contributions to the solute–solvent interaction energy (analogous to those given by the $E_{\text{coul}} + E_{\text{pol}}$ terms in the dimer).
- 3) V_{rep} regards repulsive interaction between solute and solvent. In the dimer it was given by the exchange term. In solution, part of the repulsion is given by the cavity formation term while the remaining part comes from this term.
- 4) V_{dis} regards the dispersion interactions, a term we replaced in the dimer decomposition with an empirical function.

Once an effective Hamiltonians is introduced to describe the solvated system we need to define the thermodynamic status of the effective operator describing the interaction with the medium. In our case the most convenient choice is given by the Gibbs free energy (G). This was already done for classical electrostatic models. Within this framework, the free energy becomes the main output of the QM calculations, in contrast with standard calculations for isolated molecules which use internal energy. In using solvation models (continuum as well as discrete) attention must be constantly paid to the proper definition of the thermodynamic status of all the components of the systems. For example, the molecular partition function of statistical mechanics has to be redefined using free energy as starting element (zero-point vibrational corrections to the energy are to be measured as free energies). In addition, the functional used in the variational theorem, which we have mentioned above, and which we shall also use here, has to be modified [29].

The next step is to define the reference energy of the model system we are trying to describe. As stated above, the model system is composed of two parts: the solvent (S) and the solute (M). For the solvent the best choice is a pure solvent in equilibrium at a given T and P , with a free energy (G) set equal to zero. For the solute we have different choices: we are here describing a full *ab initio* procedure, and so the best choice for M is given by the appropriate number of non-interacting electrons and nuclei necessary to describe the solute, all elements at zero kinetic energy. This is the starting state, at zero energy, in the *ab initio* QM studies of molecules. In other words we are

describing the process of formation of a molecule inside the solvent, taking into consideration all the energetic quantities involved in this process.

The expression of the free energy at equilibrium is:

$$G = G_{\text{cav}} + G_{\text{el}} + G_{\text{rep}} + G_{\text{dis}} + G_{\text{tm}} \quad (4.12)$$

where the last term includes contributions due to thermal motions of the solute, as well as zero-point vibrational energies, vibrational and rotational contributions, which, as in standard QM calculations *in vacuo*, are separately computed and added to the result obtained as solution of the Schrödinger equation based on Hamiltonian (4.10).

Let us examine now each individual element of Equation 4.12:

- 1) G_{cav} : our definition of the reference energy explains why there is the need of an energy associated with the formation of the cavity. In the unperturbed solvent, at equilibrium, there are no empty spaces sufficient to accommodate M and so we need to spend energy (under the form of a reversible work) to form a cavity of suitable size and shape. The parameters for the definition of this free energy (called G_{cav}) must be derived from properties of the solvent accompanied by information about the size and shape of the cavity. We examined different alternative definitions and our final choice was the use of the scaled particle theory (SPT) [30], which can be used for solvents of any nature. Actually, the validity of this SPT model was not initially granted, but the cavity formation energy is the only part of the solvation energy that can be determined independently, via computer simulations with discrete solvent models, and now it may be said that evaluation of the problem was correct and the choice we made was appropriate, at least for small-medium molecules.

The other components of the solvation interaction cannot be singled out and checked with independent experimental data. The only direct checks are on the total solvation energy. Indirect checks on the properties of solvated molecules (spectroscopic properties, reaction mechanisms, etc.) are more abundant as well as comparisons with computational results obtained with other methods.

All the free energy contributions deriving from the potential of Equation 4.11 can be formally defined in statistical mechanics by assigning a varying parameter λ , specific for each term, with values ranging from zero (no interaction) to 1 (full interaction) and performing a charging process. The charging process is the integral along all the values of λ ; at the end of this process the distribution of matter is changed and this new distribution has to be used for the next charging process. This formal definition can also be applied in actual calculations. Indeed, it is applied in PCM to compute some contributions (namely G_{dis} and G_{rep}) in a semiclassical formulation; alternatively, there is now a PCM formulation that inserts all perturbation terms (cavity formation apart, which is still determined with a charging process) in the calculation of the electron density of M. This latter formulation avoids the problems related to the couplings of the charging processes.

- 2) G_{el} : The electrostatic term is the most important one in almost all cases – the exception being non-polar solutes in non-polar solvents: systems that are not generally regarded in quantum biochemistry. Physically this term describes the effects due to the so-called solvent reaction field, a concept introduced by Onsager [31], that is a cornerstone in theories of solutions. The concept is simple: the molecular charge distribution polarizes the liquid (described as a dielectric), the polarization of which in turn acts on the solute, inducing a polarization. This is an implicitly iterative definition; the polarized molecule modifies the solvent reaction field, and so on.

The electrostatic operator can be derived starting from the Poisson equation for the classical analogous system (a distribution of charges within a limited space delimited by a closed surface, i.e., by the cavity surface, surrounded by a dielectric in which there are no charges) and applied to our system in the QM formalism. Among the various possible procedures of solution of the resulting integral equation (almost all are at present used in the various continuum solvation methods) we were the first to select a method based on the apparent charges (i.e., polarization charges) appearing on the cavity surface as consequence of the jump condition on the gradient of the total electrostatic potential on the surface separating the molecule from the dielectric medium. We called this approach the ASC (apparent surface charge) method but it is now better known as BEM (boundary element method), a name introduced in the scientific literature a few years later. This method presents remarkable advantages with respect to alternative procedures. The solute–solvent interaction is reduced to that of the solute charge density with solvent point charges solely placed on the cavity surface. These charges are easily computed because they only depend on the electric field generated by all the charges, real or apparent, present in the system. The limitation of the allocation of charges on a finite 2D surface allows a detailed numerical description of the potential acting on M, even for molecules of large size and with a corrugated surface, which can be refined by increasing the number of point charges without appreciable numerical trouble.

In the PCM scheme, the cavity is given in terms of interlocking spheres centered on the nuclei of the solute; its surface is divided into small elements (called “tesserae”) and the apparent surface charge is computed at the center of each tessera; in the ensuing calculation the charge is weighted by the surface of the tessera. This is the simplest way of making an allocation; in fact, the function to be allocated could be described within each tessera with a polynomial expansion. The use of a single charge corresponds to a polynomial of order zero (hence called P_0); we have not considered convenient to use polynomial of higher degree (say P_1 or P_2) with the tessellations currently used for molecules. The option of higher P_s perhaps will be useful for larger and more ordered surfaces (a possible example is the β -sheets in proteins).

More details on the mathematics of the integral equation methods exploited in the PCM can be found in the clear review given by Cancès [32] in a chapter of a book we edited recently. In another chapter of the same book Pomelli [33] gives useful additional information about the cavity surface discretization and its use.

- 3) G_{rep} : the repulsion term gives a positive (i.e., repulsive) contribution to the energy. The physical origins of this term are the antisymmetry requirements interacting electrons must respect. We recall that a large portion of the repulsion between solute and solvent has been already described by the cavity formation energy term G_{cav} . Actually G_{rep} regards a residual contribution well described by a response operator with kernel based on the overlap between the solute electron distribution and an electron distribution for the solvent drawn from the numeral density of atoms in the liquid combined with their averaged charge distribution [34]. The resulting integral equation is described with a BEM procedure as for the electrostatic term, with contributions expressed on the same tesserae. The repulsion contribution to the solvation energy can be also computed with a semiclassical formula of similar nature and treated in the same way [35]. In this case the coupling of this term with the others is neglected, and repulsion cannot affect the solute charge distribution. Under real conditions (as those corresponding to systems in living organisms) the effect repulsion has on the charge distribution is negligible; this is no longer true for systems at high pressure, where repulsion becomes the dominant term to be carefully treated with a QM description [36].
- 4) G_{dis} : an exact closed-form for the expression of the dispersion energy in dimeric interactions is not available, but several approximate expressions exist. The expression used in PCM [34] derives from the formulation given by McWeeny [37]. We do not summarize this elaboration, which is quite complex. The final expression can be treated with the BEM formalism, as done for the preceding contributions. There is also a semiclassical formula for dispersion that is reduced again to contributions to the single tessera, without coupling with the electrostatic terms and without effect on the solute charge distribution [38]. A systematic study on a sizeable number of systems [39] has shown that the semiclassical approach produces effects on the solvation energy quite similar to the QM one, at least for molecules in their ground electronic state.

More details on what is reported in this section can be found in our review [28].

4.3.2

Beyond the Basic Formulation

The versatility of PCM is remarkable. We quote here some example applications in the biochemical field.

4.3.2.1 Dielectric Function

The available versions of PCM in the Gaussian package [40] have several sets of internal solvent input data, including the dielectric constant. Other inputs provided by the user are also accepted, so that non-standard environments can also be described, as shown in a recent paper with input for the blood serum [41]:

- The dielectric response function appearing in the $Q_{\text{el}}(\mathbf{r}, \mathbf{r}')$ operator of Equation 4.11 is not limited to being a constant. For example, its partition into two

components, corresponding to two different response regimes (slow and fast) is used to describe solvation effects connected to sudden changes in the charge distribution of the solute, as in spectroscopic properties. See Sections 4.4.5 and 4.4.7 for more details.

- Its description as a function depending on the ionic strength can be used to include the effects of dissolved salts in the medium. This has been carried out within the IEF-PCM scheme, to describe low ionic strengths (i.e., in the limit of validity of the linearized Poisson–Boltzmann) [42].
- Its description as a tensor function can be used to describe solvation effects in ordered liquids. This has been performed within the IEF-PCM scheme to describe nematic liquid crystals [43].
- Its description as a position dependent function finds applications in several cases. A sigmoidal dielectric function $\epsilon(r)$ is used to describe cases in which it is convenient to introduce a local disturbance to the uniform continuum (supercritical solutions, charged solutes, but also sometimes recommended for polar biomaterials). A function $\epsilon(z)$ is used to describe solvent effects near a flat phase separation (liquid/air, liquid/liquid, liquid/metal) [44].

4.3.2.2 Cavity Surface

It is worth mentioning the use of PCM with two cavities (or more). This feature is useful to describe dissociation processes in which a single cavity splits into two when the separation between the fragment increases, and aggregation processes when two or more solutes merge into a single aggregate.

4.3.2.3 Definition of the Apparent Charges

There are different versions for the description of the apparent charges. We introduced the original version [45] based on the values of the potential gradient at the surface: this version is now called D-PCM. Alternatively, for polar solvents, a version of the model based on the potential at the surface has been proposed. This version, known as C-PCM [46], has been obtained by reformulating the conductor-like screening model (COSMO) (see below) within the PCM framework. Finally, a third and more general version of PCM has been proposed that uses a completely new integral equation formalism. This version, called IEF-PCM [47], can treat equally well polar and non-polar solvents, isotropic and non-isotropic dielectric descriptions, salt solutions and other more complex environments.

4.3.2.4 Description of the Solute

Almost all levels of the *ab initio* QM theory have now been combined with PCM. The largest part of these levels also contains expressions for analytical derivatives, an essential tool for geometry optimization and calculation of spectroscopic properties. In addition, semiempirical QM methods have been implemented with PCM (we profitably use a PCM-ZINDO version for the excited states [48]). Solvent descriptions in terms of molecular fragments in the version elaborated by Gordon and coworkers [49] have been also proposed [50–52].

Another application deserving mention in the field of quantum biochemistry is that with Morokuma's integrated method (called ONIOM) [53, 54]; this is in fact a quite useful approach to describe large molecules (see Chapter 3, and also Chapter 2).

4.3.3

Other Continuum Solvation Methods

We limit ourselves here to indicating the most used models.

4.3.3.1 Apparent Surface Charge (ASC) Methods

Among ASC methods, the PCM family of methods (D-PCM, C-PCM, IEF-PCM) represents one of the most popular choices; there are, however, other important (and well diffused) approaches. Among them we quote here three alternatives:

- 1) **COSMO**: In this method, originally devised by Klamt and Schüürman [55], who also elaborated further versions [56, 57], the dielectric constant of the medium is changed from the specific finite value, characteristic of each solvent, to infinity. This value corresponds to that of a conductor, and this change strongly modifies the boundary conditions of the electrostatic problem. The most important effect is that the total potential $V(r)$ cancels out on the cavity surface. From this condition it follows that the ASC is determined by the local value of the electrostatic potential instead of the normal component of its gradient. To recover the effects of the finite value of the dielectric constant of the medium, the ideal unscreened charge density corresponding to $\epsilon = \infty$ is finally scaled by a proper function of ϵ . This rescaling (which is of minor relevance in a polar solvent at high dielectric constant) has given rise to several variants, among which we cite GCOSMO of Truong and Stefanovich [58] and C-PCM of Barone and Cossi [46].
- 2) **MST**: The so-called Miertuš–Scrocco–Tomasi code is derived from PCM and has been developed by Orozco, Luque and coworkers [59]. The main difference with respect to standard PCM is in the calculation of dispersion and repulsion contributions parameterized with a simple analytical function based on the surface tension. The more recent MST parameterizations regard water, chloroform, octanol and carbon tetrachloride.
- 3) **SVPE**: is the initialism (surface and volume polarization for electrostatic) of one of several versions elaborated by Chipman in a careful analysis of the effects of the electronic charge lying out of the cavity [60]. The exact SVPE method is laborious to implement and time-consuming, because it utilizes a volume polarization potential arising from a discontinuous volume charge density. To avoid this large complexity a simpler approximate solution that involves only apparent surface charge distributions was subsequently introduced, denoted as "surface and simulation of volume polarization for electrostatics," SS(V)PE. It has been shown that such an approximation is equivalent to IEF-PCM (see Reference [28] for an analysis).

4.3.3.2 Multipole Expansion Methods (MPE)

The multipole expansion is the standard formulation in classical electrostatics. It is different for the portion of space containing the confined charge distribution and the exterior. The limitations of the single center multipole expansion (notably convergence limited to the space out of the largest sphere containing all the charges) makes it necessary to use multicenter multipole expansion. In some cases the abbreviation SCRF (self-consistent reaction field) is also used to indicate this specific approach.

The most complete, and also the most largely used, MPE-SCRF method is that developed in Nancy by Rivail and coworkers [61]. This last version of the method contains several important new features (already present in PCM) as the analytic form of nuclear geometry derivatives – essential for an extensive use of the code. An earlier version of the Nancy code was the first continuum solvation program using a QM description of the solute (a single multipolar expansion into a sphere and later into an ellipsoid of a CNDO wavefunction) [62].

An alternative version of the MPE method has been proposed by Mikkelsen and collaborators [63]. Such a version of the model has been implemented in the Dalton quantum chemistry program, and it has been used to study solvent effects on a large number of molecular response properties; it has also been generalized to several QM approaches including multiconfigurational self-consistent field and coupled cluster methods.

4.3.3.3 Generalized Born Model

Several continuum solvation procedures can be collected under the heading generalized Born. The distinguishing feature of the approach is the use of an empirical (but effective) formula to express the Coulomb interaction among Born models. The Born model is a very simple expression giving the free energy of solvation of a single point solute (i.e., an atomic charge) within a sphere of given radius immersed into a continuum dielectric.

Surely the most known and used version of the generalized Born methods is the family of approaches elaborated by Cramer and Truhlar and indicated with the abbreviation SMx. The latest version is called SM8 [64].

The strategy used by Cramer and Truhlar since the first version of the method [65] consists of a systematic parameterization of several contributions describing as a whole the free energy of solvation. Within the SMx framework, the expression of the solvation free energy becomes:

$$\Delta G_s^\circ = \Delta G_{\text{ENP}} + G_{\text{CDS}} + \Delta G_{\text{conc}}^\circ \quad (4.13)$$

where the first term is the change in the component of the solvation energy associated with electrostatic contribution (subscript ENP denotes the electronic, nuclear and polarization components of the solvation energy), the second is a collective term, interpreted as due to cavitation, dispersion and contributions due to solvent structuring around the solute, whereas the third term accounts for the difference in molar densities of the two phases in their standard states.

This strategy of combining non-electrostatic terms coherently parameterized with *ab initio* calculations for the ENP terms makes SMx programs unique in the literature

and a quite accurate approach to computing solvation free-energies in aqueous and organic solvents [66].

There are many other generalized Born codes that generally tend to simpler computational schemes, based on a more restricted parameterization, and in most cases avoiding the use of a QM description.

4.3.3.4 Finite Element Method (FEM) and Finite Difference Method (FDM)

These methods, as their names imply, use a discretized numerical approach to solve the electrostatic problem represented by the Poisson–Boltzmann (PB) equation. The best known examples of finite element method (FEM) codes have been elaborated by Friesner’s group with the abbreviation PBF [67] and by McCammon’s group with the name of adaptive Poisson–Boltzmann solver (APBS) [68]. The two methods have important difference in their implementation, but in both cases the discretization is extended to the whole space and not limited to the surface only as in the BEM methods discussed above.

Finite difference methods (FDMs) are based on a different mathematical approach. The basic PB equation is expressed in a differential form, with a discrete approximation using a grid of points distributed on the whole volume. The most used codes are DelPhi [69], Mead [70], UHBD [71] and that implemented by Friesner and coworkers [72] in Jaguar [73].

4.4

Applications of the PCM Method

4.4.1

Solvation Energies

A considerable number of users of computational tools for applications to chemistry and biology are simply searching for reliable values of the solvation energy ΔG_{sol} and nothing more. There is available a large variety of methods and codes, and it is convenient to make here a few remarks as a guide for the selection of the method best corresponding to the needs of the user.

To summarize this problem to audiences of non-specialists, we found it convenient to condense the large wealth of proposals into four classes, with several internal partitions, or levels. A short synopsis of this classification is presented here.

The simplest approaches (class I) regard methods in which the calculation of solute–solvent interaction is replaced by a single molecular parameter, in general the surface of the solute accessible to the solvent (SAS). These methods require a reliable geometry of the solute (it should be that present in the given solvent), accompanied by a code computing the SAS.

The published codes are calibrated on a training set of experimental values. In general the training set regards a single solute, in some cases a couple. Such codes are often claimed to be valid for all solutes, even when the training sets are limited to a few types of solutes.

The second class in order of increasing computational demand (class II) collects the methods using a continuum description of the solvent, but limiting the solute–solvent interaction to an electrostatic term. Within this class three levels can be singled out: methods based on a semiclassical description of the solute, methods based on a semiempirical QM approach and methods based on an *ab initio* QM method.

For polar solutes in polar solvents the electrostatic component of the interaction is the dominant one; the others compensate each other to a good extent.

From what we reported in Section 4.2, it may be argued that a semiclassical approach accompanied by a reliable library of fragments could be a reasonable option. A good calibration is, however, compulsory, and the remarks expressed for class I methods are still valid.

The semiempirical level avoids the use of libraries, and it is of more immediate use, but a little more computer demanding. Calibrations are also necessary at this level.

The full *ab initio* level demands even more computer resources and calibration is still necessary. Methods belonging to this class often neglect solute polarization effects induced by the solvent. This additional simplification is almost compulsory for the semiclassical methods but it just an option for the other two levels, introduced to reduce computational times.

PCM codes can (and are) be used for methods of this class, semiclassical included.

Class III covers full continuum solvent approaches. In this chapter we only consider the *ab initio* versions of PCM, corresponding to the highest level of this class. Almost nothing has been said about calculations performed with semiempirical QM methods and with fragments. Both approaches can be used to compute solvation energies with appreciable results (see, for example, Reference [52]).

Readers interested in the calculation of solvation energies are directed to a recent sequence of three papers in *Accounts of Chemical Research*, where the performances of some continuum methods are compared [66, 74, 75]. The first comparison, given by Cramer and Truhlar [66], gave rise to a comment [74], followed by a final remark by the authors of the first account [75]. The conclusion of this series of analyses, which also gives other interesting details, is that several methods, when calibrated, give values of ΔG_{sol} at the level of chemical accuracy. Of course, the introduction of a calibration for the solvation energy does not imply that the descriptions of other properties are improved in parallel.

Class IV collects simulations with discrete solvent molecules. The first three levels correspond to MM/MM, semiempirical QM/MM, and *ab initio* QM/MM simulations, amply used, and all give ΔG_{sol} as a part of a larger output.

The higher level regards simulations in which all the elements, solute and solvent molecules are described at the *ab initio* level. These last calculations are the nearest to the ideal of full accuracy in the description (the limitations are due to the level of the QM calculations and mainly to the number of solvent molecules used in the simulation) but are considerably expensive.

It is not easy to compare the computational cost of each method because published papers rarely report the pertinent information. Information given by colleagues currently using simulations, and an appraisal of the number of elementary math-

ematical operations necessary to reach the required result, lead to a conservative estimate of the ratio of a class I calculation to a calculation at the highest level of one to several million.

4.4.2

About the PES

One of the most important contributions given by quantum mechanics to chemical enquiry is the concept of potential energy surface (PES). It is well known that this is a surface, or better a set of surfaces, one for each electronic state, defined over the $3N$ -dimensional space spanned by the nuclear coordinates of the material system. Each point of this surface corresponds to a specific collocation of nuclei, and the local value of the PES defines a specific energy, the electronic energy of the systems, which in QM defines the potential for the full QM description of the system (including the nuclear part). The usual chemical use of PES neglects this second part of the QM problem, preferring to use a semiclassical approach following a picture based on the use of the gradient of this energy and all the related conceptual and mathematical tools to track the most important position on this surface, and their relative energies.

In solution, things are by far more complex. The number N of nuclear coordinates is enormously larger and the use of a complete PES is impossible. Fortunately, in continuum solvent approaches the effective Hamiltonian we have introduced (see Equation 4.10) is given in terms of the N coordinates of the solute alone. The semiclassical approach to exploiting PES is exactly the same as for a molecular system *in vacuo*. The only difference is that the energetic quantity is no longer the internal energy (E), but rather the free energy (G). This function should be called the “free energy potential surface” (FEPS) but almost nobody uses this term. The function $G(R)$ (were R stands for the nuclear coordinates) depends on some parameters set in the QM calculations of PES, namely the temperature and the related value of the dielectric constant. If one changes the temperature in a study, the PES will also change. A PES in solutions contains a fair amount of entropy, which is not present in the energy function $E(R)$ defining the PES *in vacuo*. This means that the surface $G(R)$ is flatter than $E(R)$, and the higher the temperature the flatter the surface.

The solvent coordinates are not completely lost, however. In some more refined studies, addressing specific problems in a semi-dynamic approach, solvent collective coordinates may be profitably added. A detailed presentation of the topic concerning the chemical use of $G(R)$ can be found in Reference [76].

Our short considerations will be limited to two aspects of the use of PES, chemical equilibria and reaction mechanisms.

4.4.3

Chemical Equilibria

For a reaction the problem is formally simple. What is needed is the free energy of the reactant A and of the product B, both computed at the respective equilibrium geometries. Note that both $G_A(r_{eq})$ and $G_B(r_{eq})$ must contain all the contributions,

including those not directly given by the QM calculation of the energy, namely the G_{tm} term of Equation 4.12. Because the PES is relatively shallow, particular attention must be exercised for the proper description of vibrations, especially with regards to those connected to low frequency motions.

4.4.3.1 Tautomeric Equilibria

Tautomeric equilibria are of considerable importance in biochemical problems. They have also been extensively studied in relatively distant years. Cramer and Truhlar have given detailed reviews about these calculations, in particular in their review published in 1999 [77].

4.4.3.2 Equilibria in Molecular Aggregation

Another class of equilibria includes molecular aggregations. These phenomena are even more frequent and important in biochemistry than tautomerism. They also are more subtle and more difficult to describe properly. Since we have dedicated a relatively large space to noncovalent interactions isolated in space there is no need to repeat these concepts and analyses here. We limit ourselves instead to consider the supermolecular descriptions, without alluding to the use of many-body interaction potentials, which have problems of their own.

The region of the PES describing clustering is extremely flat and, in general, different clusters are very near in energy, which explains why accurate studies exploring their PES are not frequent. A specific problem is the description of the contributions to the entropy of the modes of motion passing from translation and rotation (before the aggregation) to internal vibrational modes (in the aggregate). These very low frequency vibrations, which give important contributions to the entropy, cannot be treated with the harmonic approximation, and even anharmonic corrections are often not sufficient. We have studied these phenomena in several cases, as a corollary to the main subject of the inquiry [78, 79], finding difficulties in reaching a sufficiently accurate description. There is surely a need for a computational approach able to give multimode interactions, but, to the best of our knowledge, a version for continuum solvation methods is not yet available.

4.4.3.3 pK_a of Acids

To close this section, we discuss the calculation of pK_a of acids.

The calculation of this quantity, of fundamental importance in biochemistry, has given rise to a deluge of papers. In this literature the subject is split into different parts, the calculation of relative values, the calculation of absolute values and the calculation of pK_a for residues buried in a protein or other biological material.

While pK_a calculations belong to the family of reaction equilibria discussed at the beginning of this section, there are good reasons to treat them separately. The main reason is that one of the products of acid dissociation, the proton, does not exist as a separate entity in solution. Limiting ourselves to aqueous solutions, the proton will appear as hydroxonium ion, H_3O^+ , which in turn aggregates more water molecules, to first form well-recognized aggregates such as the Zundel ($H_5O_2^+$) and Eigen ($H_9O_4^+$) complexes, and then form other larger aggregates. As mentioned above, a

rigorous study of the energetics of these equilibria is a delicate and hard task, subject to possible inaccuracies. The shift from continuum to discrete models is not of much help: the interaction potentials of water, considered in Section 4.2.4 should be extended with inclusion of the most important protonated species, and such enlarged potentials are not currently available. Resolution of this computational problem is of course possible, and there are several studies on this subject. A different way that has been selected by almost all the calculations of pK_a is to adopt thermodynamic cycles including the parallel dissociation process in gas phase and introducing here some experimental values. The simplest approach is to take the solvent protonation free energy from experiments, leaving to experimentalists the responsibility of fixing this value. Unfortunately, the definition of the absolute solvation energy of the solvent is not an easy task. We have remarked with satisfaction that in recent years the recommended value for this quantity has been changed. We expressed, several times, our opinion that the generally adopted choice was not satisfactory and that another suggested value was to be preferred. The choice of the value proposed by Tissandier *et al.* [80] satisfies our remarks. As just mentioned, several thermodynamic cycles were suggested and used. Among them we cite that proposed by Clarissa da Silva and used in her studies on acidity. The experimental value she selected was the vaporization energy of a water molecule [81]. This choice was adopted by others, and also criticized, notably by Pliego [82], opening up a debate in which others intervened. It is beyond the scope of this section to review and analyze this debate and its several elements: quality of the QM calculations, cavity radii, use of thermodynamic definitions and standard, and so on. It is sufficient to emphasize that this debate has not yet ended.

To conclude our discussion we note that absolute pK_a calculations in bulk water have almost reached chemical accuracy [i.e., a ΔG corresponding to a fraction of the pK_a unit, which is equal to $1.36 \text{ kcal mol}^{-1}$ ($1 \text{ kcal} = 4.184 \text{ kJ}$)]. However, it is not clear yet if this agreement is due to error cancellation more than to a formally exact elaboration of the procedure.

Calculation of the pK_a of acidic groups buried in a biological matrix is far more complex. The continuum approach clearly cannot describe highly structured and quite specific local media, but it may be profitably employed as complement in methods giving a sufficiently detailed atomistic description of the biological structure [52].

4.4.4

Reaction Mechanisms

Reaction mechanism studies make greater use of the PES features than the chemical equilibria we have thus far considered. The standard techniques for the study of reaction mechanisms in the gas phase are well known. First, the equilibrium geometries of reactant and products have to be determined (they are critical points on the surface corresponding to local minima). Second, the transition states have to be determined (they are critical points of another type, namely saddle point of the first type, usually called SP1 critical points). Simple reactions have just one SP1 point, but

many reactions have several saddle points; in this search the concept of reaction coordinate may be usefully employed. Third, the study can be refined with the determination of geometry and energy of the reaction intermediates, always present in the case of multiple saddle points. In continuum solvation models the procedure is analogous and so no further comments are needed on these aspects of such studies.

Comments are necessary, though, for special cases of frequent occurrence in biological problems. The medium encircling the reacting system is not an inert spectator but some molecules quite often have an active role. The distinction between different degrees of involvement of specific solvent molecules is an important step in a full description of the mechanism. In a recent paper [83] we tried to give a classification, which surely can be refined, but is sufficient to give a first insight into the problem. There are water molecules that exhibit a persistent interaction with the solute, but without an active role in the reaction mechanism, and others with persistent interactions and an active role. The molecules with an active role must be inserted in the QM description of the effective Hamiltonian, in other words, the QM portion of the system must be enlarged to include these molecules. To finish the classification, there are some other molecules that may play a specific role in the reaction without necessarily entering the QM description of the transition state. For such molecules we have coined the name “actively assisting molecules.” How many active molecules are present in the system and where they are placed with respect to the molecular framework is not dictated by the stoichiometry of the reaction. Chemical intuition may assist, but we feel that only during the calculations is it possible to precisely define the number and position of active molecules. A strategy to do this has been sketched in the quoted reference. We have found it profitable to use the decomposition of the forces acting on specific nuclei of the QM portion, a procedure derived from the semiclassical analyses mentioned above, which are easily implemented in continuum solvation codes. The information gained by adding solvent effects to the analysis is remarkable; local solvation forces push the reacting chemical group towards completion of the reaction, while for other groups a counteracting effect can occur, with a distortion of the geometry of the transition state with respect to that found without solvent. The assisting molecules reinforce these effects. An example is given in Reference [84].

As a final consideration we return to the first study of a complex reaction mechanism we published over 25 years ago [85] regarding the mechanism of carbonyl reduction by LiBH_4 in ether. This was the first published *ab initio* continuum solvent study for a reaction mechanism with a complex structure (leaving aside the Menshutkin and Walden reactions previously studied, which have a simple collision-like character). It is a study with many technical limitations (duly acknowledged in the paper) but which required considerable effort. Indeed, there were no analytical derivatives, an essential tool in the study of reaction mechanisms. Despite these efforts we spent considerable energy and computational time to establish the elements of the analysis subject of this review. An analogous study on this system could now be performed in a few days without difficulty, probably uncovering other effects not yet analyzed due to the presence of a reagent behaving as an ionic pair, but we have not found studies of this kind in the literature. The evolution of compu-

tational chemistry has led towards computer simulations of chemically interesting events, paying little attention to the interpretation. This is a pity, in our opinion.

4.4.5

Solvent Effects on Molecular Properties and Spectroscopy

Recent progress in the field of derivative techniques has made it possible to calculate various molecular properties and spectroscopies, even with the inclusion of solvation effects [28, 86]. This field has developed tremendously through the implementation of solvation techniques in computational packages available to the wider scientific community.

When aiming to calculate molecular response properties or spectroscopies for systems surrounded by a given environment, a certain number of effects are to be considered, especially if a quantitative comparison with experimental findings is to be achieved. In this case, a description as accurate as possible of the physics of the “solvated” sample interacting with the external perturbing field should be considered.

When the system is not isolated, additional factors due to the interaction of the molecule with the surrounding have to be taken into account in the development of successful computational strategies. In principle, a successful comparison between calculated and experimental properties for systems in solution requires the inclusion, in the calculated data, of the maximum number of possible effects that are believed to be present in the experimental sample.

The approach generally adopted in the modeling of solvated systems consists of applying the same methodologies developed for isolated systems with the additional introduction of solvent-dependent peculiarities.

By taking as reference a calculation *in vacuo*, the presence of the environment introduces some complication in the analysis. We certainly have some kind of “direct” effects, due to the changes in molecular electronic distribution induced by interaction with the environment. In addition, in the presence of an external perturbing field, the so-called “local-field” effects should be taken into account, these being due to changes in the external applied field induced by the presence of the environment. Furthermore, since interaction with an external field is, in general, a time-dependent phenomenon, solvent relaxation is also to be considered in addition to “specific” solvent effects depending on the nature of the solute and solvent pair.

Lastly, the presence of a solvating environment always causes changes in the geometry of the solute with respect to its geometry *in vacuo*. In addition, floppy systems generally exhibit different conformational preferences in solution as opposed to *in vacuo*. In recent years, both our group and others have developed computational strategies for evaluating all the aforementioned effects for several spectroscopic and response properties. In this section we briefly analyze this matter by resorting to few cases, which should be considered as representative of large families of systems.

Direct solvent effects on the molecular electronic distribution are always present whenever considering a solvated system. The accounting of such effects is the

primary scope in the development of any solvation model for molecular properties and spectroscopies, and it has been the subject of various contributions from both our group and others for many response properties and spectroscopies (see Reference [28] for details). In any case, accounting for such effects, if absolutely necessary, in some cases is not sufficient to gain a reliable description of the property in the condensed phase, especially if a direct comparison with experimental data is required.

Among solvent effects, a key role in the prediction and interpretation of molecular structures and properties is played by solvent-induced geometrical and conformational changes. Such effects are of general occurrence, because the presence of a solvating environment always causes a rearrangement in the molecular geometry with respect to the same system *in vacuo*. For this reason, complete geometry optimization in the presence of the real environment is generally to be recommended whenever attention is focused on a solvated system. These issues can be even more relevant in some selected cases, where such effects can be very large. To explain such a concept let us focus on a couple of case studies where the presence of the environment completely changes the picture with respect to vacuum. Such examples are of course very specific, but the findings of these studies should be kept in mind because they are expected to carry-over to a wider, more general range of systems.

4.4.5.1 *N*-Acetylproline Amide (NAP)

N-Acetylproline amide (NAP) is a proline derivative where the free amine is blocked with an acetyl group, and the carboxyl group is replaced by an amide (Figure 4.1). *trans*-NAP (the most stable isomer) can formally exist in various conformations, namely α helix, 3_{10} helix I, 3_{10} helix II, P_{II} and C_7 , differing from each other by the values of the conformational ϕ and ψ angles. The coupling of molecular dynamics (MD) simulations with infrared (IR), and Raman and vibrational circular dichroism (VCD) spectroscopies demonstrated that the three-dimensional conformation of NAP strongly depends on the solvent, particularly on the hydrogen-bonding ability of the solvent molecules, so that it was suggested the predominant conformation in non-polar solvent is the C_7 structure. The conformation of NAP in polar, protic solvents, such as water or alcohols, was instead still controversial [87], the results obtained by using MD simulations being strongly dependent on the force field

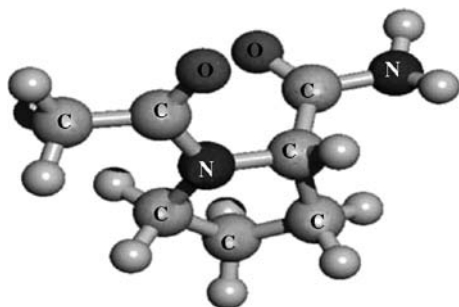


Figure 4.1 Ball stick model of *N*-acetylproline amide.

employed. Therefore, a key point in this case is certainly to obtain a reliable description of the conformational preference, or its dependence on the environment.

A recent paper by two of us has shown that continuum solvation models constitute a valid alternative to MD simulations for the investigation of these issues of NAP. In particular, the continuum strategy has been applied to evaluate the conformational distribution and its consequences on the prediction/description of various spectroscopic properties [IR, Raman, VCD, vibrational Raman optical activity (VROA), UV absorption and CD, ORD and NMR] of NAP in water solution, even with respect to experimental findings [88].

Geometry optimization showed that only the 3_{10} helix I and C_7 are stable minima in the gas phase, whereas in water three structures, that is, 3_{10} helix I, P_{II} and C_7 , coexist. At room temperature, in the gas phase, NAP assumes almost exclusively the C_7 conformation, with only 1% as 3_{10} helix I. In water solution the relative weights change, so that 68% is P_{II} , 28% is 3_{10} and C_7 is about 4%.

Accounting for the large changes in the relative conformer populations in the two environments (*in vacuo* and in water) led to completely different prediction of IR and VCD spectra; the latter showed a $-/+$ rotational strengths pattern in solution and the opposite $+/-$ alternation in gas phase and in the case of the sole consideration of solvent effects on the molecule wavefunction and PES minima. Because VCD is mostly used for assignment of the absolute conformation, such discrepancies are crucial.

4.4.5.2 Glucose

The importance of taking into account solvent-induced conformational changes has also been reported in Reference [89], where the optical rotation of the eight most abundant structures of glucose in aqueous solution is treated by resorting to time-dependent DFT (TDDFT)/GIAO coupled to PCM. The investigation had several intrinsic difficulties, related to the sensitiveness of the optical rotation (OR) theoretical values to the computational method used, the suitable description of the solvent effects and the high flexibility of the system studied, which makes glucose occur in many different possible conformations. In addition, optical rotation calculations are very sensitive to the geometry of the compound, so in this case it is a key point to recover a reliable conformational preference. Furthermore, the sampling of the large conformational space of carbohydrates complicates the prediction of OR, especially when the energy differences of the most stable structures are of the same order as the accuracy of the theoretical models employed. As reported in the literature, glucose occurs in aqueous solution with more than 99% as a six-member pyranose ring, in a stable ${}^4C_1$ chair conformation.

The results of the study showed that in aqueous solution four conformers represent more than 75% of the glucose molecule population, of which about 55% is due to the two β anomers. A similar result was also obtained in gas phase, but in this case the percentage is 38%. OR calculations on the selected conformations showed that all α structures gave a large positive contribution to the OR property, while the β structures gave both positive and negative contributions, and that had a considerable influence on the prediction of the property averaged over conforma-

tions. Also, the most abundant β anomers gave a negative contribution to the $[\alpha]_D$ net value. The calculated Boltzmann weighted $[\alpha]_D$ value of 58.75° compared quite well with the experimental value of 52.7° . The weighted value using gas-phase populations was instead 76.82° , the discrepancy being due to the fact that the α population dominates in the gas phase, while the β population dominates in solution.

To end this section, we recall that the two cases reported above are just examples representative of a large family of systems, where the effects of geometry optimization and conformational analysis are particularly relevant. Other studies on this topic have been performed, also by our group [90], leading to the general conclusion that a careful investigation of solvent-induced geometrical and conformational effects is required whenever the description and prediction of molecular properties of floppy systems is sought.

4.4.5.3 Local Field Effects

The description of molecular response and spectroscopic properties for systems surrounded by some kind of environment should also take into account the fact that, generally, the external (radiation) field experienced by a molecule in the solvent does not match the external field defined *in vacuo*. Such “local-field” effects are of general occurrence for any spectroscopic and response property, and play a role in the direct comparison of calculated and experimental findings. This problem has been faced in the past both by our group and others for several properties, by resorting to both classical or quantum-mechanical approaches [28, 91, 92].

An important point in this discussion is that classical theories for these effects based on the Onsager–Lorentz theory of dielectric polarization formulate a common scaling factor for a given property and a given solvent, which is independent of the nature of the solute, but in terms of which the property in solution is expressed in terms of the hypothetical quantity *in vacuo*. In contrast, quantum-mechanical approaches to this topic give different factors, depending on the nature of the solute and solvent pair. For this reason, the contribution due to “local-field” effects, which can be simply seen as a multiplicative factor larger than one in classical theories, can instead be very different if quantum-mechanical approaches are used.

The relevance of local-field effects (as evaluated by means of quantum-mechanical models) is difficult to assess in a general manner: suffice to say that they should always be evaluated whenever a direct comparison of calculated and experimental findings is to be achieved. However, their relevance depends on the property and on the particular system. Especially, researchers in our group have shown in the past that for some high order response properties, such as the Kerr birefringence or the electric field-induced second-harmonic generation (EFISH) process, the use of a quantum-mechanical formulation of the local field, by resorting to a PCM description of the medium, causes large differences with respect to classical theories [93, 94] and is even able to align a set of homologous compounds along their experimental behavior [94]. Less pronounced is the difference for other properties of lower order, such as optical rotation, IR, Raman, VCD and VROA intensities [90b,95–98], even if in the case of vibrational spectroscopies the factor changes, moving from a normal mode to another (the classical value is obviously the same in this case).

4.4.5.4 Dynamic Effects

Among solvent-induced effects is the dynamic (time-dependent) solvent response to changes in the solute charge distribution. Such effects, which were originally formulated for treating solvent effects of electronic absorption and emission processes (see the following section), play also a relevant role for other spectroscopic and response properties. Thus, the correct solvation regime has to be determined with care whenever a molecular system is in some way interacting with a time-dependent external field.

Starting from the assumption that the solvent polarization can be formally decomposed into different contributions, each related to the various degrees of freedom of the solvent molecules, in common practice such contributions are grouped in two terms only [28, 99]: one accounts for all motions slower than those involved in the physical phenomenon under examination (the “slow” polarization), the other includes the faster contributions (the “fast” polarization). As a further assumption, only the slow motions are instantaneously equilibrated to the momentary molecular charge distribution changing as a result of the interaction with the external field, whereas the fast processes cannot readjust, giving rise to a non-equilibrium solvent–solute system.

This partition and the following non-equilibrium approach were originally formulated and commonly applied to electronic processes, as well as to the evaluation of solute response to external oscillating fields. In this case the fast term is connected with the polarization of the electron clouds and the slow contribution accounts for all the nuclear degrees of freedom of the solvent molecules (see the following section). In the case of vibrations of solvated molecules, the same two-term partition can be assumed, but in this case the “slow” term will account for the contributions arising from the motions of the solvent molecules as a whole (translations and rotations), whereas the “fast” term will take into account the internal molecular motions (electronic and vibrational) [100]. After a shift from a previously reached equilibrium solute–solvent system, the fast polarization will be still in equilibrium with the new solute charge distribution but the slow polarization remains fixed to the value corresponding to the initial state solute charge distribution.

Such a scheme has been implemented within the PCM framework to treat non-equilibrium effects on IR frequencies and intensities, showing that non-equilibrium shifts are in very good agreement with experiment, whereas a pure equilibrium model fails in reproducing the solvent-induced effects.

We now move to Raman intensities described classically as resulting from the modulation due to vibrational motions in the electric field-induced oscillating dipole moment. Such a modulation has the frequency of molecular vibrations, whereas the dipole moment oscillations have the frequency of the external electric field. Thus, the dynamic aspects of Raman scattering are to be described in terms of two time scales. One is connected to the vibrational motions of the nuclei, the other to the oscillation of the radiation electric field (which gives rise to oscillations in the solute electronic density). In the presence of a solvent medium, both mentioned time scales give rise to non-equilibrium effects in the solvent response, being much faster than the time scale of the solvent inertial response. Both dynamic (non-equilibrium) responses

have been formulated within PCM, showing that, even though vibrational non-equilibrium effects have been shown to give substantial corrections to IR absorption intensities of molecules in solution, these effects are in general negligible for Raman intensities [101]. Non-equilibrium effects have also been tested in the case of VCD [90b] and VROA spectra [98].

4.4.6

Effect of the Environment on Formation and Relaxation of Excited States

The accurate modeling of electronically excited state formation and relaxation of molecules in solution is a very important problem not only in photochemical or spectroscopic studies but especially in material science and biology. Despite this recognized importance, the progress achieved so far is not as much as that obtained for ground state phenomena. The modeling of electronically excited molecules when interacting with an external medium requires the introduction of the concept of time progress, a concept that can be safely neglected in treating most of the properties and processes of solutes in their ground states. In these cases and also when introducing reaction processes one can always reduce the analysis to a completely equilibrated solute–solvent system. Conversely, when attention is shifted towards dynamic phenomena such as those involved in electronic transitions (absorptions and/or emissions), or towards relaxation phenomena such as those that describe the time evolution of the excited state, one has to introduce new models, in which solute and solvent have proper response times that have not to be coherent or at least not before very long times.

To better understand this point, it is convenient to introduce a classification of the sources of the dynamical behavior of the environment into two main components (for simplicity we identify the environment as an isotropic and homogeneous solvent). One is represented by the molecular motions inside the solvent due to changes in the charge distribution and/or in the geometry of the solute system. The solute when immersed in the solvent produces an electric field inside it, which can modify the structure of the liquid, for example, inducing phenomena of alignment and/or preferential orientation of the solvent molecules around the solute. Specific time scales of the order of the rotational and translational times proper of liquids characterize these molecular motions. Analogously, we can assume that the solvent molecules are subjected to internal geometrical variations (i.e., vibrations) that will be described by specific shorter time scales. Translational, rotational and/or vibrational motions involve nuclear displacements and therefore they are collectively indicated as “nuclear motions.” The other important component of the dynamic nature of the medium, complementary to the nuclear one, is that induced by motions of the electrons inside each solvent molecule; these motions are extremely fast and they represent the electronic polarization of the solvent. These nuclear and electronic components, owing to their different dynamic behavior, will give rise to different effects. In particular, the electronic motions can be considered as instantaneous and, as a result, the part of the solvent response they give rise to is always readjusted to any change, even if fast, in the charge distribution of the solute. Conversely, solvent

nuclear motions, by far slower, can be delayed with respect to fast changes in the solute. As a consequence, they can give rise to solute–solvent systems not completely equilibrated in the time interval relevant to the dynamical phenomenon under consideration. This condition of non-equilibrium will successively evolve towards a more stable and completely equilibrated state in a time interval that will depend on the specific system under scrutiny.

This very general picture of the coupling of the solute and the environment dynamics clearly has important consequences in all the processes that involve formation and relaxation of electronically excited states in solvated systems. In the following two sections we focus in particular on two families of such processes, namely, electronic transitions (UV-VIS and fluorescence spectroscopies) and photoinduced electron and energy transfers.

4.4.7

Electronic Transitions and Related Spectroscopies

It has long been well known that solvents strongly influence the electronic spectral bands of individual species measured by various spectrometric techniques (UV/visible, fluorescence spectroscopy, etc.). Broadening of the absorption and fluorescence bands results from fluctuations in the structure of the solvation shell around the solute (this effect, called inhomogeneous broadening, superimposes homogeneous broadening because of the existence of a continuous set of vibrational sublevels) [102]. Moreover, shifts in absorption and emission bands can be induced by a change in the solvent nature or composition; these shifts, called solvatochromic shifts, are experimental evidence of changes in solvation energy. In fact, when surrounded by solvent molecules, solute ground and excited states are differently stabilized depending on the chemical nature of both solute and solvent molecules [103, 104].

As the solvatochromic shifts are experimental evidence of changes in solvation energy, they have been largely used to construct empirical polarity scales for the different solvents. The use of solvatochromism of betaine dyes (Figure 4.2) as a probe of solvent polarity proposed by Reichardt is worth mentioning [105]. The exceptionally large solvatochromism shown by these compounds can be explained considering that in their ground state they are zwitterions while, upon excitation, electron transfer occurs exactly in the direction of canceling this charge separation. As a result, the dipole moment, which is very large in the ground state, becomes nearly zero in the excited state and thus solvent interactions drastically change, leading to the observed negative solvatochromism.

An alternative approach to quantifying polarity effects was proposed by Kamlet and Taft [106]. The scale of Kamlet and Taft deserves special recognition not only because it has been successfully applied in many studies (not limited to UV or fluorescence spectra, and including many other physical or chemical parameters like reaction rate, equilibrium constant, etc.) but also because it introduces a useful partition of the solvent effect into distinct contributions such as polarity/polarizability effects and hydrogen bonding.

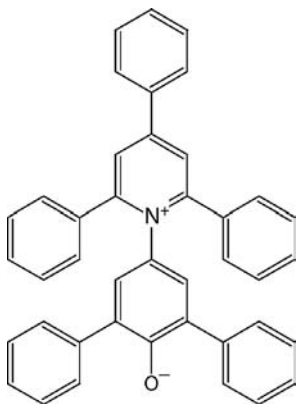


Figure 4.2 Example of a betaine dye.

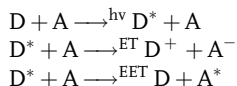
Obviously, these empirical polarity scales cannot be used as real predictive and interpretative tools to study environment effects on electronic transitions (and the related spectroscopies) of solvated molecules. These tools to be accurate need to explicitly account for the quantum nature of the process. However, due to the complexity and the extremely large dimensions of the whole system a detailed quantum-mechanical description of all the components of the system is prohibitive and in general focused approaches are introduced. Within these approaches an accurate QM description is used for the molecular system of interest (the chromophore, possibly including small portions of the environment) and a less accurate description for the remainder. Hybrid QM/MM is a well-known example but also QM/continuum approaches belong to the same class of focused models. The different philosophy beyond the two classes of methods leads to important differences in both the physical and the computational aspects of their applications, as well as in their range of applicability. The methods based on explicit representations of the environment yield information on specific configurations of the environment around the chromophore, while the continuum models give only an averaged picture of it. However, QM/MM requires many more calculations than continuum models to obtain a correct statistical description. This much larger computational cost of QM/MM is particularly disadvantageous in the study of excited states as the QM level required is generally quite expensive even for a single calculation on an isolated system; thus, the necessity to repeat the calculation many times makes the approach very expensive (or even not feasible). Conversely, in QM/continuum approaches the additional computational cost with respect to gas-phase calculations remains very limited whatever the QM level adopted. In addition, continuum solvation models include effects of mutual polarization between the solute and the environment (also those due to a possible non-equilibrium solvation), while standard QM/MM methods are based on non-polarizable force fields. QM/MM approaches including environment polarization have been proposed and also applied to the study of excited states of solvated systems but they still represent a minority [107].

Among the possible continuum solvation models, the PCM approach described in Section 4.3.1 remains one of the most widely used in computational studies on UV-VIS (and fluorescence) spectra of solvated systems (see, for example, two recent reviews [108]). In particular, the coupling of PCM with the time-dependent density functional theory (TDDFT) can indeed represent an efficient general approach to describing solvent effects on the spectral properties and, more generally, on the structures and other properties of solvated excited states. TDDFT can include correlation effects through the exchange-correlation potential for both the ground and the excited states without adding a significant computational effort, whereas PCM can include all required solvation effects necessary to describe excited states in a formally simple and computationally efficient way. For example, possible non-equilibrium solvation can be easily included within the PCM framework just by introducing a partition of the apparent charges representing the response of the solvent into two sets, the dynamic and inertial charges, which account for the fast (electronic) and slow (nuclear) part of the polarization, as discussed above.

4.4.8

Photoinduced Electron and Energy Transfers

In excitation (or resonance) energy transfers, the excitation energy from a donor system in an electronic excited state (D^*) is transferred to a sensitizer (or acceptor) system (A). Alternatively, in photoinduced electron transfer (ETs), a donor (D) transfers an electron to an acceptor (A) after photoexcitation of one of the components:



Owing to an ever-growing interest in both ET [109, 110] and excitation energy transfer (EET) processes [111, 112] there have been significant efforts to develop efficient and reliable theoretical and computational tools. In particular, as the most important quantity involved in the two processes is the electronic coupling between donor and acceptor moieties, the most proper approach is represented by quantum mechanics. For both ET and EET processes, a further theoretical issue to be considered is the inclusion of environment effects. For ET processes, for example, solvent effects have received extensive attention regarding so-called solvent reorganization energies and reaction free energies.

Solvent reorganization energy (at the heart of Marcus ET theory) [113] represents the energy involved in disrupting the local solvent structure during any charge transfer between two moieties. According to the well-known Marcus formula, this energy can be calculated by assuming that D and A are point charges at the center of spheres of known radius within a continuum dielectric with given static and the optical dielectric permittivities. Over the years, several generalizations of the Marcus formula have been proposed: Most of them are based on the concept of non-equilibrium solvation free energies [114]. Such generalizations have been also used

within the PCM framework and in other continuum solvation models using a QM description of the D/A pair. In contrast, a QM description of the coupling including the effect of the environment is less common. Solvation models (combined with linear response approaches for the characterization of excited states) have been combined with the generalized Mulliken–Hush [115] approach to ET, showing that the polarity of the solvent can significantly modulate the electronic coupling between a donor and an acceptor.

In contrast, EET has been historically modeled in terms of two main schemes: the Förster transfer [116], a resonant dipole–dipole interaction, and the Dexter transfer [117], based on wavefunction overlap. The effects of the environment were early recognized by Förster in his unified theory of EET, where the Coulomb interaction between donor and acceptor transition dipoles is screened by the presence of the environment (represented as a dielectric) through a screening factor $1/n^2$, where n is the solvent refractive index. This description is clearly an approximation of the global effects induced by a polarizable environment on EET. In fact, the presence of a dielectric environment not only screens the Coulomb interactions as formulated by Förster but also affects all the electronic properties of the interacting donor and acceptor [118].

More accurate descriptions of the effects of dielectric environments on EET have been successively given using classical dielectric theory [119] and quantum electrodynamics (QED) theory [120]. In all these theories, however, point dipole (or higher multipole) levels of description of the chromophores result in essentially Förster’s coupling scaled by a prefactor (generally a screening contribution multiplied by the square of the local field factor) that does not depend on the orientation and alignment of the two transition dipoles. Only more recently have more accurate quantum-mechanical descriptions of the interacting chromophores appeared, using either semiempirical or *ab initio* approaches [121].

In particular, an advance in this field has been the development of a general QM theory to study EET in solution developed in the context of continuum solvation models. Such a theory is based on the linear response (LR) approach (either within Hartree–Fock or DFT) and introduces the solvent effects in terms of the PCM approach. A unique characteristic of this model is that both polarizing effects on the interacting molecules and screening effects are included in a coherent and self-consistent way [122]. This model has been applied to the study of EET between molecules in liquid solutions [123] and at liquid/gas interfaces [124], and to exciton splitting in conjugated molecular materials [125]. More recently, the model has been applied to examine the screening for a set of over 100 pairs of chromophores (chlorophylls, carotenoids and bilins) taken from structural models of photosynthetic proteins [126]. In that study, we found a striking exponential attenuation of the screening factor (s) at separations less than about 20 Å, thus interpolating between the limits of no apparent screening and a significant attenuation of the EET rate. This observation reveals a previously unidentified contribution to the distance dependence of the EET rate, which has particularly important consequences for the development of quantitative EET models such as those actively pursued in the study of photosynthetic light-harvesting systems and conjugated polymers.

References

- 1 Bonaccorsi, R., Scrocco, E., and Tomasi, J. (1970) *J. Chem. Phys.*, **52**, 5270–5284.
- 2 (a) Scrocco, E. and Tomasi, J. (1973) *Top. Curr. Chem.*, **42**, 95–170; (b) Scrocco, E. and Tomasi, J. (1978) *Adv. Quantum Chem.*, **11**, 116–193.
- 3 Ghio, C., Tomasi, J., Weill, J., and Sillion, B. (1986) *J. Mol. Struct. (Theochem)*, **135**, 299–328.
- 4 Murray, J.S. and Sen, K. (eds) (1996) *Molecular Electrostatic Potentials: Concepts and Applications*, Elsevier, Amsterdam.
- 5 Pullman, B. (1990) *Int. J. Quantum Chem.*, **38**, 81–92.
- 6 Boys, S.F. (1966) *Quantum Theory of Atoms, Molecules and the Solid State* (ed. P.O. Löwdin), Academic Press.
- 7 Reed, A.E. and Weinhold, F. (1985) *J. Chem. Phys.*, **83**, 735–746.
- 8 Jeziorski, B., Moszynski, R., and Szalewicz, K. (1994) *Chem. Rev.*, **94**, 1887–1930.
- 9 Bonaccorsi, R., Scrocco, E., and Tomasi, J. (1977) *J. Am. Chem. Soc.*, **99**, 4546–4554.
- 10 Agresti, A., Bonaccorsi, R., and Tomasi, J. (1978) *Theor. Chim. Acta*, **53**, 215–220.
- 11 (a) Bader, R.F.W. and Matta, C.F. (2004) *J. Phys. Chem. A*, **108**, 8385–8394; (b) Matta, C.F. and Bader, R.F.W. (2006) *J. Phys. Chem. A*, **110**, 6365–6371; (c) Bader, R.W.F. (1990) *Atoms in Molecules: A Quantum Theory*, Oxford University Press, Oxford; (d) Matta, C.F. and Boyd, R.J. (eds) (2007) *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*, Wiley-VCH Verlag GmbH, Weinheim.
- 12 Mulliken, R. (1955) *J. Chem. Phys.*, **23**, 1833–1840.
- 13 Momany, F.A. (1978) *J. Phys. Chem.*, **82**, 592–601.
- 14 Williams, D.E. (1988) *Adv. At. Mol. Phys.*, **23**, 87.
- 15 Tomasi, J., Bonaccorsi, R., and Cammi, R. (1991) *Theoretical Models of Chemical Bonding Part 4* (ed. Z.B. Maksic), Springer, Berlin, pp. 229–268.
- 16 Santra, B., Michaelides, A., Fuchs, M., Tkatchenko, A., Filippi, C., and Schleffer, M. (2008) *J. Chem. Phys.*, **129**, 194111.
- 17 Kitaura, K. and Morokuma, K. (1976) *Int. J. Quantum Chem.*, **10**, 325–340.
- 18 Tomasi, J., Mennucci, B., and Cammi, R. (1996) *Molecular Electrostatic Potentials. Concepts and Applications* (eds J.S. Murray and K. Sen), Elsevier, Amsterdam, pp. 1–103.
- 19 Tomasi, J. (1982) *Molecular Interactions*, vol 3 (eds H. Ratajczak and W. Orville-Thomas), John Wiley & Sons, Inc., New York, pp. 119–181.
- 20 Alagona, G., Ghio, C., Cammi, R., and Tomasi, J. (1988) *Molecules in Physics, Chemistry, and Biology*, vol 2 (ed. J. Maruani), Kluwer, Dordrecht, pp. 507–559.
- 21 Burnham, C.J. and Xantheas, S.S. (2002) *J. Chem. Phys.*, **116**, 1479–1492.
- 22 Car, R. and Parrinello, M. (1985) *Phys. Rev. Lett.*, **55**, 2471–2474.
- 23 Xantheas, S.S., Burnham, C.J., and Harrison, R.J. (2002) *J. Chem. Phys.*, **116**, 1493–1499.
- 24 Liu, Y.P., Kim, K., Berne, B.J., Friesner, R.A., and Rick, S.W. (1998) *J. Chem. Phys.*, **108**, 4739–4755.
- 25 Tolosa Arroyo, S., Sanson Martin, J.A., and Hidalgo Garcia, A. (2007) *J. Phys. Chem. A*, **111**, 339–344.
- 26 Akin-Ojo, O., Song, Y., and Wang, F. (2008) *J. Chem. Phys.*, **129**, 064108.
- 27 Tomasi, J. and Persico, M. (1994) *Chem. Rev.*, **94**, 2027–2094.
- 28 Tomasi, J., Mennucci, B., and Cammi, R. (2005) *Chem. Rev.*, **105**, 2999–3093.
- 29 Cammi, R. and Tomasi, J. (1996) *Int. J. Quantum Chem.*, **60**, 297–306.
- 30 Pierotti, R.A. (1976) *Chem. Rev.*, **76**, 717–726.
- 31 Onsager, L. (1936) *J. Am. Chem. Soc.*, **58**, 1486–1493.
- 32 Cancès, E. (2007) *Continuum Solvation Models in Chemical Physics: from Theory to Applications* (eds B. Mennucci and R. Cammi), John Wiley & Sons, Ltd., Chichester, pp. 29–48.
- 33 Pomelli, C.S. (2007) *Continuum Solvation Models in Chemical Physics: from Theory to Applications* (eds B. Mennucci and R. Cammi), John Wiley & Sons, Ltd., Chichester, pp. 48–63.

- 34 Amovilli, C. and Mennucci, B. (1997) *J. Phys. Chem. B*, **101**, 1051–1057.
- 35 Floris, F.M., Tomasi, J., and Pascual-Ahuir, J.L. (1991) *J. Comput. Chem.*, **12**, 784–791.
- 36 Cammi, R., Verdolino, V., Mennucci, B., and Tomasi, J. (2008) *Chem. Phys.*, **344**, 135–141.
- 37 Mc Weeny, R. (1992) *Methods of Molecular Quantum Mechanics*, Academic Press, London.
- 38 Floris, F. and Tomasi, J. (1989) *J. Comput. Chem.*, **10**, 616–627.
- 39 Curutchet, C., Orozco, M., Luque, J.F., Mennucci, B., and Tomasi, J. (2006) *J. Comput. Chem.*, **27**, 1769–1780.
- 40 Frisch, M.J., Trucks, G.W., Schlegel, H.B. et al. (2004) Gaussian 03, Revision C. 02, Gaussian, Inc., Wallingford CT.
- 41 Wujec, M., Siwek, A., Dzierzawska, J., Rostkowski, M., Kaminski, R., and Paneth, P. (2008) *J. Phys. Chem. B*, **112**, 12414–12419.
- 42 Cossi, M., Barone, V., Mennucci, B., and Tomasi, J. (1998) *Chem. Phys. Lett.*, **286**, 253–260.
- 43 Mennucci, B. and Cammi, R. (2003) *Int. J. Quantum Chem.*, **93**, 121–130.
- 44 Frediani, L., Cammi, R., Corni, S., and Tomasi, J. (2004) *J. Chem. Phys.*, **120**, 3893–3907.
- 45 Miertus, S., Scrocco, E., and Tomasi, J. (1981) *Chem. Phys.*, **55**, 117–129.
- 46 Barone, V. and Cossi, M. (1998) *J. Phys. Chem. A*, **102**, 1995–2001.
- 47 Cancès, E., Mennucci, B., and Tomasi, J. (1997) *J. Chem. Phys.*, **107**, 3032–3041.
- 48 Caricato, M., Mennucci, B., and Tomasi, J. (2004) *J. Phys. Chem. A*, **108**, 6248–6256.
- 49 Gordon, M.S., Freytag, M.A., Bandyopadhyay, P., Jensen, H.J., Kayris, V., and Stevens, W.J. (2001) *J. Phys. Chem. A*, **105**, 293–307.
- 50 Bandyopadhyay, P., Gordon, M.S., Mennucci, B., and Tomasi, J. (2002) *J. Chem. Phys.*, **116**, 5023–5032.
- 51 Li, H., Pomelli, C.S., and Jensen, J.H. (2003) *Theor. Chem. Acc.*, **109**, 71–84.
- 52 Jensen, J.H., Li, H., Robertson, A.D., and Molina, P.A. (2005) *J. Phys. Chem. A*, **109**, 6634–6643.
- 53 Maseras, F. and Morokuma, K. (1995) *J. Comput. Chem.*, **16**, 1170–1179.
- 54 Vreven, T., Mennucci, B., da Silva, C.O., Morokuma, K., and Tomasi, J. (2001) *J. Chem. Phys.*, **115**, 62–72.
- 55 Klamt, A. and Schüürman, G. (1993) *J. Chem. Soc. Perkins Trans. 2*, 799–805.
- 56 Klamt, A. (2005) *COSMO-RS from Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*, Elsevier, Amsterdam.
- 57 Eckert, F. and Klamt, A. (2007) COSMOtherm, Version C2. 1, COSMOlogic GmbH & Co KG, Leverkusen, Germany.
- 58 Stefanovich, E.V. and Truong, T.N. (1995) *Chem. Phys. Lett.*, **244**, 65–74.
- 59 Soteras, I., Curutchet, C., Bidon-Chanal, A., Orozco, M., and Luque, F.J. (2005) *J. Mol. Struct. (THEOCHEM)*, **727**, 29–40.
- 60 Chipman, D.M. (2002) *J. Chem. Phys.*, **116**, 10129.
- 61 Rinaldi, D., Bouchy, A., Rivail, J.L., and Dillet, V. (2002) *J. Chem. Phys.*, **104**, 2343–2350.
- 62 Rivail, J.L. and Rinaldi, D. (1976) *Chem. Phys.*, **18**, 233–242.
- 63 Mikkelsen, K.V., Jørgensen, P., and Jensen, H.J. (1994) *J. Chem. Phys.*, **100**, 6597–6607.
- 64 Marenich, A.V., Olson, R.M., Kelly, C.P., Cramer, C.J., and Truhlar, D.G. (2007) *J. Chem. Theory Comput.*, **3**, 2011–2033.
- 65 Cramer, C.J. and Truhlar, D.G. (1991) *J. Am. Chem. Soc.*, **113**, 8305–8311.
- 66 Cramer, C.J. and Truhlar, D.G. (2008) *Acc. Chem. Res.*, **41**, 760–768.
- 67 Cortis, C.M., Langlois, J.M., Beachy, M.D., and Friesner, R.A. (1996) *J. Chem. Phys.*, **105**, 5472–5484.
- 68 Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001) *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 10037–10041.
- 69 Honig, B. et al., DelPhi: http://wiki.c2b2.columbia.edu/honiglab_public/index.php/Software:DelPhi. Last accessed: 14/10/2009.
- 70 Bashford, D. (1997) An object-oriented programming suite for electrostatic effects in biological molecules. In *Scientific Computing in Object-Oriented Parallel Environments*, Vol. 1343; Ishikawa, Y., Oldehoeft, R.R., Reynnders, J.V.W., and Tholburn, M. (eds), Springer, Berlin.

- 71 Briggs, J.M., Madura, J.D., Davis, M.E., Gilson, M.K., Antosiewicz, J., Luty, B.A., Wade, R.C., Bagheri, B., Ilin, A., Tan, R.C., and McCammon, J.A., UHBD, University of California, <http://mccammon.ucsd.edu/uhbd.html>. Last accessed: 14/10/2009.
- 72 Marten, B., Kim, K., Cortis, C., Friesner, R.A., Murphy, R.B., Ringnalda, M.N., Sitkoff, D., and Honig, B. (1996) *J. Phys. Chem.*, **110**, 11775–11788.
- 73 Jaguar, Schroedinger Inc., <http://www.schroedinger.com>. Last accessed: 14/10/2009.
- 74 Klamt, A., Mennucci, B., Tomasi, J., Barone, V., Curutchet, C., Orozco, M., and Luque, F.J. (2009) *Acc. Chem. Res.*, **42**, 489–492.
- 75 Cramer, C.J. and Truhlar, D.G. (2009) *Acc. Chem. Res.*, **42**, 493–497.
- 76 Tomasi, J., Mennucci, B., Cammi, R., and Cossi, M. (1997) *Computational Approaches to Biochemical Reactivity* (eds G. Naray-Szabo and A. Warshel) Kluwer, Dordrecht, pp. 1–102.
- 77 Cramer, C.J. and Truhlar, D.G. (1999) *Chem. Rev.*, **99**, 2161–2200.
- 78 Cappelli, C., Mennucci, B., da Silva, C.O., and Tomasi, J. (2000) *J. Chem. Phys.*, **112**, 5382–5392.
- 79 Mennucci, B. (2002) *J. Am. Chem. Soc.*, **124**, 1506–1515.
- 80 Tissandier, M.D., Cowen, K.A., Feng, W.Y., Gundlach, E., Cohen, M.H., Earhart, A.D., Coe, J.V., and Tuttle, T.R. Jr. (1998) *J. Phys. Chem. A*, **102**, 7787–7794.
- 81 (a) da Silva, C.O., da Silva, E.C., and Nascimento, M.A.C. (1999) *J. Phys. Chem. A*, **103**, 11194–11199; (b) Silva, C.O., da Silva, E.C., and Nascimento, M.A.C. (2000) *J. Phys. Chem. A*, **104**, 2402–2409; (c) da Silva, C.O. and Nascimento, M.A.C. (2002) *Adv. Chem. Phys.*, **123**, 423–468.
- 82 Pliego, J.R. (2003) *Chem. Phys. Lett.*, **367**, 145–149.
- 83 Tomasi, J. (2007) *Continuum Solvation Models in Chemical Physics: from Theory to Applications* (eds B. Mennucci and R. Cammi), John Wiley & Sons, Ltd., Chichester, pp. 1–28.
- 84 Coitino, E.L., Tomasi, J., and Ventura, O.N. (1994) *J. Chem. Soc., Faraday Trans.*, **90**, 1745–1755.
- 85 Bonaccorsi, R., Cimbrigaglia, R., Tomasi, J., and Miertus, S. (1983) *J. Mol. Struct. (Theochem)*, **94**, 11–23.
- 86 Barone, V., Impropa, R., and Rega, N. (2008) *Acc. Chem. Res.*, **41**, 605–616.
- 87 Lee, K.-K., Hahn, S., Oh, K.-I., Choi, J.S., Joo, C., Lee, H., Han, H., and Cho, M. (2006) *J. Phys. Chem. B*, **110**, 18834–18843.
- 88 Cappelli, C. and Mennucci, B. (2008) *J. Phys. Chem. B*, **112**, 3441–3450.
- 89 da Silva, C.O., Mennucci, B., and Vreven, T. (2004) *J. Org. Chem.*, **69**, 8161–8164.
- 90 (a) Cappelli, C., Mennucci, B., and Monti, S. (2005) *J. Phys. Chem. A*, **109**, 1933–1943; (b) Cappelli, C., Corni, S., Mennucci, B., Cammi, R., and Tomasi, J. (2002) *J. Phys. Chem. A*, **106**, 12331–12339; (c) Cappelli, C., Monti, S., and Rizzo, A. (2005) *Int. J. Quantum Chem.*, **104**, 744–757; (d) Cappelli, C., Bronco, S., and Monti, S. (2005) *Chirality*, **17**, 577–589.
- 91 Cappelli, C. (2007) *Continuum Solvation Models in Chemical Physics: from Theory to Applications* (eds B. Mennucci and R. Cammi), John Wiley & Sons, Ltd., Chichester, pp. 167–179.
- 92 Cammi, R. and Mennucci, B. (2007) *Continuum Solvation Models in Chemical Physics: from Theory to Applications* (eds B. Mennucci and R. Cammi), John Wiley & Sons, Ltd., Chichester, pp. 238–251.
- 93 Cappelli, C., Mennucci, B., Cammi, R., and Rizzo, A. (2005) *J. Phys. Chem. B*, **109**, 18706–18714.
- 94 Ferrighi, L., Frediani, L., Cappelli, C., Saøek, P., Agren, H., Helgaker, T., and Ruud, K. (2006) *Chem. Phys. Lett.*, **425**, 267–272.
- 95 Mennucci, B., Tomasi, J., Cammi, R., Cheeseman, J.R., Frisch, M.J., Devlin, F.J., Gabriel, S., and Stephens, P.J. (2002) *J. Phys. Chem. A*, **106**, 6102–6113.
- 96 Cammi, R., Cappelli, C., Corni, S., and Tomasi, J. (2000) *J. Phys. Chem. A*, **104**, 9874–9879.
- 97 Corni, S., Cappelli, C., Cammi, R., and Tomasi, J. (2001) *J. Phys. Chem. A*, **105**, 8310–8316.

- 98 Pecul, M., Lamparska, E., Frediani, L., Cappelli, C., and Ruud, K. (2006) *J. Phys. Chem. A*, **110**, 2807–2815.
- 99 Mennucci, B., Cammi, R., and Tomasi, J. (1998) *J. Chem. Phys.*, **109**, 2798–2807.
- 100 Cappelli, C., Corni, S., Cammi, R., Mennucci, B., and Tomasi, J. (2000) *J. Chem. Phys.*, **113**, 11270–11279.
- 101 Cappelli, C., Corni, S., and Tomasi, J. (2001) *J. Chem. Phys.*, **115**, 5531–5535.
- 102 Nemkovich, N.A., Rubinov, A.N., and Tomin, I.T. (1991) Topics in fluorescence spectroscopy, in *Principles*, vol 2 (ed. J.R. Lakowicz), Plenum Press, New York.
- 103 Valeur, B. (2001) *Molecular Fluorescence: Principles and Applications*, Wiley-VCH Verlag GmbH.
- 104 Suppan, P. and Ghoneim, N. (1997) *Solvatochromism*, The Royal Society of Chemistry, Cambridge, UK.
- 105 Reichardt, C. (1990) *Solvents and Solvent Effects in Organic Chemistry*, 2nd edn, VCH, Weinheim.
- 106 Kamlet, M.J., Abboud, J.L., and Taft, R.W. (1977) *J. Am. Chem. Soc.*, **99**, 6027–6038.
- 107 (a) Kongsted, J., Osted, A., Mikkelsen, K.V., Astrand, P.O., and Christiansen, O. (2004) *J. Chem. Phys.*, **121**, 8435–8445; (b) Jacob, C.R., Neugebauer, J., Jensen, L., and Visscher, L. (2006) *Phys. Chem. Chem. Phys.*, **8**, 2349–2359; (c) Ohrn, A. and Karlstrom, G. (2006) *Mol. Phys.*, **104**, 3087–3099; (d) Lin, Y.L. and Gao, J.L. (2007) *J. Chem. Theory Comput.*, **3**, 1484–1493.
- 108 (a) Mennucci, B. (2007) *Continuum Solvation Models in Chemical Physics: from Theory to Applications* (eds B. Mennucci and R. Cammi), John Wiley & Sons, Ltd., Chichester, pp. 110–122; (b) Cammi, R., Mennucci, B., and Shukla, M.K. (2008) *Radiation Induced Molecular Phenomena in Nucleic Acids: A Comprehensive Theoretical and Experimental Analysis; Challenges and Advances in Computational Chemistry and Physics*, vol. 5 (ed. J. Leszczynski), Springer, Berlin.
- 109 Newton, M.D. (1991) *Chem. Rev.*, **91**, 767–792.
- 110 Balzani, V. (ed.) (2001) *Electron Transfer in Chemistry*, vols. 1–5, Wiley-VCH Verlag GmbH, Weinheim.
- 111 Scholes, G.D. (2003) *Annu. Rev. Phys. Chem.*, **54**, 57–87.
- 112 Andrews, D.L. and Demidov, A.A. (eds) (1999) *Resonance Energy Transfer*, John Wiley & Sons, Inc., New York.
- 113 (a) Marcus, R.A. (1956) *J. Chem. Phys.*, **24**, 966–978; (b) Marcus, R.A. (1956) *J. Chem. Phys.*, **24**, 979–989.
- 114 (a) Newton, M.D. and Friedman, H.L. (1988) *J. Chem. Phys.*, **88**, 4460–4472; (b) Liu, Y.-P. and Newton, M.D. (1995) *J. Phys. Chem.*, **99**, 12382–12386; (c) Basilevsky, M.V., Chudinov, G.E., and Newton, M.D. (1994) *Chem. Phys.*, **179**, 263–278; (d) Basilevsky, M.V., Chudinov, G.E., Rostov, I.V., Liu, Y.-P., and Newton, M.D. (1996) *J. Mol. Struct. (THEOCHEM)*, **371**, 191–203; (e) Basilevsky, M.V., Parsons, D.F., and Vener, M.V. (1998) *J. Chem. Phys.*, **108**, 1103–1114.
- 115 Cave, R.J. and Newton, M.D. (1996) *Chem. Phys. Lett.*, **249**, 15–19.
- 116 Forster, T. (1948) *Annalen der Physik*, **437**, 55–75.
- 117 Dexter, D.L. (1953) *J. Chem. Phys.*, **21**, 836–850.
- 118 Knox, R.S. and van Amerongen, H. (2002) *J. Phys. Chem. B*, **106**, 5289–5293.
- 119 Agranovich, V.M. and Galanin, M.D. (1982) *Electronic Excitation Energy Transfer in Condensed Matter*, North-Holland, Amsterdam.
- 120 (a) Juzeliunas, G. and Andrews, D.L. (1994) *Phys. Rev. B*, **49**, 8751–8763; (b) Andrews, D.L. and Juzeliunas, G. (1994) *J. Lumin.*, **60/61**, 834–837.
- 121 (a) Tretiak, S., Middleton, C., Chernyak, V., and Mukamel, S. (2000) *J. Phys. Chem. B*, **104**, 9540–9553; (b) Hsu, C.-P., Fleming, G.R., Head-Gordon, M., and Head-Gordon, T. (2001) *J. Chem. Phys.*, **114**, 3065–3072.
- 122 Iozzi, M.F., Mennucci, B., Tomasi, J., and Cammi, R. (2004) *J. Chem. Phys.*, **120**, 7029–7040.
- 123 (a) Curutchet, C. and Mennucci, B. (2005) *J. Am. Chem. Soc.*, **127**, 16733–16744; (b) Russo, V., Curutchet, C., and Mennucci, B. (2007) *J. Phys. Chem. B*, **111**, 853–863.

- 124 Curutchet, C., Cammi, R., Mennucci, B., and Corni, S. (2006) *J. Chem. Phys.*, **125**, 054710.
- 125 Mennucci, B., Tomasi, J., and Cammi, R. (2004) *Phys. Rev. B*, **70**, 205212.
- 126 (a) Scholes, G.D., Curutchet, C., Mennucci, B., Cammi, R., and Tomasi, J. (2007) *J. Phys. Chem. B*, **111**, 6978–6982; (b) Curutchet, C., Scholes, G.D., Mennucci, B., and Cammi, R. (2007) *J. Phys. Chem. B*, **111**, 13253–13265.

5

The Fast Marching Method for Determining Chemical Reaction Mechanisms in Complex Systems

*Yuli Liu, Steven K. Burger, Bijoy K. Dey, Utpal Sarkar,
Marek R. Janicki, and Paul W. Ayers*

5.1

Motivation

Suppose that one is given a set of molecules (the reagents) and their reaction conditions (solvent or gas phase, temperature, etc.). Does a chemical reaction occur? What kind of reaction? What is (are) the product(s)? How and why does the reaction happen? These are the fundamental problems of chemistry.

The theoretical solution to these problems requires finding the chemical reaction pathway. For example, given the minimum energy path (MEP), one can determine molecular structures and energies of the reactants, products and transition states. The difference in energy between the reactants and products is the reaction energy; the difference in energy between the reactants and the transition state structure is the activation energy, which is related to the rate of reaction. The MEP provides key information about reaction thermodynamics and kinetics. In addition, tracing the MEP from the reactant, through reactive intermediates and the transition state, to the products gives us the chemical reaction mechanism. The reaction mechanism is the key to understanding how and why a reaction occurs, and it is important for optimizing reaction conditions and designing catalysts.

This chapter reviews our recent work on computational algorithms for finding the MEP, with particular emphasis on the fast marching method (FMM). In Section 5.2 we present more information about MEPs and review alternatives to the FMM. Section 5.3 provides algorithmic details about the FMM and some applications to small systems. Section 5.4 reviews the development of the quantum mechanics/molecular mechanics (QM/MM) methods for studying enzyme-catalyzed reactions and presents the idea of, and some preliminary work on, incorporating FMM with QM/MM methods. Section 5.5 summarizes our results to date and presents our perspective on future research directions.

5.2

Background

5.2.1

Minimum Energy Path

The reaction path is usually identified with the steepest descent path linking a transition state structure and its adjacent minima (such as reactant, product and reactive intermediates). When there is more than one steepest descent path, the one with the lowest energy barrier is the MEP. The steepest descent path defines the intrinsic reaction coordinate for a chemical reaction [1].

There are two main families of algorithms for finding the MEP: two end methods [2] and surface walking methods. The two end methods require a good guess for the path linking the reactant and product; if the mechanism in the initial guess is qualitatively correct [i.e., the path threads its way through the correct “mountain passes” on the potential energy surface (PES)] the right MEP will be located. Surface walking methods do not require an initial guess. They start exploring the PES from the reactant configuration, and eventually predict the products and the mechanism of the chemical reaction. Unfortunately, surface walking algorithms are usually either very expensive or, if a heuristic is used to simplify the calculation, they tend to be unreliable for complicated systems. The two end methods have great advantages from the viewpoint of computational cost and numerical stability.

5.2.2

Two End Methods

A simple example of a two end method is the nudged elastic band (NEB) method [3–8]. In this method an initial guess of the path is given that is divided up into a series of beads, with springs in between each bead. The beads are then propagated down the PES. One of the significant improvements of NEB over previous method is it decouples the problem, by projecting the spring force parallel to the path and the force from the potential perpendicular to the path. This preventing things like corner cutting of the path and ensures that the path will eventually converge to the MEP.

String methods are similar to NEB, but they do not use a fictitious force to ensure that the molecular conformations that define the reaction path are well-spaced. In the following paragraphs, we discuss the original string method of Ren and Vanden-Eijnden [9–11] and two improved string methods: the growing string method (GSM) [12] and the quadratic string method (QSM) [13].

Ren and Vanden-Eijnden proposed a zero-temperature string method [9, 14] for finding the MEP on the PES. Like the NEB method, the string method drives the initial path to the MEP by the perpendicular forces on the bead. The continuity of the path is ensured by reparameterizing the approximate path at each iteration so that the nodes are spaced evenly along the path.

GSM has the same algorithmic structure as the string method. The difference is that the string grows from two ends of the reaction path (the reactant and product)

toward the transition state along an interpolated pathway until the growing ends meet. However, the growth of the string depends on the interpolated pathway, which is determined by the update to all nodes in the previous iteration. The dependence on the previous iteration makes it very difficult to parallelize this method. Furthermore, the growing two ends will not meet unless the original interpolated pathway is a good guess for the MEP.

QSM uses the local quadratic approximation of the PES [13]. Compared with the string method, it is more accurate and it converges faster. QSM applies an adaptive step-size Runge–Kutta method and accordingly removes the need for the user to decide the step size [15, 16]. Formulated as a multi-objective optimization problem, it can be easily parallelized. QSM is considered one of the most efficient two end methods for large reaction systems.

Unfortunately, even the best two-end methods require that one have enough prior knowledge of the PES to guess an accurate initial path. Guessing an initial path is almost impossible when exploring new chemistry, in which case one could use the surface walking algorithms instead.

5.2.3

Surface Walking Algorithms

Surface walking algorithms usually start from a stationary point and search for energy minima and transition state (TS) by walking on the PES. Some popular surface walking methods are the eigenvector following (EF) method [17], the gradient extremal following (GEF) method [18], the reduced gradient following (RGF) method [19], the scaled hypersphere search (SHS) method [20–23] and the fast marching method (FMM) [24–31]. Since walking uphill is much more difficult than downhill, most surface walking algorithms focus on the uphill walking algorithm, and aim at global mapping of the PES. The fundamental problem with walking uphill is deciding which walking direction leads from the minimum to the TS.

The eigenvector following method can locate local minima and first-order saddle points by walking through the PES. Starting from an arbitrary point on the PES, the EF method locates the stationary points by walking along an eigenvector of the Hessian (second-derivative) matrix. By walking along all $2(3N-6)$ eigenvector directions, the EF method can potentially find all local minima and saddle points in an N -atom molecular system.

The gradient extremal following method walks uphill and downhill by following the extreme absolute values of the gradient along the potential contours [17]. Gradient extremals are curves that intersect the potential energy isosurfaces, $V(\mathbf{R}) = k$, where the curvature of these contour surfaces is an extremum. Since the curvature of an isosurface at a stationary point is infinity, the gradient extremal curves (e.g., the gradient maximum and gradient minimum) are supposed to cross at the stationary points. Consequently, finding the crossing points of gradient extremals will give the stationary points. One problem with the GEF method is that sometimes gradient extremals also intersect at points other than the stationary points [23].

The idea of the reduced gradient following method comes from the zero gradient criterion for stationary points [19]. Starting from a minimum, RGF finds the set of points whose potential gradients are all aligned to the direction of a chosen coordinate. RGF curves connect stationary points differing in their index by 1 and they intersect at the stationary points. The index of a stationary point is the number of negative eigenvalues of the Hessian matrix at this point [32]. Examples of searching for saddle points using the crossing points of RGF curves are shown in Reference [19]. RGF curves have been extended to a more general concept: the Newton trajectory (NT) [33, 34]. The searching direction of NT is not limited to one of the coordinates. It could be any direction. To avoid constructing trajectories that wander around the high energy regions of the PES, Quapp applied the growing string algorithm to find the NT [35–37]. A NT without a turning point can be used as approximation to the reaction path. Unfortunately, because there are infinitely many searching directions, it is sometimes difficult to locate a NT that approximates the reaction path.

The scaled hypersphere searching method [20, 21, 23] is based on the chemical intuition that energy-lowering interactions distort the potential surface downwards as one moves towards the TS [21]. SHS can walk towards the TS by following the extreme magnitude of anharmonicity from the second-order surface expanded at the starting minimum. The efficiency of SHS method is claimed to be $2(3N - 6)$ energy minimization calculations on each hypersphere, but expensive calculations of the Hessian matrix are required.

5.2.4

Metadynamics Methods

Other energy minima searching approaches, such as the free-energy minima escaping method proposed by Laio and Parrinello [38], and the conformation flooding approach by Grubmüller [39], are based on self-avoiding molecular dynamics trajectories on the potential energy surface. These trajectories do not pass precisely through the TS and reactive intermediate structures, so they do not provide a satisfactory representation for the reaction path. However, a reactive trajectory from these methods can be used as an initial guess for a two end method.

5.2.5

Fast Marching Method

FMM is a wavefront propagation method that solves the nonlinear eikonal equation [27–30]. FMM has been successfully applied to find the MEP on the PES. As previously mentioned, uphill walking on the PES is more troublesome than downhill walking. FMM avoids the uphill walking problem and transforms the multi-well PES into a single-well energy cost surface by solving the eikonal equation. The only well on the energy cost surface is the starting point, where the cost is defined to be zero. Then the MEP from any point on the PES to the starting point can be found by a downhill

backtracing from this point to the bottom of the energy cost surface. Unlike the two end methods, FMM does not need an initial guess of the path and it always converges to the MEP. If the ending point of the path is not specified, FMM will eventually evaluate the whole PES. Details of the FMM algorithm are presented in next section.

5.3

Fast Marching Method

5.3.1

Introduction to FMM

We define the cost function at $\mathbf{R}$ as the “minimum cost” required to attain this configuration starting from the reactant configuration $\mathbf{R}_0$ [26, 31]:

$$U(\mathbf{R}) = \min_{\mathbf{C}_{\mathbf{R}_0, \mathbf{R}}(s)} \int_0^L \left\{ \sqrt{2(E - V[\mathbf{C}(s)])} \right\}^n ds \quad (5.1)$$

Here the minimization is over all paths, $\mathbf{C}_{\mathbf{R}_0, \mathbf{R}}(s)$, that start at $\mathbf{R}_0$ and end at $\mathbf{R}$, E is the total energy of the system, $V(\mathbf{R})$ is the potential energy and L is the path length. [The variable s parameterizes the path so that $\mathbf{C}_{\mathbf{R}_0, \mathbf{R}}(0) = \mathbf{R}_0$ and $\mathbf{C}_{\mathbf{R}_0, \mathbf{R}}(L) = \mathbf{R}$.] The path integral problem (5.1) can be conveniently restated as an eikonal equation, namely:

$$|\nabla U(\mathbf{R})| = \left\{ \sqrt{2[E - V(\mathbf{R})]} \right\}^n \quad (5.2)$$

The energy cost of the reactant is zero by definition [$U(\mathbf{R}_0) = 0$]; this is the boundary condition for the eikonal equation.

This eikonal equation describes wavefront propagation with the local speed function:

$$\frac{1}{\left\{ \sqrt{2[E - V(\mathbf{R})]} \right\}^n}$$

To locate the MEP, we need the cost of molecular configurations with higher potential energy to be infinitely larger than the cost of configurations with lower potential energy. (Equivalently, we need for configurations that are lower in energy to be attained infinitely faster than configurations that are higher in energy.) This can be achieved by letting $n \rightarrow -\infty$, which ensures that higher energy paths in Equation 5.1 are cut off from the set of paths ($\mathbf{C}_{\mathbf{R}_0, \mathbf{R}}$), giving only the MEP. Of course, in computational implementations we will choose n to be a sizeable (but non-infinite) negative number. In practice, results are usually good when $n < -10$.

Solving this eikonal equation transforms a multi-well potential energy surface, $V(\mathbf{R})$, into a conical energy cost surface $U(\mathbf{R})$. The numerical algorithms for solving the eikonal equation are discussed in the following sections.

5.3.2

Upwind Difference Approximation

We need to solve the eikonal equation using an “upwind” finite difference approximation that preserves the causality of the solutions. To do this, we discretize the eikonal equation as follows:

$$\left(\frac{(U-a_1)^+}{dR(1)}\right)^2 + \left(\frac{(U-a_2)^+}{dR(2)}\right)^2 + \cdots + \left(\frac{(U-a_d)^+}{dR(d)}\right)^2 = [2(E-V(\mathbf{R}))]^n \quad (5.3)$$

Here $dR(i)$ is the i -th component of the grid size vector $d\mathbf{R}$; a_i is the smaller cost value of point $\mathbf{R}$'s two neighboring points in direction i , $a_i = \min(U_{\text{left}}, U_{\text{right}})$. The upwind finite difference approximation defines $(U-a_i)^+ = (U-a_i)$ if $U > a_i$ and $(U-a_i)^+ = 0$ otherwise [i.e., $(U-a_i)^+ = \max(0, U-a_i)$].

The upwind finite difference enforces the causality condition in the fast marching method, which means the cost can only increase while the wavefront moves outward. In other words, for the point in question, its unknown cost value U has to be greater than the cost value, a_i , of its known neighboring point; if $a_i > U$, then the cost value of this neighboring point must not be known either. We cannot use an unknown point, so we discard it by letting $(U-a_i)^+ = 0$. This is the idea behind the upwind finite difference approximation.

Equation 5.3 can be solved in an iterative way. First, sort the a_i 's in increasing order. Second, start from $j = 1$ and solve the truncated equation:

$$\left[\frac{(U-a_1)^+}{dR(1)}\right]^2 = \{2[E-V(\mathbf{R})]\}^n \quad (5.4)$$

If the solution $U_1 \leq a_2$, then $U_1 \leq a_3 \leq \cdots \leq a_d$ and thus $U = U_1$ is also the solution to Equation (5.3). If $U_1 > a_2$, then let $j = j + 1$, and continue to solve the truncated equation with two terms on the left-hand side. This process is repeated until we find the j -th solution $U_j \leq a_{j+1}$, $1 \leq j \leq d$. $U = U_j$ is the solution to Equation 5.3 [40].

5.3.3

Heapsort Technique

As the wavefront propagates outward, energy cost values of grids on the wavefront are computed by solving the discretized eikonal equation. After computing the energy cost values of all points on the wavefront, we need to identify the point with minimum energy cost value because this is the point that the wavefront is going to pass next. The heapsort technique is used to sort these values.

Heapsort is an “in-place” sorting algorithm, requiring no auxiliary storage [41]. It has a runtime of $O(N \log_2 N)$ for the worst case, where N is the number of data. A “sift-up” process is applied to arrange the input data into a binary heap. The sift-up

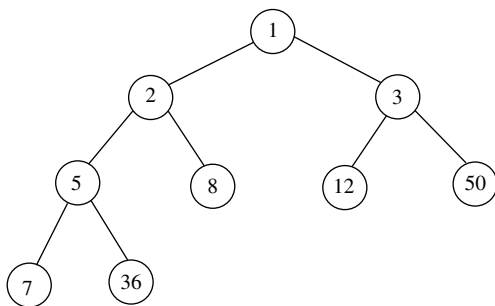


Figure 5.1 Binary Min-heap.

process is analogous to corporate promotion. It can be described as the following two parts.

“Add to heap” process: We can imagine the first data added to the heap as the first employee. Once we “hire” another one, he will temporarily be the subordinate to the first.

“Update heap” process: We compare the newly-hired employee with his supervisor, if he is more capable, swap their positions, and repeat this comparison until we reach the top of the heap; if not, he stays put. This “update heap” process ensures that the most capable employee always stays at the top and that each upper level employee is always more capable than his subordinates. If the capability of the employees is evaluated by numbers, the sift-up process gives us a min-heap such as that in Figure 5.1.

5.3.4

Shepard Interpolation

The computational cost of the FMM is dominated by the potential energy calculation. One Gaussian calculation for a five-atom reaction system takes about 3 minutes using B3LYP/6-311 + + G**. At a reasonable grid size, a two-dimensional PES consists of thousands of points, so it might take several weeks to compute the entire potential energy surface. FMM does not need the entire potential energy surface, but only a narrow band along the reaction path. This saves up to 70% of Gaussian calculations for two-dimensional PES and even more for higher dimensional PES.

The number of Gaussian calculations can be reduced even further by building the PES using Shepard interpolation. Based on N accurately calculated points (we call them “reference” points), we can approximate the potential energy at another point, $\mathbf{R}$, using the Taylor series [24, 42]:

$$(T^{(i)}(\mathbf{R}) = V(\mathbf{R}^{(i)}) + (\mathbf{R} - \mathbf{R}^{(i)}) \cdot \nabla V(\mathbf{R}^{(i)}) + \frac{1}{2}(\mathbf{R} - \mathbf{R}^{(i)}) \cdot \nabla \nabla V(\mathbf{R}^{(i)}) (\mathbf{R} - \mathbf{R}^{(i)}) + \cdots)_{i=1}^N. \quad (5.5)$$

Owing to different distances of the reference points from $\mathbf{R}$, the Taylor series from each of these points makes a different contribution to $V(\mathbf{R})$. If we model their

contribution using a weight function, $\omega^{(i)}(\mathbf{R})$, then the interpolated potential for point $\mathbf{R}$ is:

$$\tilde{V}(\mathbf{R}) = \sum_{i=1}^N \omega^{(i)}(\mathbf{R}) T^{(i)}(\mathbf{R}) \quad (5.6)$$

where $T^{(i)}(\mathbf{R})$ is the Taylor series expansion given in Equation (5.5), and the weight function $\omega^{(i)}(\mathbf{R})$ is a non-negative, normalized function. Normalization can be enforced by:

$$\omega^{(i)}(\mathbf{R}) = \frac{v^{(i)}(\mathbf{R})}{\sum_{j=1}^N v^{(j)}(\mathbf{R})} \quad (5.7)$$

It is well known that the asymptotic form of $v^{(i)}(\mathbf{R})$ should be $\|\mathbf{R} - \mathbf{R}^{(i)}\|^{-(n+1)}$ if $T^{(i)}(\mathbf{R})$ is truncated at the n -th order term. We use the following form:

$$v^{(i)}(\mathbf{R}) = \frac{e^{-\frac{1}{2} \sum_{k=1}^d \left(\frac{R_k - R_k^{(i)}}{\sigma_k^{(i)}} \right)^2}}{\sum_{k=1}^d \left(\frac{R_k - R_k^{(i)}}{\sigma_k^{(i)}} \right)^{n+1}} \quad (5.8)$$

where $\sigma_k^{(i)}$ is the trust radius of reference point i in the k -th dimension. Rather than using the Bettens–Collins isotropic formula [43] to calculate the trust radius:

$$\sigma^{(i)} = \left[\frac{1}{M} \sum_{j=1}^M \frac{[V(\mathbf{R}^{(j)}) - T(\mathbf{R}^{(i)})]^2}{(\epsilon_v)^2 \|\mathbf{R}^{(j)} - \mathbf{R}^{(i)}\|^{2n+2}} \right]^{-\frac{1}{2n+2}} \quad (5.9)$$

we use the direction dependent formula of the form:

$$\sigma_k^{(i)} = \left[\frac{1}{M} \sum_{j=1}^M \frac{\left[\left(\frac{\partial V(\mathbf{R}^{(j)})}{\partial R_k} - \frac{\partial T(\mathbf{R}^{(i)})}{\partial R_k} \right) (R_k^{(j)} - R_k^{(i)}) \right]^2}{(\epsilon_v)^2 \|\mathbf{R}^{(j)} - \mathbf{R}^{(i)}\|^{2n+2}} \right]^{-\frac{1}{2n+2}} \quad (5.10)$$

Given the interpolated potential value, $\tilde{V}(\mathbf{R})$, the error of the Shepard interpolant is estimated by:

$$\text{error} = \sqrt{\sum_{i=1}^N \omega^{(i)}(\mathbf{R}) [\tilde{V}(\mathbf{R}) - T^{(i)}(\mathbf{R})]^2} \quad (5.11)$$

If the estimated error given by Equation 5.11 is less than the error threshold, we accept $\tilde{V}(\mathbf{R})$ as the potential for point $\mathbf{R}$. If the error is too large, then we do not use the Shepard interpolant $\tilde{V}(\mathbf{R})$; instead we calculate the PES at this point using Gaussian.

If we truncate the Taylor series at higher order terms, we expect the accuracy of Shepard interpolation to be improved. Unfortunately, computing the higher-order

derivatives is very expensive. Instead we use the interpolating moving least-squares method and the potential and gradient values from Gaussian calculation to fit the higher-order derivatives.

5.3.5

Interpolating Moving Least-Squares Method

For the interpolated moving least-squares method, the basic equation we need to solve is [24, 42]:

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\| \quad (5.12)$$

where $\mathbf{x}$ is the vector of higher order derivatives at the point $\mathbf{X}^{(j)}$. For these equations we assume that the energy and the first-order derivatives are available at all calculated points. If we denote the set of M neighbor points for the j -th point as $\mathbf{Q}^{(j)}$ then we can write the vector of known data $\mathbf{b}$ in the form:

$$\begin{aligned} b_1 &= V(\mathbf{X}^{(j)}) - V(\mathbf{X}^{[\mathbf{Q}_1^{(j)}]}) - (\mathbf{X}^{[\mathbf{Q}_1^{(j)}]} - \mathbf{X}^{(j)}) \cdot \nabla V(\mathbf{X}^{[\mathbf{Q}_1^{(j)}]}) \\ &\vdots \\ b_M &= V(\mathbf{X}^{(j)}) - V(\mathbf{X}^{[\mathbf{Q}_M^{(j)}]}) - (\mathbf{X}^{[\mathbf{Q}_M^{(j)}]} - \mathbf{X}^{(j)}) \cdot \nabla V(\mathbf{X}^{[\mathbf{Q}_M^{(j)}]}) \\ b_{M+1} &= \frac{\partial V(\mathbf{X}_j)}{\partial \mathbf{X}_{j,1}} - \frac{\partial V(\mathbf{X}_{\mathbf{Q}_1^{(j)}})}{\partial \mathbf{X}_{\mathbf{Q}_1^{(j)},1}} \\ &\vdots \\ b_{(d+1)M} &= \frac{\partial V(\mathbf{X}_j)}{\partial \mathbf{X}_{j,d}} - \frac{\partial V(\mathbf{X}_{\mathbf{Q}_M^{(j)}})}{\partial \mathbf{X}_{\mathbf{Q}_M^{(j)},d}} \end{aligned} \quad (5.13)$$

which has $(d+1)M$ elements. The unknown vector $\mathbf{x}$ contains the derivatives of the potential with the redundant elements removed:

$$\mathbf{x} = \begin{bmatrix} \frac{1}{2} \frac{\partial^2 V(\mathbf{X}^{(j)})}{\partial \mathbf{X}_1^{(j)} \partial \mathbf{X}_1^{(j)}} \\ \frac{\partial^2 V(\mathbf{X}^{(j)})}{\partial \mathbf{X}_1^{(j)} \partial \mathbf{X}_2^{(j)}} \\ \vdots \\ \frac{1}{2} \frac{\partial^2 V(\mathbf{X}^{(j)})}{\partial \mathbf{X}_d^{(j)} \partial \mathbf{X}_d^{(j)}} \\ \frac{1}{6} \frac{\partial^3 V(\mathbf{X}^{(j)})}{\partial \mathbf{X}_1^{(j)} \partial \mathbf{X}_1^{(j)} \partial \mathbf{X}_1^{(j)}} \\ \frac{1}{2} \frac{\partial^3 V(\mathbf{X}^{(j)})}{\partial \mathbf{X}_1^{(j)} \partial \mathbf{X}_1^{(j)} \partial \mathbf{X}_2^{(j)}} \\ \vdots \end{bmatrix} \quad (5.14)$$

which has $\frac{1}{n!} \prod_{i=0}^{n-1} d + i$ elements. The matrix **A** takes the form:

$$\mathbf{A} = \begin{pmatrix} \left(X_1^{(j)} - X_1^{Q_1(j)}\right)^2 & \cdots & \left(X_d^{(j)} - X_d^{Q_1(j)}\right)^n \\ \vdots & \ddots & \vdots \\ \left(X_1^{(j)} - X_1^{Q_M(j)}\right)^2 & \cdots & \left(X_d^{(j)} - X_d^{Q_M(j)}\right)^n \\ \frac{1}{2} \left(X_1^{(j)} - X_1^{Q_1(j)}\right) & \cdots & \frac{1}{n} \left(X_1^{(j)} - X_1^{Q_1(j)}\right)^{n-1} \\ \vdots & \ddots & \vdots \\ \frac{1}{2} \left(X_1^{(j)} - X_1^{Q_M(j)}\right) & \cdots & \frac{1}{n} \left(X_d^{(j)} - X_d^{Q_M(j)}\right)^{n-1} \end{pmatrix} \quad (5.15)$$

At order n each point contributes $d + 1$ elements to Equation 5.13 – so the number of points available to solve for **x** needs to be at least:

$$\frac{1}{(d+1)n!} \prod_{i=0}^{n-1} (d+i)$$

5.3.6

FMM Program

To apply FMM to real chemical systems, we need to interface FMM with a quantum chemistry package to compute the potential energy. In this section, we discuss how FMM is interfaced to the Gaussian quantum chemistry program.

5.3.6.1 Setup, Definitions and Notation

- 1) **Define the grid space:** Given a chemical reaction, the first step is to determine the dimensionality of the PES that will be used in Equation (5.3) of the FMM program. To minimize the computational cost, we use a reduced PES by choosing a few key coordinates that are essential for describing the reaction coordinate. The dimension of reduced PES is the number of key coordinates. We denote it as d . We also need to decide the minimum and maximum values of all key coordinates, so that we can limit our calculation to the region of the PES that we are interested in.
- 2) **Categorize the grid points:** The wavefront starts from a point (usually the reactant) and propagates outward. We need to categorize the grid points inside (evaluated) and outside (unevaluated) of the wavefront and points on the wavefront (being evaluated).
 - “Alive” points: points inside the wavefront. The energy cost values of the “alive” points have been evaluated and will no longer change.
 - “Near” points: points on the wavefront. These points are under evaluation and their energy cost values are temporary. The energy cost of these points will be updated whenever the cost of one of their neighboring points changes.

- “Far” points: points outside of the wavefront. These points will not be evaluated until the wavefront moves close. The energy cost values of all “far” points are assigned as infinity.

5.3.6.2 Initialize the Calculation

- 1) Tag all points as “far”, and set their energy cost values as infinity.
- 2) Call Gaussian to compute the potential energy and gradient of the starting point. Set the energy cost of the starting point to zero and tag it as “alive.”
- 3) Tag the $2d$ neighboring points of this first alive point as “near” and add them to the heap. Call Gaussian to compute the potential energy and potential energy gradient of each “near” point, and calculate the energy cost by solving the discretized eikonal equation, Equation (5.3). Update the heap according to the updated energy cost values, so that the point with minimum energy cost value is at the top of the heap.
- 4) Initialize the Shepard interpolation. We call these Gaussian points “reference points” because they will be used to approximate the potential values of nearby points. For each reference point, we need a neighbor list. This neighbor list contains M points that are used to determine the trust radius of Shepard interpolation weights and to calculate higher-order derivatives by using interpolated moving least squares. Once we have a new Gaussian point, we compare its distance to the existing reference points. If it lies within an acceptable distance of a reference point, then we add it to the neighbor list of this reference point.

5.3.6.3 Updating the Heap

- 1) Tag the top point of the heap as “alive”, and tag its “far” neighboring point(s) as “near.” Add them to the heap.
- 2) For each of these new “near” points, call Shepard interpolation to approximate the potential energy. If the estimated error is acceptable, then use the potential energy from the Shepard interpolant. If the estimated error is over the error threshold, call Gaussian to compute the potential energy and gradient. Use the potential energy to compute the energy cost and then update the heap.
- 3) Repeat the above steps 1 and 2 until the product is found or another stopping criterion is met.

5.3.6.4 Backtracing from the Ending Point to the Starting Point on the Energy Cost Surface

Owing to the causality condition of the eikonal equation, as the wavefront moves outwards the energy cost will always increase, which ensures that the energy cost surface is a one-well conical surface. The starting point is at the bottom. So a simple steepest descent path from the ending point to the starting point on the energy cost surface will give the MEP. In our program this is done with Euler integration [17, 26]:

$$R_{k+1} = R_k - h \frac{\nabla U(R_k)}{\|\nabla U(R_k)\|} \quad (5.16)$$

where, for simplicity, we use a fixed step size $h = \|d\mathbf{R}\|/20$. To compute the energy cost and gradient at point $\mathbf{R}$ we can use its 2^d nearest neighbor grid points to form a linear set of equations:

$$U(\mathbf{R}_q) = b + \sum_{i=1}^d a_i(\mathbf{R}_q - \mathbf{R}) \quad (5.17)$$

where $\nabla U(\mathbf{R}) = \mathbf{a}$, $U(\mathbf{R}) = b$, $q = 1 \dots 2^d$ and $\mathbf{R}_q$ are the coordinates of the nearest neighbor grid points. Since the energy cost at the neighbor grid points $U(\mathbf{R}_q)$ are known, the energy cost and gradient values at point $\mathbf{R}$ can be fitted by solving this linear set of equations.

5.3.7

Application

5.3.7.1 Four-Well Analytical PES

The four-well PES is defined by the following analytical function [26]:

$$V(R_1, R_2) = V_0 + a_0 e^{-(R_1 - b_1)^2 - (R_2 - b_2)^2} - \sum_{i=1}^4 a_i e^{-p_i(R_1 - \alpha_i)^2 - q_i(R_2 - \beta_i)^2} \quad (5.18)$$

where all parameters are listed in Table 5.1.

The four-well PES is a standard test system for the FMM. There are four minima on this PES, and four transition states between each pair of minima. If we choose the minimum at the bottom right as the reactant and the one at the top right as the product, then there are two possible pathways: (i) the “direct” one-step pathway and (ii) the “C-shaped” three-step pathway passing through by two intermediates (I and II) and three transition states (TS1, TS2 and TS3) (Figure 5.2). Starting from the reactant, we imagine the FMM procedure as slowly adding water to the reactant valley [31]; the “water” level can be considered as the propagating wavefront. The water level will keep going up, wetting the contours of the potential energy surface as it does so. Eventually the water level will rise to the level of the lowest-energy TS, which is the lowest “mountain pass” for exiting the reactant valley. At this stage a

Table 5.1 Parameters for the four-well analytical PES.

Parameter	Value	Parameter	Value	Parameter	Value
V_0	5.0 kcal mol ^{-1a)}	p_1	0.3 Å ⁻²	α_1	1.3 Å
a_0	0.6 kcal mol ⁻¹	p_2	1.0 Å ⁻²	α_2	-1.5 Å
a_1	3.0 kcal mol ⁻¹	p_3	0.4 Å ⁻²	α_3	1.4 Å
a_2	1.5 kcal mol ⁻¹	p_4	1.0 Å ⁻²	α_4	-1.3 Å
a_3	3.2 kcal mol ⁻¹	q_1	0.4 Å ⁻²	β_1	-1.6 Å
a_4	2.0 kcal mol ⁻¹	q_2	1.0 Å ⁻²	β_2	-1.7 Å
b_1	0.1 Å	q_3	1.0 Å ⁻²	β_3	1.8 Å
b_2	0.1 Å	q_4	0.1 Å ⁻²	β_4	1.23 Å

a) 1 kcal = 4.184 kJ.

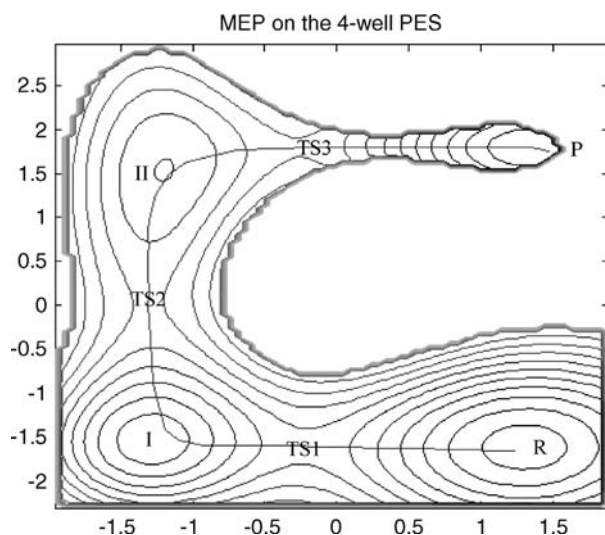


Figure 5.2 MEP on the four-well PES. The grid sizes on both dimensions are $dR = 0.05$.

narrow thread of water will follow the steepest-descent path to the bottom of the next valley. The water keeps flooding mountain valleys in this way until the product is found. In FMM, the “energy cost” contours record which portions of the PES are “flooded” at any given point in time (Figure 5.3). Notice that only the “flooded” portion of the surface needs to be computed. This reduces the computational cost significantly.

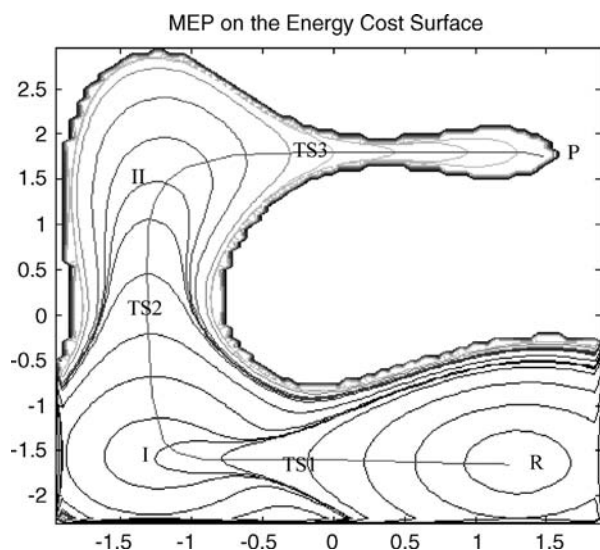
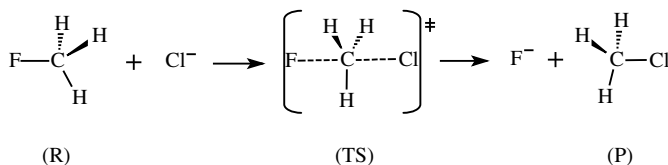


Figure 5.3 MEP on the energy cost surface transformed from the four-well PES by solving the eikonal equation. The MEP is determined by backtracing from the product to the reactant along the steepest descent path on the energy cost surface.

The lower energy region of the PES in Figure 5.2 is transformed into an energy cost surface in Figure 5.3, and the higher energy part is cut off. Backtracing from the product to the reactant along the steepest descent path gives the MEP.

5.3.7.2 S_N2 Reaction [31]

The mechanism of the S_N2 reaction (e.g., Scheme 5.1) has been studied intensively by experimental and theoretical methods, so it is a good test for FMM. This is a one-step reaction, so we expect two minima (the reactant R and product P) and one TS on the PES.



Scheme 5.1

In this reaction, only C–F and C–Cl bonds are involved in bond-breaking and bond-forming, so the PES of this reaction can be modeled using a two-dimensional reduced PES based on the C–F and C–Cl coordinates. At each grid point, we will freeze the C–F and C–Cl bond lengths at the given values and minimize the energy with respect to the other coordinates. Figure 5.4 depicts the two-dimensional reduced PES and the MEP computed by the FMM program. About 20% of grid points are in the “flooded” region and are computed by Gaussian 03. Figure 5.5 shows the energy-

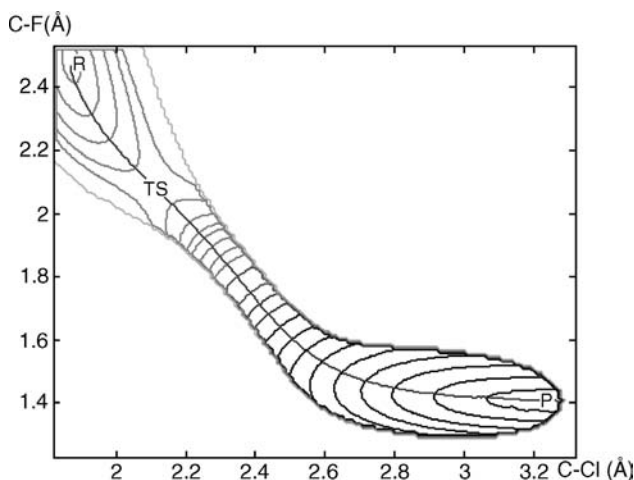


Figure 5.4 PES of the S_N2 reaction based on C–F and C–Cl bond lengths. The grid sizes on both dimensions are $dR = 0.01 \text{ \AA}$. The calculation starts from the reactant (R), fills the reactant valley, breaches the reaction barrier at the transition state (TS) and then “flows” down to the product (P). The FMM program

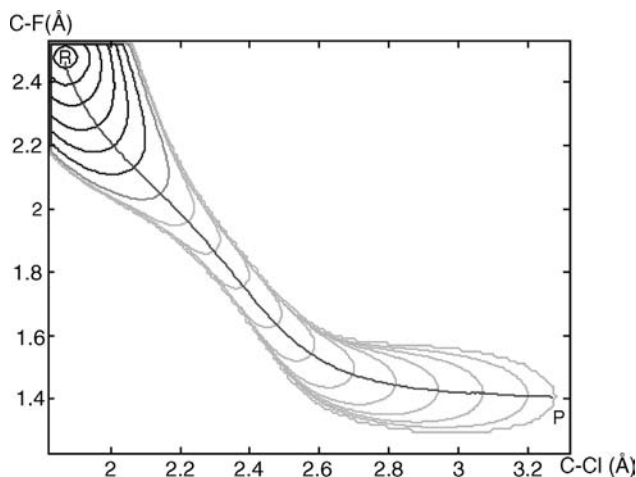


Figure 5.5 Energy-cost surface transformed from the PES in Figure 5.4. The MEP is determined by backtracing from the product to the reactant along the steepest descent path on the energy-cost surface.

cost surface with the reactant (R) as starting point and the MEP found on this surface. Plotting the change in potential energy along the MEP gives the energy profile of the reaction coordinate (Figure 5.6).

5.3.7.3 Dissociation of Ionized *O*-Methylhydroxylamine [31]

The PES of $[\text{CH}_5\text{NO}]^+$ has been studied using mass spectroscopy and computational methods [44]. The following dissociation reaction has been observed:

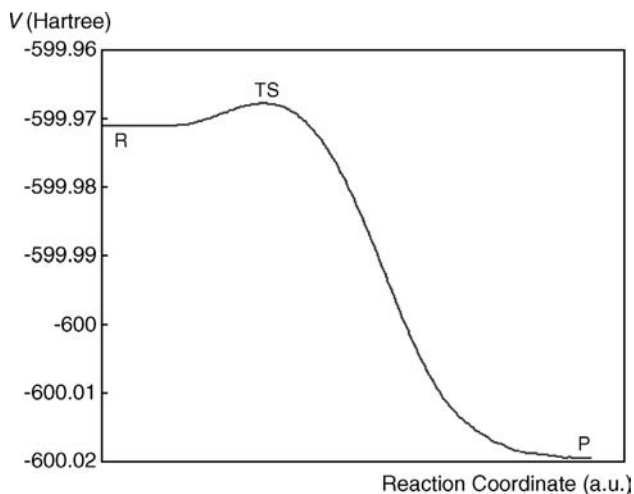
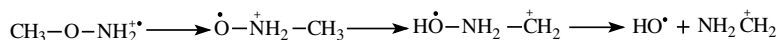


Figure 5.6 Energy profile of the $\text{S}_{\text{N}}2$ reaction.



Scheme 5.2



Terlouw and coworkers have proposed the mechanism shown in Scheme 5.2 for this dissociation reaction [44].

Using bond lengths C–N, N–O and O–H as key coordinates, FMM finds a reduced three-dimensional PES. The three-dimensional equipotential surfaces have onion-like structures. Each layer of the “onion” represents a certain value of the potential energy. Figure 5.7 shows one layer of the “onion” with a potential value of -170.534 hartree. The cores of the onions represent minima on the PES (Figure 5.8). We can see that there are four minima on this PES, the reactant (R), two intermediates (I, II) and the product (P). The coordinates of the minima show that the structures of intermediate (I) and (II) coincide with $\dot{\text{O}}\text{--NH}_2\text{--CH}_3$ and $\text{H}\dot{\text{O}}\text{--NH}_2\text{--C}^+\text{H}_2$, respectively. The FMM calculation confirms that the mechanism in Scheme 5.2 is the minimum energy reaction pathway.

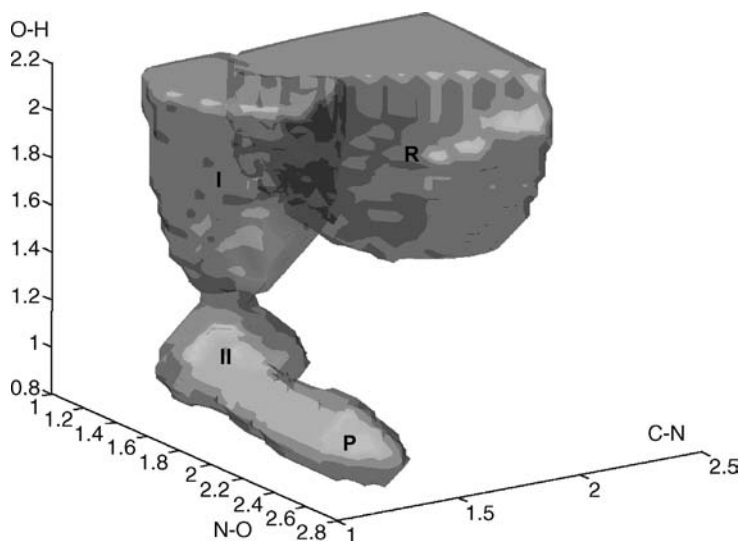


Figure 5.7 Isosurface with a potential value of -170.534 hartree, which is one layer of the reduced three-dimensional PES for the dissociation reaction of ionized O-methylhydroxylamine. The three-dimensional

equipotential surfaces have an onion-like structure. Each layer of the “onion” represents a certain value of the potential energy. The cores of the “onions” are minima on the PES.

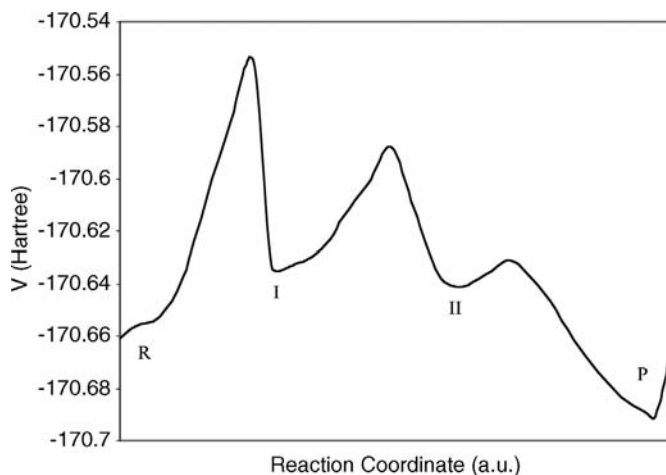


Figure 5.8 Energy profile of the dissociation reaction of ionized *O*-methylhydroxylamine.

5.4

Quantum Mechanics/Molecular Mechanics (QM/MM) Methods Applied to Enzyme-Catalyzed Reactions

5.4.1

QM/MM Methods

Enzyme-catalyzed reactions are of great importance in the biological sciences and pharmaceutical industry because of their efficiency and specificity. Using computational tools to study the mechanism of enzyme-catalyzed reactions is one of our ultimate goals. Even with the advances of modern computers and new computational methods, studying the mechanism of enzyme-catalyzed reactions is still a great challenge due to the large size of the enzyme system. QM methods are accurate but expensive, and so are generally limited to systems of less than 100 atoms. For enzyme-catalyzed reactions that involve thousands of atoms, it is impossible to apply QM methods to the entire system. To deal with larger systems molecular mechanics is commonly employed. The accuracy of MM methods can be poor and it is unsuitable for studying bond-breaking and bond-forming processes in chemical reactions.

In a typical enzyme-catalyzed reaction, only a small number of atoms are involved directly in the bond-breaking and bond-forming processes; the primary role of the other atoms is to provide a favorable steric and electrostatic environment. This realization led Warshel and Levitt to propose the hybrid QM/MM approach [45, 46]. In QM/MM the enzyme reaction system is divided into two parts: the atoms that are directly involved in the reaction are evaluated quantum mechanically, while the rest of the atoms are treated with MM methods. This approach combines the advantages of the high accuracy of QM methods for the small QM subsystem with the computational affordability of MM methods for the remainder of the molecules (Figure 5.9).

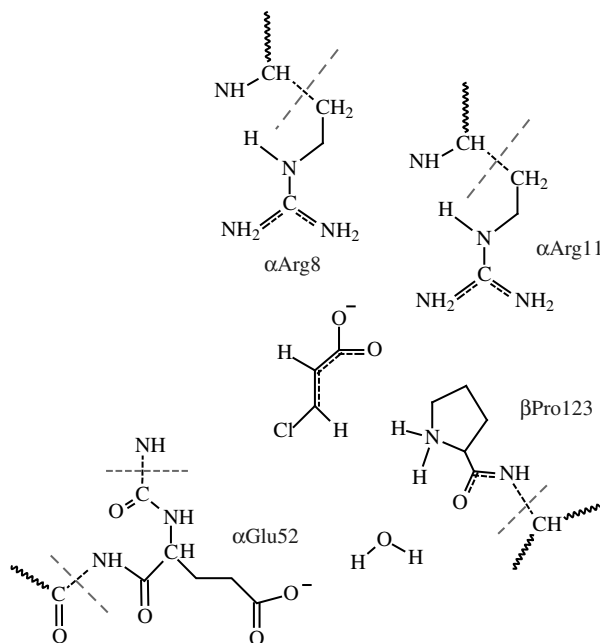


Figure 5.9 QM subsystem (the substrates, part of residue α -arginine8, α -arginine11, α -glutamine52 and β -proline123) and MM subsystem (the rest of the system) of the dechlorination of *trans*-3-chloroacrylate catalyzed by *trans*-3-chloroacrylic acid dehalogenase (CAAD).

After three decades of development, QM/MM methods have been successfully applied in simulations of various enzyme-catalyzed reactions [47–57].

One important problem associated with QM/MM methods is how to deal with the QM and MM covalent boundary. The link atom approach is one of the most commonly used methods [58–61]. In the link atom approach, link atoms like hydrogen or pseudohalogen atoms are inserted to cover the free valence of the QM subsystem so that the QM subsystem will still be a closed-shell system. The problem with the link atom approach is that it introduces additional degrees of freedom and some double counting of the interactions into the system, which can be difficult to correct for. Owing to the deficiency of the link atom approach, in the following discussion we focus on the pseudobond QM/MM method developed by Yang's group [54, 55, 57]. The pseudobond approach does not introduce additional atoms to the system. Instead this approach replaces the MM boundary atom with a seven-valence-electron atom with an effective core potential and forms a pseudobond between this atom and the QM boundary atom [58, 59]. The pseudobond approach gives a smooth interface between the QM and MM subsystems and provides a consistent and well-defined *ab initio* QM/MM potential energy surface.

QM/MM methods can be categorized into two types: semiempirical QM/MM methods and *ab initio* QM/MM methods, depending on the level of QM theory used.

Semiempirical QM/MM methods are much faster computationally so that classical statistical sampling can be applied. However, semiempirical QM/MM methods are often not sufficiently accurate to give reliable free energies [62]. *Ab initio* QM/MM methods are accurate but expensive, so reaction path ensemble sampling is not feasible. The QM/MM free-energy perturbation (QM/MM-FEP) method developed by Yang's group utilizes the pseudobond approach to form a smooth interface between the QM and MM subsystem, then applies an efficient, iterative optimization procedure [56, 63] to optimize the QM and MM subsystem of a given conformation independently and iteratively until convergence. Incorporated with a reaction path optimization method [3, 8, 13, 63, 64], the reaction path can be found on the PES. The last step is to perform free-energy perturbation calculations on the reaction path to give the free-energy profile of the reaction. The problem with the QM/MM-FEP method is that the optimization of the reaction path depends on the PES of a single MM conformation [51]. To eliminate this dependence one can instead perform a direct path optimization on the free energy surface.

The most recent QM/MM minimum free-energy path (QM/MM-MFEP) method [51, 53] is one of the more efficient and reliable *ab initio* QM/MM methods. Unlike other *ab initio* QM/MM methods, the free energy profile obtained in the QM/MM-MFEP method is not built from a previously sampled PES of a random chosen initial conformation of the system, instead it is generated naturally because the reaction path is optimized on the potential of mean force (PMF) surface, which is the free energy expression of the QM subsystem with the MM contributions averaged out. Thus, the problem of finding the reaction path in a complicated phase space with the same number of degrees of freedom of the entire QM/MM system is simplified to a problem of exploring the PMF surface depending on just the QM degrees of freedom [51].

5.4.2

Incorporating the QM/MM-MFEP Methods with FMM

QM/MM-MFEP methods [51, 53] have been incorporated with several path optimization methods, such as NEB [3], the Ayala–Schlegel second-order MEP method [64] and QSM [13]. All these methods aim to find the local MEP of the enzyme reaction. To ensure the convergence to the global MEP, we can implement the QM/MM-MFEP methods with FMM.

To carry out reaction path optimization on the PMF surface, the relative free energies between adjacent QM conformations and free-energy gradients for each individual QM conformation need to be computed [51]. The relative free energies between adjacent QM conformations are computed by the QM/MM-FEP method. The free energy difference is defined as [53]:

$$\Delta A = A^{(n)}(r_{\text{QM}}) - A_{\text{ref}} = -\frac{1}{\beta} \ln \left[\frac{1}{N} \sum_{\tau=1}^N \exp \{ -\beta [E(r_{\text{QM}}, \mathbf{r}_{\text{MM}}^{(n)}(\tau)) - E_{\text{ref}}(\mathbf{r}_{\text{MM}}^{(n)}(\tau))] \} \right] \quad (5.19)$$

where an MD simulation is performed on the MM subsystem with the QM conformation frozen. Then FMM is performed within a trust radius using the same MD ensemble. Outside of this trust radius the FMM algorithm can continue only when a new MD simulation is performed with a new QM conformation.

The free-energy gradients of the QM subsystem are computed through molecular dynamics sampling of the MM environment. The free-energy gradient associated with Equation 5.19 is computed as [53]:

$$\frac{\partial A(\mathbf{r}_{\text{QM}})}{\partial \mathbf{r}_{\text{QM}}} = \frac{\frac{1}{N} \sum_{\tau=1}^N \left(\frac{\partial E(\mathbf{r}_{\text{QM}}, \mathbf{r}_{\text{MM}})}{\partial \mathbf{r}_{\text{QM}}} \exp \left\{ -\beta [E(\mathbf{r}_{\text{QM}}, \mathbf{r}_{\text{MM}}^{(n)}(\tau)) - E_{\text{ref}}(\mathbf{r}_{\text{MM}}^{(n)}(\tau))] \right\} \right)}{\frac{1}{N} \sum_{\tau=1}^N \exp \left\{ -\beta [E(\mathbf{r}_{\text{QM}}, \mathbf{r}_{\text{MM}}^{(n)}(\tau)) - E_{\text{ref}}(\mathbf{r}_{\text{MM}}^{(n)}(\tau))] \right\}} \quad (5.20)$$

The rest of the FMM algorithm is the same as in Section 5.3.

5.4.3

Application of the Incorporated FMM and QM/MM-MFEP Method to Enzyme-Catalyzed Reactions

Using FMM as the path optimization algorithm, the QM/MM-MFEP method can be applied to find the global MEP for solution-phase reactions and enzyme-catalyzed reactions. Below we present a representative sample of the applications we are currently pursuing using this new methodology.

5.4.3.1 S_N2 Reaction in Solvent

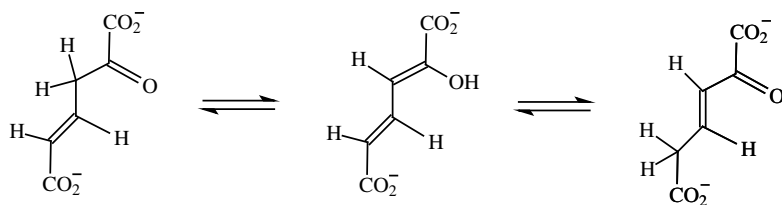
The solvent S_N2 reaction is a good test for the incorporated FMM and QM/MM-MFEP method. This reaction has been studied intensively by experimental and theoretical methods, so there is plenty of data to compare with.

Because of the rapid exchange of solvent molecules, QM/MM methods that depend on the initial conformation of the system cannot give reliable results because the initial conformation does not reflect the rapid change of solvent. Since the QM/MM-MFEP method eliminates this dependence, we expect better results for this solvent reaction.

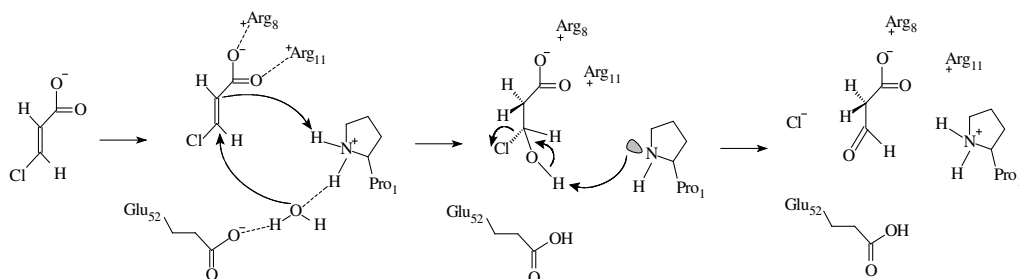
5.4.3.2 Isomerization Reaction Catalyzed by 4-Oxalocrotonate

Tautomerase (4-OT) [48, 65]

The mechanism of this reaction (Scheme 5.3) has been studied using the QM/MM-MFEP method. The incorporated FMM and QM/MM-MFEP method can confirm whether the reaction path is a global MEP.



Scheme 5.3



Scheme 5.4

5.4.3.3 Dechlorination Reaction Catalyzed by *trans*-3-Chloroacrylic Acid Dehalogenase (CAAD)

3-Chloroacrylic acid is an unnatural substance degraded from the active ingredient of the nematocides Shell D-D and Telone II. Its uncatalyzed half-life is about 10 000 years [66]. When catalyzed by CAAD, this hydrolytic dechlorination reaction proceeds with a rate enhancement of 2×10^{12} . The X-ray structure of *trans*-3-chloroacrylic acid dehalogenase gives some hint of the mechanism of the dechlorination reaction of *trans*-3-chloroacrylic acid (Scheme 5.4) [67]. We are planning to apply the incorporated FMM and QM/MM-MFEP method to study the mechanism of this reaction.

5.5

Summary

In this chapter we have reviewed briefly some numerical methods that locate the MEP on the PES or free energy surface. We focused on the FMM, which is one of the most general and reliable methods for finding the chemical reaction path. Unlike most competing methods, FMM always finds the global MEP. Some proof-of-principle examples of applying FMM to small gas-phase reactions were shown in Section 5.3. Most reactions are more complicated than this. Our ultimate goal is to study the mechanism of more realistic systems such as solution-phase or enzyme-catalyzed

reactions. To deal with the effects of the complicated molecular environment, QM/MM methods were introduced. A brief history of the development of QM/MM methods was given in Section 5.4, followed by the key ideas required to merge FMM with the recent QM/MM potential of mean force-based free energy path finding methods. The combination of QM/MM methods with FMM is a promising approach for determining chemical reaction mechanisms in complex reaction systems.

References

- 1 Fukui, K. (1981) The path of chemical-reactions – the IRC approach. *Acc. Chem. Res.*, **14** (12), 363–368.
- 2 Koslover, E.F. and Wales, D.J. (2007) Comparison of double-ended transition state search methods. *J. Chem. Phys.*, **127** (13), 134102.
- 3 Jonsson, H., Mills, G., and Jacobsen, K.W. (1998) Nudged elastic band method for finding minimum energy paths of transitions, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, World Scientific, Singapore, pp. 385–404.
- 4 Alfonso, D.R. and Jordan, K.D. (2003) A flexible nudged elastic band program for optimization of minimum energy pathways using ab initio electronic structure methods. *J. Comput. Chem.*, **24** (8), 990–996.
- 5 Chu, J.-W., Trout, B.L., and Brooks, B.R. (2003) A super-linear minimization scheme for the nudged elastic band method, in *A Super-Linear Minimization Scheme for the Nudged Elastic Band Method*, 119 (24), 12708–12717.
- 6 Henkelman, G. and Jonsson, H. (2000) Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.*, **113** (22), 9978–9985.
- 7 Trygubenko, S.A. and Wales, D.J. (2004) A doubly nudged elastic band method for finding transition states. *J. Chem. Phys.*, **120** (5), 2082–2094.
- 8 Xie, L., Liu, H.Y., and Yang, W.T. (2004) Adapting the nudged elastic band method for determining minimum-energy paths of chemical reactions in enzymes. *J. Chem. Phys.*, **120** (17), 8039–8052.
- 9 Weinan, E., Ren, W.Q., and Vanden-Eijnden, E. (2002) String method for the study of rare events. *Phys. Rev. B*, **66** (5), 052301.
- 10 Weinan, E., Ren, W.Q., and Vanden-Eijnden, E. (2007) Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *J. Chem. Phys.*, **126** (16), 164103.
- 11 Ren, W., Vanden-Eijnden, E., Maragakis, P., and Weinan, E. (2005) Transition pathways in complex systems: application of the finite-temperature string method to the alanine dipeptide. *J. Chem. Phys.*, **123** (13), 134109.
- 12 Peters, B., Heyden, A., Bell, A.T., and Chakraborty, A. (2004) A growing string method for determining transition states: comparison to the nudged elastic band and string methods. *J. Chem. Phys.*, **120** (17), 7877–7886.
- 13 Burger, S.K. and Yang, W.T. (2006) Quadratic string method for determining the minimum-energy path based on multiobjective optimization. *J. Chem. Phys.*, **124** (5), 054109.
- 14 Weinan, E. and Ren, Weiqing (2005) Finite temperature string method for the study of rare events. *J. Chem. Phys.*, **109** (14), 6688–6693.
- 15 Burger, S.K. and Yang, W.T. (2006) Automatic integration of the reaction path using diagonally implicit Runge-Kutta methods. *J. Chem. Phys.*, **125** (24), 244108.
- 16 Burger, S.K. and Yang, W.T. (2006) A combined explicit-implicit method for high accuracy reaction path integration. *J. Chem. Phys.*, **124** (22), 224102.

- 17 Tsai, C.J. and Jordan, K.D. (1993) Use of an eigenmode method to locate the stationary-points on the potential-energy surfaces of selected argon and water clusters. *J. Phys. Chem.*, **97** (43), 11227–11237.
- 18 Sun, J.Q. and Ruedenberg, K. (1993) Gradient extremals and steepest descent lines on potential-energy surfaces. *J. Chem. Phys.*, **98** (12), 9707–9714.
- 19 Quapp, W., Hirsch, M., Imig, O., and Heidrich, D. (1998) Searching for saddle points of potential energy surfaces by following a reduced gradient. *J. Comput. Chem.*, **19** (9), 1087–1100.
- 20 Maeda, S., Watanabe, Y., and Ohno, K. (2005) A scaled hypersphere interpolation technique for efficient construction of multidimensional potential energy surfaces. *Chem. Phys. Lett.*, **414** (4–6), 265–270.
- 21 Maeda, S. and Ohno, K. (2005) Global mapping of equilibrium and transition structures on potential energy surfaces by the scaled hypersphere search method: applications to *ab initio* surfaces of formaldehyde and propyne molecules. *J. Phys. Chem. A*, **109** (25), 5742–5753.
- 22 Maeda, S. and Ohno, K. (2005) A new approach for finding a transition state connecting a reactant and a product without initial guess: applications of the scaled hypersphere search method to isomerization reactions of HCN, (H₂O) (2), and alanine dipeptide. *Chem. Phys. Lett.*, **404** (1–3), 95–99.
- 23 Ohno, K. and Maeda, S. (2004) A scaled hypersphere search method for the topography of reaction pathways on the potential energy surface. *Chem. Phys. Lett.*, **384** (4–6), 277–282.
- 24 Burger, S.K., Liu, Y., Sarkar, U., and Ayers, P.W. (2009) Moving least-squares enhanced Shepard interpolation for the fast marching and string methods. *J. Chem. Phys.*, **130**, 024103.
- 25 Dey, B.K., Janicki, M.R., and Ayers, P.W. (2004) Hamilton-Jacobi equation for the least-action/least-time dynamical path based on fast marching method. *J. Chem. Phys.*, **121** (14), 6667–6679.
- 26 Dey, B.K. and Ayers, P.W. (2006) A Hamilton-Jacobi type equation for computing minimum potential energy paths. *Mol. Phys.*, **104** (4), 541–558.
- 27 Sethian, J.A. (1996) A fast marching level set method for monotonically advancing fronts. *Proc. Natl. Acad. Sci. U.S.A.*, **93** (4), 1591–1595.
- 28 Sethian, J.A. and Adalsteinsson, D. (1997) An overview of level set methods for etching, deposition, and lithography development. *IEEE T. Semiconduct. M.*, **10** (1), 167–184.
- 29 Sethian, J.A. (1999) Fast marching methods. *SIAM Rev.*, **41** (2), 199–235.
- 30 Sethian, J.A. and Vladimirsky, A. (2000) Fast methods for the Eikonal and related Hamilton-Jacobi equations on unstructured meshes. *Proc. Natl. Acad. Sci. U.S.A.*, **97** (11), 5699–5703.
- 31 Liu, Y. and Ayers, P.W. (2009) Finding minimum energy reaction paths on *ab initio* potential energy surfaces using the fast marching method, submitted.
- 32 Hirsch, M. and Quapp, W. (2004) Reaction channels of the potential energy surface: application of Newton trajectories. *J. Mol. Struct. (THEOCHEM)*, **683** (1–3), 1–13.
- 33 Quapp, W. (2004) Newton trajectories in the curvilinear metric of internal coordinates. *J. Math. Chem.*, **36** (4), 365–379.
- 34 Quapp, W., Hirsch, M., and Heidrich, D. (2004) An approach to reaction path branching using valley-ridge inflection points of potential-energy surfaces. *Theor. Chem. Acc.*, **112** (1), 40–51.
- 35 Quapp, W. (2004) Reaction pathways and projection operators: Application to string methods. *J. Comput. Chem.*, **25** (10), 1277–1285.
- 36 Quapp, W. (2005) A growing string method for the reaction pathway defined by a Newton trajectory. *J. Chem. Phys.*, **122** (17), 174106.
- 37 Quapp, W. (2007) Finding the transition state without initial guess: the growing string method for Newton trajectory to isomerization and enantiomerization reaction of alanine dipeptide and poly(15) alanine. *J. Comput. Chem.*, **28** (11), 1834–1847.
- 38 Laio, A. and Parrinello, M. (2002) Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.*, **99** (20), 12562–12566.

- 39 Grubmuller, H. (1995) Predicting slow structural transitions in macromolecular systems – conformational flooding. *Phys. Rev. A*, **52** (3), 2893–2906.
- 40 Zhao, H.K. (2005) A fast sweeping method for Eikonal equations. *Math. Comput.*, **74** (250), 603–627.
- 41 Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. (1989) *Numerical Recipes (FORTRAN)*, Cambridge University Press, Cambridge, New York, pp. 229–232.
- 42 Collins, M.A. (2002) Molecular potential-energy surfaces for chemical reaction dynamics. *Theor. Chem. Acc.*, **108** (6), 313–324.
- 43 Bettens, R.P.A. and Collins, M.A. (1999) Learning to interpolate molecular potential energy surfaces with confidence: A Bayesian approach. *J. Chem. Phys.*, **111** (3), 816–826.
- 44 Burgers, P.C., Lifshitz, C., Ruttink, P.J.A., Schaftenaar, G., and Terlouw, J.K. (1989) The [Ch5No] + . potential-energy surface – distonic ions. ion dipole complexes and hydrogen-bridged radical cations. *Org. Mass. Spectrom.*, **24** (8), 579–590.
- 45 Warshel, A. and Levitt, M. (1976) Theoretical studies of enzymic reactions – dielectric, electrostatic and steric stabilization of carbonium-ion in reaction of lysozyme. *J. Mol. Biol.*, **103** (2), 227–249.
- 46 Warshel, A., Hwang, J.K., and Aqvist, J. (1992) Computer-simulations of enzymatic-reactions – examination of linear free-energy relationships and quantum-mechanical corrections in the initial proton-transfer step of carbonic-anhydrase. *Faraday Discuss.*, **93**, 225–238.
- 47 Bentzien, J., Muller, R.P., Florian, J., and Warshel, A. (1998) Hybrid *ab initio* quantum mechanics molecular mechanics calculations of free energy surfaces for enzymatic reactions: the nucleophilic attack in subtilisin. *J. Phys. Chem. B*, **102** (12), 2293–2301.
- 48 Cisneros, G.A., Liu, H.Y., Zhang, Y.K., and Yang, W.T. (2003) *Ab initio* QM/MM study shows there is no general acid in the reaction catalyzed by 4-oxalocrotonate tautomerase. *J. Am. Chem. Soc.*, **125** (34), 10384–10393.
- 49 Cisneros, G.A., Wang, M., Silinski, P., Fitzgerald, M.C., and Yang, W.T. (2004) The protein backbone makes important contributions to 4-oxalocrotonate tautomerase enzyme catalysis: Understanding from theory and experiment. *Biochemistry*, **43** (22), 6885–6892.
- 50 Cisneros, G.A., Wang, M., Silinski, P., Fitzgerald, M.C., and Yang, W.T. (2006) Theoretical and experimental determination on two substrates turned over by 4-oxalocrotonate tautomerase. *J. Phys. Chem. A*, **110** (2), 700–708.
- 51 Hu, H., Lu, Z.Y., and Yang, W.T. (2007) QM/MM minimum free-energy path: Methodology and application to triosephosphate isomerase. *J. Chem. Theor. Comput.*, **3** (2), 390–406.
- 52 Hu, H. and Yang, W.T. (2008) Free energies of chemical reactions in solution and in enzymes with *ab initio* quantum mechanics/molecular mechanics methods. *Annu. Rev. Phys. Chem.*, **59**, 573–601.
- 53 Hu, H., Lu, Z.Y., Parks, J.M., Burger, S.K., and Yang, W.T. (2008) Quantum mechanics/molecular mechanics minimum free-energy path for accurate reaction energetics in solution and enzymes: sequential sampling and optimization on the potential of mean force surface. *J. Chem. Phys.*, **128** (3).
- 54 Zhang, Y.K. and Yang, W.T. (1999) Studying enzyme reactions with a pseudobond *ab initio* QM/MM approach. *Abstr. Pap. Am. Chem. Soc.*, **218**, U528.
- 55 Zhang, Y.K., Lee, T.S., and Yang, W.T. (1999) A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *J. Chem. Phys.*, **110** (1), 46–54.
- 56 Zhang, Y.K., Liu, H.Y., and Yang, W.T. (2000) Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined *ab initio* QM/MM potential energy surface. *J. Chem. Phys.*, **112** (8), 3483–3492.
- 57 Zhang, Y.K. (2005) Improved pseudobonds for combined *ab initio* quantum mechanical/molecular mechanical methods. *J. Chem. Phys.*, **122** (2).

- 58 Gao, J.L., Amara, P., Alhambra, C., and Field, M.J. (1998) A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *J. Phys. Chem. A*, **102** (24), 4714–4721.
- 59 Eurenus, K.P., Chatfield, D.C., Brooks, B.R., and Hodoscek, M. (1996) Enzyme mechanisms with hybrid quantum and molecular mechanical potentials. 1. Theoretical considerations. *Int. J. Quantum. Chem.*, **60** (6), 1189–1200.
- 60 Das, D., Eurenus, K.P., Billings, E.M., Sherwood, P., Chatfield, D.C., Hodoscek, M., and Brooks, B.R. (2002) Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method. *J. Chem. Phys.*, **117** (23), 10534–10547.
- 61 Amara, P. and Field, M.J. (2003) Evaluation of an ab initio quantum mechanical/molecular mechanical hybrid-potential link-atom method. *Theor. Chem. Acc.*, **109** (1), 43–52.
- 62 vi-Kesavan, L.S., Garcia-Viloca, M., and Gao, J. (2003) Semiempirical QM/MM potential with simple valence bond (SVB) for enzyme reactions. Application to the nucleophilic addition reaction in haloalkane dehalogenase. *Theor. Chem. Acc.*, **109** (3), 133–139.
- 63 Liu, H.Y., Lu, Z.Y., Cisneros, G.A., and Yang, W.T. (2004) Parallel iterative reaction path optimization in ab initio quantum mechanical/molecular mechanical modeling of enzyme reactions. *J. Chem. Phys.*, **121** (2), 697–706.
- 64 Ayala, P.Y. and Schlegel, H.B. (1997) A combined method for determining reaction paths, minima, and transition state geometries. *J. Chem. Phys.*, **107** (2), 375–384.
- 65 Wang, S.C., Johnson, W.H., and Whitman, C.P. (2003) The 4-oxalocrotonate tautomerase- and YwhB-catalyzed hydration of 3E-haloacrylates: implications for the evolution of new enzymatic activities. *J. Am. Chem. Soc.*, **125** (47), 14282–14283.
- 66 Horvat, C.M. and Wolfenden, R.V. (2005) A persistent pesticide residue and the unusual catalytic proficiency of a dehalogenating enzyme. *Proc. Natl. Acad. Sci. U.S.A.*, **102** (45), 16199–16202.
- 67 de Jong, R.M., Brugman, W., Poelarends, G.J., Whitman, C.P., and Dijkstra, B.W. (2004) The X-ray structure of trans-3-chloroacrylic acid dehalogenase reveals a novel hydration mechanism in the tautomerase superfamily. *J. Biol. Chem.*, **279** (12), 11546–11552.

Part Two

Nucleic Acids, Amino Acids, Peptides and Their Interactions

6

Chemical Origin of Life: How do Five HCN Molecules Combine to form Adenine under Prebiotic and Interstellar Conditions

Debjani Roy and Paul von Ragué Schleyer

6.1

Introduction

Experiments show that simple molecules combine under prebiotic conditions to give the fundamental building blocks of life – amino acids, nucleotides, carbohydrates and other essential compounds. However, the origin of even the simplest of these biomolecules remains a fascinating but unsolved puzzle [1, 2]. They could have formed from smaller molecules present on primitive earth, either very slowly over millions of years or rapidly before the earth cooled down. Asteroids may have brought them from outer space. This chapter provides a detailed examination of the chemical processes involved in the genesis of adenine, one of the four building blocks of DNA, and an abundant biochemical molecule found on earth. First we mention a chronology of research in prebiotic chemistry, followed by the details of the experiments that led to the detection of biomolecules under simulated prebiotic conditions. We then elucidate the quantum chemical investigation for a thermodynamically viable, step by step pathway, for the formation of adenine under prebiotic conditions.

6.1.1

Prebiotic Chemistry: Experimental Endeavor to Synthesize the Building Blocks of Biopolymers

In 1828, F. Wöhler reported the first chemical synthesis of a simple organic molecule (urea) from inorganic starting materials (silver cyanate and ammonium chloride) [3]. Wöhler's work led to a new era in prebiotic chemistry. Figure 6.1 demonstrates the occurrences of these research milestones in a somewhat chronological order. The first observations relating to the prebiotic syntheses of ribose date back to 1861 when Butlerow showed that sugars could be made by mild heating of formaldehyde in the presence of $\text{Ca}(\text{OH})_2$ catalyst; this became known as the formose reaction [4].

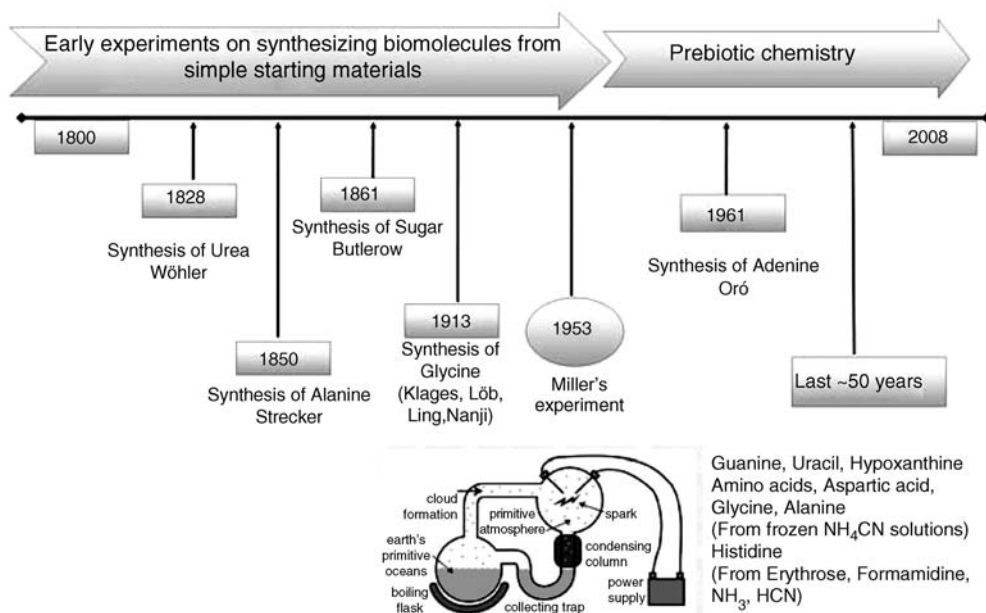


Figure 6.1 Progress in prebiotic chemistry for last two centuries.

By the end of the nineteenth century, a large amount of research on organic synthesis had been performed that led to the abiotic formation of fatty acids and sugars using electric discharges with various gas mixtures [5]. This work was continued into the twentieth century by Löb, Baudish and others on the synthesis of amino acids by exposing wet formamide (HCONH_2) to a silent electrical discharge [6] and to UV light [7]. These efforts heralded the dawn of prebiotic chemistry (Figure 6.1).

It was not until 1953 that the first successful synthesis of organic compounds under plausible *primordial conditions* was accomplished, by electric discharges acting for a week on a mixture of CH_4 , NH_3 , H_2 and H_2O ; racemic mixtures of several protein amino acids were produced, as well as hydroxy acids, urea and other organic molecules [8, 9]. Miller's first paper on the formation of amino acid under possible primitive earth condition was published only a few weeks after Watson and Crick reported their DNA double-helix model in *Nature*. The link between the two nascent fields began to develop a few years later when Juan Oró demonstrated the remarkable ease by which adenine, one of the nucleobases in DNA and RNA, could be produced through the oligomerization of hydrogen cyanide under basic conditions [10–12]. In 1968, Orgel *et al.* showed that cyanoacetylene is a major product of the action of an electric discharge on a mixture of methane and nitrogen, and that cyanoacetylene is a possible source of the pyrimidine bases, uracil and cytosine [13]. Purines like adenine [14], guanine [15], uracil [16] and hypoxanthine [17] have been detected in concentrated ammonium cyanide (NH_4CN) solution after acid hydrolysis

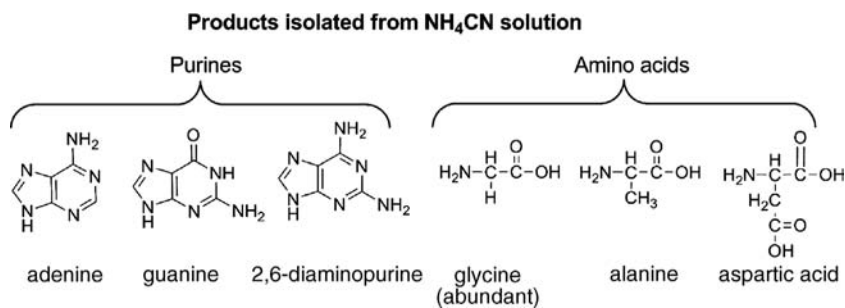


Figure 6.2 Products isolated from HCN in water–ammonia solution.

(Figure 6.2). Some pyrimidine derivatives, such as orotic acid, 5-hydroxyuracil and 4,5 dihydroxypyrimidine, along with adenine, were detected in a solution of 0.1 M NH_4CN kept for 4–12 months after neutral or acid hydrolysis [18].

The amino acids alanine, glycine and aspartic acid were identified among the products formed by the condensation of hydrogen cyanide in aqueous ammonia [18–23]. Histidine, a basic amino acid containing an imidazole group, was synthesized from erythrose and formamidine through condensation reaction and a Strecker synthesis without the isolation of any intermediate [24]. The precursor molecules (erythrose, formamidine, hydrogen cyanide and ammonia) used for the synthesis of histidine are all considered to be prebiotic compounds. A recent noteworthy achievement should be mentioned that concludes this section on a very exciting note. The 1953 Miller–Urey synthesis had two sibling studies, neither of which was published. Vials containing the products from those experiments were recently recovered and reanalyzed using modern technology and the results have been reported in *Science* [25]. Miller identified five amino acids: aspartic acid, glycine, α -aminobutyric acid and two versions of alanine. The eleven vials scientists recovered from the unpublished aspirator experiment, however, produced 22 amino acids and the same five amines at yields comparable to the original experiment [25].

6.1.2

Key Role of HCN as a Precursor for Prebiotic Compounds

Owing to its high degree of unsaturation, HCN is an energy-rich, reactive molecule that undergoes addition reactions exothermically. This high energy prebiotic precursor is produced in appreciable amounts, for example, by the action of electric discharge on simulated prebiotic atmosphere [26]. Aqueous solutions of cyanides are frequently used in experiments related to prebiotic chemistry. It is remarkable how some of HCN's reactions appear to aim at “bio-molecular building blocks.” Similar processes are expected to be present planetary environments, such as Titan [27–29], Europa [20, 30–32], Ganymede and Callisto [33, 34], but in the absence of liquid water, due to abundance of CN group containing molecules in interstellar space [35], in comets [36, 37] and in the atmosphere of Titan [27, 38].

6.1.3

Prebiotic Experiments and Proposed Pathways for the Formation of Adenine

The HCN pentamer, adenine (a constituent of DNA, RNA and many coenzymes), is one of the most abundant biochemical molecules. The synthesis of adenine from hydrogen cyanide in water–ammonia systems under conditions assumed to have existed on prebiotic earth, first demonstrated by Oró in 1960 [10–12, 21], is so remarkable that we believe it to have some relevance to the prebiotic accumulation of purines. In Oró's experiment, adenine was formed in 0.5% yield by heating solutions of ammonium cyanide (>1.0 M) at 70°C for several days. Since then, the abiotic synthesis of adenine from the polymerization of HCN under various conditions has been achieved many times [11, 13, 39–43]. Adenine also has been obtained in a very high yield (20%) by heating HCN with liquid ammonia in a sealed tube [44]. In 1978 Ferris *et al.* detected 0.04% adenine from 0.1 M NH_4CN kept in dark at room temperature for 4–12 months [45]. To simulate prebiotic synthetic processes on Europa and other ice-covered planets and satellites, Levy *et al.* have investigated the prebiotic synthesis of organic compounds from dilute solutions of NH_4CN frozen for 25 years at -20 and -78°C [30]. They found that both adenine and guanine, as well as a simple set of amino acids dominated by glycine, are produced in substantial yields under these conditions [30]. Moreover, Since 1961, several reaction pathways have been reported in the literature to account for the formation of adenine under prebiotic earth conditions (Figure 6.3) [11, 13, 39–43]. However, the puzzle: “How do five HCN molecules combine to form adenine under prebiotic conditions?” remained unsolved. An *ab initio* mechanistic investigation of the HCN dimerization and an adenine protonation study do shed some light on the problem [46, 47]. We are the first research group to systematically investigate the step-by-step mechanisms for the oligomerization of HCN, by applying quantum-chemical methods, and propose a viable route for formation of adenine under prebiotic earth conditions [48]. Rainer Glaser *et al.* have published parallel research, in the journal *Astrobiology*, that also sheds light on the abiotic origins of adenine, although from a different perspective [49].

6.2

Computational Investigation

Experimental investigation for a thermodynamically feasible pathway for the formation of adenine under prebiotic conditions would be very difficult, since adenine is not formed cleanly, yields are small and many steps are involved. Some have considered it too difficult for scientific study because the direct evidence is long gone and we can only work by plausible inference. Quantum chemical computations are powerful methods to study the structure and behavior of molecules related to prebiotic origin of life. Reaction mechanisms involving several intermediates and transition states (which are difficult to detect and identify experimentally) can be studied effectively computationally. These allow selection

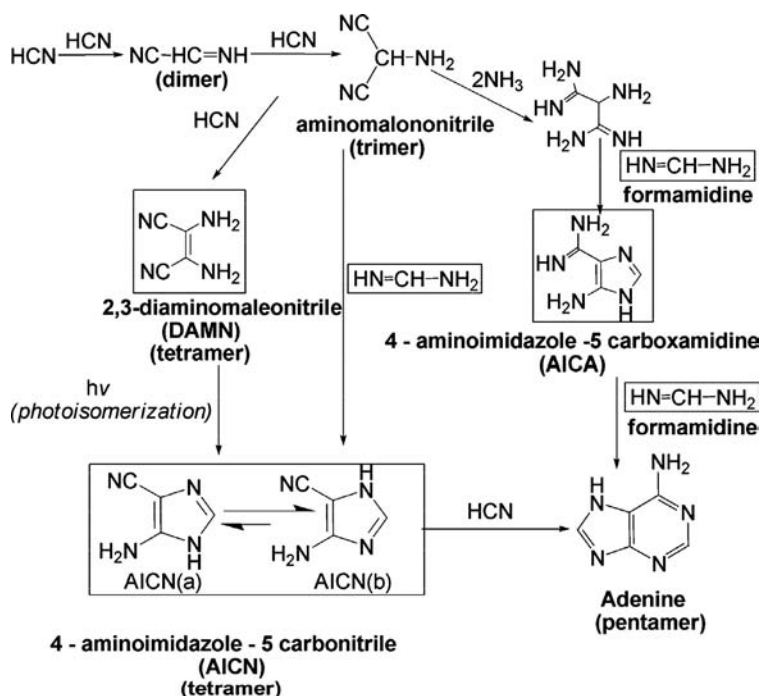


Figure 6.3 Summary of pathways proposed to account for the formation of adenine under prebiotic conditions [11, 13, 39–43]. Experimentally detected intermediates, which provide a clue to understanding the overall mechanism, are enclosed in boxes. Our computational investigation infers that two

tautomers of AICN can lead to adenine and AICN (b) is the more stable [48]. An alternative photoisomerization route is proposed for the formation of AICN from another tetramer DAMN. (Figure reproduced under Author Rights (3b) PNAS copyright [48].)

among various possibilities. Furthermore, quantum chemical computations are advantageous since interstellar or prebiotic conditions (solvation and high temperature, for instance) are implicitly considered in calculations using the quantum chemical techniques.

Although adenine is a difficult problem to approach, several striking observations over the years have allowed the formulation of plausible scenarios for its prebiotic synthesis. Clues for finding a viable pathway are provided by four putative intermediates detected in the product mixtures (Figure 6.3): formamidine, diaminomaleonitrile (DAMN), 4-aminoimidazole-5-carbonitrile (AICN) [13, 39, 41–43] and 4-aminoimidazole-5-carboxamide (AICA) [12, 39]. We propose a possible mechanism for prebiotic synthesis of adenine with calculated low energy barriers, where solvent participation is important, especially of the first H₂O or ammonia molecule [based on density functional theory (DFT) computations described in the next section] [50].

6.2.1

Method

The computations were performed at the B3LYP level of theory using 6-31G and 6-311 + G** basis sets. In addition, the results of CCSD (T) (*ab initio*) theory have been reported where necessary (simulated the first HCN + RNH₂ addition step with the highest barrier). Stable structures and transition states were fully optimized at the mentioned level of theory. All critical points were further characterized by analytic computations of harmonic vibrational frequencies at the same level/basis set. Transition states have been characterized by one imaginary frequency (first order saddle point) on the potential energy surface (PES). Intrinsic reaction coordinate (IRC) analysis was performed to determine the minimum energy pathways (MEPs) at the mentioned level of theory. Transition states were found to connect proper reactants and products. Vibrational frequencies were calculated within the harmonic approximation at the same level of theory as for geometries. The reaction barriers are defined as the difference in sum of electronic and zero-point energies ($\epsilon_0 + \epsilon_{\text{ZPE}}$) of the reactant-complex and transition state. ΔG is defined as difference between the sum of electronic and thermal free energies ($\epsilon_0 + G_{\text{corr}}$) of reactant and product.

Instead of modeling bulk solvation by an explicit shell involving many solvent molecules surrounding the solute [51], we employed the polarizable continuum model (PCM) implemented in the Gaussian 98 program [52]. The PCM “bulk solvent medium” is simulated as a continuum of the dielectric constant (ϵ). This surrounds a solute cavity, which is defined by the union of a series of interlocking spheres centered on the atoms. Our computed single-point PCM bulk solvent simulations employed the optimized equilibrium geometries. We have added an estimated $\sim 2.5 \text{ kcal mol}^{-1}$ (1 kcal = 4.184 kJ) ZPVE correction to the single-point PCM energy.¹⁾

More sophisticated solvation treatments are not called for in the absence of experimental data.

6.2.2

Thermochemistry of Pentamerization

The pentamerization of HCN to give adenine is very exothermic overall ($\Delta G^{298} = 53.7 \text{ kcal mol}^{-1}$ Figure 6.4). Moreover, all the successive oligomerization steps (from

1) This corrected energy value was derived as follows: The ZPVE corrections to the gas-phase energies are nearly the same with different numbers of solvent molecules, zero, one or two. For example, the relative energy of the transition structure (TS) for the first step without solvent is 62.9 (without) and 60.4 kcal mol⁻¹ (with ZPVE correction). The relative energy of the TS with

one water molecule is 40.5 (without) and 38.0 kcal mol⁻¹ (with ZPVE). With two water molecules, the TS relative energy is 39.9 (without) and 37.6 kcal mol⁻¹ (with ZPVE correction). Consequently, the average ZPVE correction ($\sim 2.5 \text{ kcal mol}^{-1}$) was added to the PCM single-point energy.

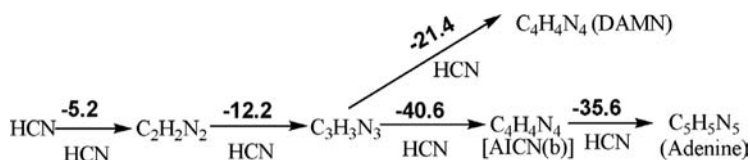


Figure 6.4 Overall energy for pentamerization of HCN to adenine (5HCN $\rightarrow$ C₅H₅N₅) is -93.6 kcal mol⁻¹ (enthalpy) [$\Delta G^{298} = -53.7$ kcal mol⁻¹]. Note that the last crucial step for formation of the pentamer (adenine) from tetramer [AICN(b)] is highly exergonic. (Figure reproduced under Author Rights (3b) PNAS copyright [48].)

monomer to dimer, dimer to trimer, etc.) are quite exothermic as well. Entropy is unfavorable. However, the reaction energetics is mostly governed by the enthalpy change, especially at the low temperatures for some of the experiments. However, one has to keep in mind that although adenine formation from five HCN is highly exothermic, it does not insure that the reaction will actually proceed. Some intermediate steps might be associated with huge barriers.

6.2.3

Detailed Step by Step Mechanism

6.2.3.1 DAMN vs AICN as Adenine Precursors

Our study began with the last, crucial step, in which HCN tetramer AICN, upon reaction with HCN, leads to the formation adenine. This is the key part of the general adenine formation starting from HCN. Although both DAMN and AICN have the same empirical formula, (HCN)₄, AICN was selected as the adenine precursor because of its greater structural similarity to adenine. (Like adenine, AICN has a C–C–C–N sequence, whereas DAMN has a C–C–C–C backbone, Figure 6.5.) Therefore DAMN cannot be directly involved in the pathway. Moreover, AICN is thermodynamically more stable than DAMN (the energy difference being 19.2 kcal mol⁻¹ according to our computations Figure 6.4). DAMN may be a kinetically-controlled side product.

Furthermore, many details of *photoisomerization* of DAMN to AICN in water are established, but no similar information exists for the *non-photolytic* conversion of DAMN into AICN or to adenine [39, 41, 42]. Since we are investigating the mechanism for formation of adenine under non-photochemical conditions (reactions in the dark), consideration of HCN tetramer DAMN as adenine precursor is out of context.

6.2.3.2 Is an Anionic Mechanism Feasible in Isolation?

Since HCN is a weak acid and adenine forms in ammoniacal solution, an anionic mechanism for the abiotic formation of adenine seems plausible from a physical organic chemist's point of view. However, the anionic mechanism proved not to be

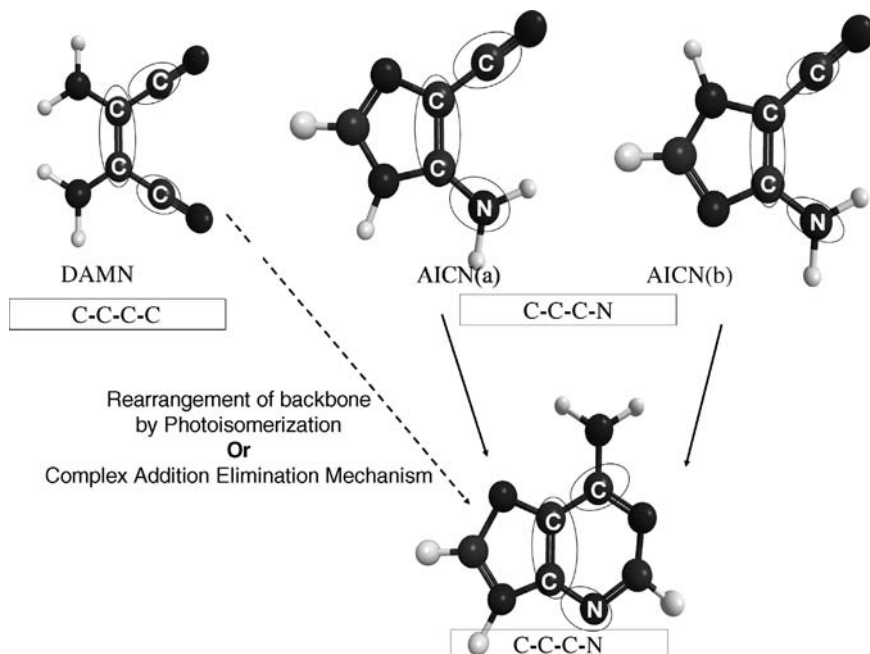


Figure 6.5 DAMN and AICN are both $C_4H_4N_4$ isomers. Like adenine, both AICN-(a) and -(b) have a $C-C-C-N$ sequence, whereas DAMN has a $C-C-C-C$ backbone not present in adenine.

feasible computationally!²⁾ The anionic intermediates ii and iii (Figure 6.6) fall apart and revert back to the starting materials (ring-opened isomer) on optimization. Mechanisms involving free radical intermediates also were without prospect. In contrast, we were successful in optimizing the reactant, product and transition states for *neutral* mechanism. However, the first step of addition of HCN to the NH_2 group of AICN proceeds via a four-center transition state and has a prohibitively high gas-phase activation barrier relative to those of the 1 : 1 (HCN) (AICN) complex ($60.4 \text{ kcal mol}^{-1}$, Figure 6.6). Moreover, an intermediate step for neutral uncatalyzed mechanism (the concerted six-membered ring closure and H transfer step) also has a huge in reaction barrier (53 kcal mol^{-1} , Figure 6.6). Therefore, the neutral uncatalyzed mechanism can also be ruled out as a possibility for the formation of adenine.

2) K. Najafian, our collaborator in the original work [48], carried out extensive computations on intermediates and transition states that might be involved in possible anionic and radical mediated pathways.

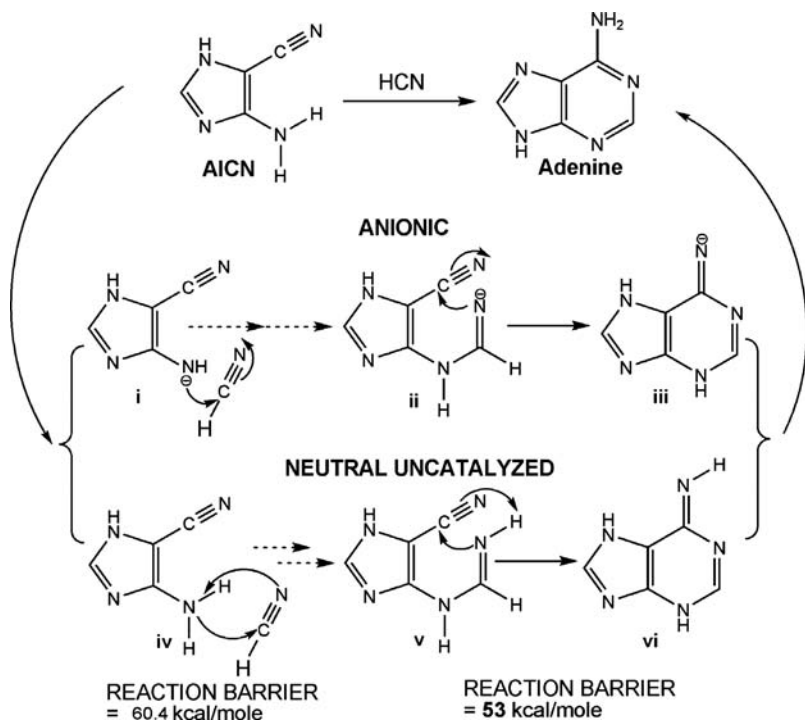


Figure 6.6 Anionic and free radical mediated mechanisms are unfeasible in the gas phase (interstellar condition) (see footnote 2). Upon optimization anionic intermediates ii and iii revert back to i (reactants). In contrast, the reactant, product and transition state geometries

for a neutral mechanism can be optimized in isolation. However, two large reaction barriers associated with the two steps shown makes neutral *uncatalyzed* mechanisms unfavorable too. (Figure reproduced under Author Rights (3b) PNAS copyright [48].)

6.2.3.3 Two Tautomeric forms of AICN: Which one is the Favorable Precursor for Adenine Formation under Prebiotic Conditions?

Tautomerization of imidazoles is indeed known, but not the relative energies of the AICN isomers. We first point out explicitly that two AICN tautomeric forms, AICN(a) and AICN(b) (Figure 6.3), can exist and that the latter is favored strongly at equilibrium (by $\Delta G^{298} = 3.73 \text{ kcal mol}^{-1}$ in the gas phase and $1.73 \text{ kcal mol}^{-1}$ with bulk solvation) [48]. Since the energetically more favorable isomer AICN(b) is not the one assumed in pathways proposed by experimentalists, we explicitly called attention to the structures and energies of both isomers. The pathway to adenine from the less stable AICN(a) is precluded by the second high reaction barrier ($41.6 \text{ kcal mol}^{-1}$) associated with the six-membered ring closure step. Figure 6.7 depicts the reaction profile. Note that the reaction barriers are reported with a smaller basis set, 6-31G*. Since we did not consider this pathway to be the viable one, we believe that this is sufficient to explain the qualitative nature of the pathway.

The favorable mechanism of adenine formation starting from the thermodynamically more stable AICN(b) tautomer is reported here. The first step of addition of

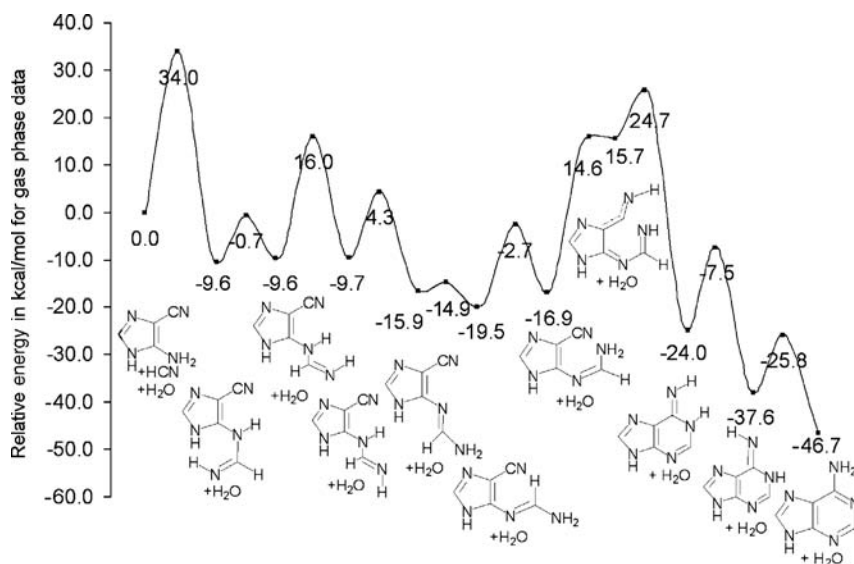


Figure 6.7 Reaction profile for adenine formation from the less stable AICN(a) isomer with one H₂O acting as a catalyst (gas phase). This pathway is precluded by the second high

barrier, which requires over 40 kcal mol⁻¹ activation from the preceding lowest energy minimum. (Figure reproduced under Author Rights (3b) PNAS copyright [48].)

HCN to the NH₂ group of AICN(b) proceeds via a four-centered transition state (TS0) and has a high activation barrier relative to those of the 1 : 1 (HCN) (AICN) complex (60.4 kcal mol⁻¹) (Figures 6.6 and 6.8). Moreover, a prohibitive reaction barrier (53 kcal mol⁻¹) also precludes a subsequent step in the neutral uncatalyzed pathway (involving the concerted six-membered ring closure and H transfer, Figure 6.6). The reaction barriers are much too high for this neutral uncatalyzed pathway to be viable for adenine formation at room temperature or below. The involvement of additional molecules serving as catalysts is required to lower the high barriers.

The unfavorable four-center transition structure (TS0) associated with the first step from AICN(b) is shown in Figure 6.8. A six-center transition state formed via catalytic participation of another molecule (e.g., water/ammonia) can considerably reduce the barrier. The importance of water-assisted proton transfer is well known in keto–enol tautomerization [53, 54]. The water bridge connects the donor and acceptor sites and stabilizes the transition structure. Hence we incorporated water in the mechanism. Several transition states with a specific water molecule leading to the addition product were computed. The energetically most favorable six-center cyclic transition structure leading to the HCN–AICN(b) addition product (TS1) is depicted in Figure 6.8. The water molecule in TS1 transfers its hydrogen-bonded proton to the HCN nitrogen concertedly with the formation of the new bond between the AICN amine nitrogen and the electron-deficient HCN carbon.

The inclusion of one specific water molecule decreases the reaction barrier drastically, from 60.4 (without water) to 38.0 kcal mol⁻¹ (with a single H₂O)

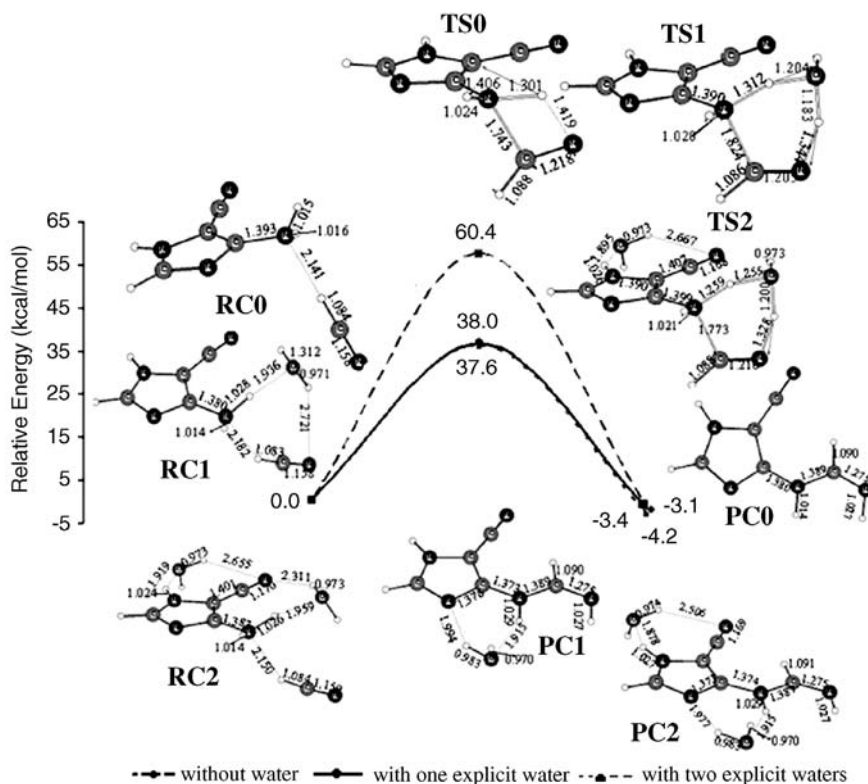


Figure 6.8 Gas-phase reaction barrier for the key first rate-determining step with high barrier (B3LYP/6-311 + C** + ZPE). All the reactant complexes are stabilized relative to their isolated components. The reaction barrier for the neutral uncatalyzed pathway is ~ 60.4 kcal mol $^{-1}$. Water (solvent) is a positive participant in the transition step. Incorporation of one and two water

molecules in the system reduces the reaction barrier to 38.0 and 37.6 kcal mol $^{-1}$, respectively. Optimized geometries for reactant complexes (RC), transition states (TS) and product complexes (PC) for systems with 0, 1 and 2 explicit water molecules are depicted in the figure. (Figure reproduced under Author Rights (3b) PNAS copyright [48].)

(Figure 6.8). Consequently, we also explored the effect of more than one solvent molecule in the mechanism. A second H₂O does not participate in the proton relay effectively, but can form relatively strong hydrogen bonds stabilizing the reactant, product and the transition state complexes. The energetically most favorable transition structure with two water molecules (TS2) is also depicted in Figure 6.8. However, inclusion of two explicit H₂O molecules only decreases the reaction barrier by an additional 0.4 kcal mol $^{-1}$ (to 37.6 kcal mol $^{-1}$). It does not seem likely that additional explicit H₂O molecules would have much of a further effect. However, a complete solvation shell does have a significant influence (see below).

The reported syntheses of adenine under simulated prebiotic conditions were carried out with HCN dissolved in water–ammonia solutions. Besides maintaining

the pH of the medium, ammonia might also participate mechanistically. We also investigated the effect of ammonia as a catalyst on the first addition step. Indeed, an explicit NH_3 molecule is as good a catalyst as an explicit H_2O molecule. The reaction barrier for the first step, the NH_3 -catalyzed addition of HCN to AICN(b) (with one explicit NH_3), is $37.1 \text{ kcal mol}^{-1}$, as compared to $38.0 \text{ kcal mol}^{-1}$ with an explicit H_2O . While computations based on species in isolation (gas phase) may simulate extraterrestrial conditions, they are not adequate to predict prebiotic processes on the primitive earth, where reactions might have taken place in solution. In addition to the explicit solvent modeling (see above) we simulated the bulk solvation effect by employing the polarizable continuum model (PCM) [50], which considers the solvent as a macroscopic continuum of dielectric constant. Bulk solvation stabilizes structures involving greater charge separation (e.g., transition states) preferentially. Indeed, bulk solvation reduces the reaction barriers for the rate-determining step to $33.9 \text{ kcal mol}^{-1}$ (from 37.6) and to $35.2 \text{ kcal mol}^{-1}$ (from 37.6) for specific $\text{H}_2\text{O}/\text{NH}_3$ catalyzed mechanisms, respectively. These barriers are low enough to be consistent with the experimental observations as well as conjectures regarding the abiotic genesis of adenine.

The subsequent steps from the HCN–AICN adduct are depicted in Figure 6.9 (for catalysis by two H_2O) and Figure 6.10 (for catalysis by two NH_3). The effects

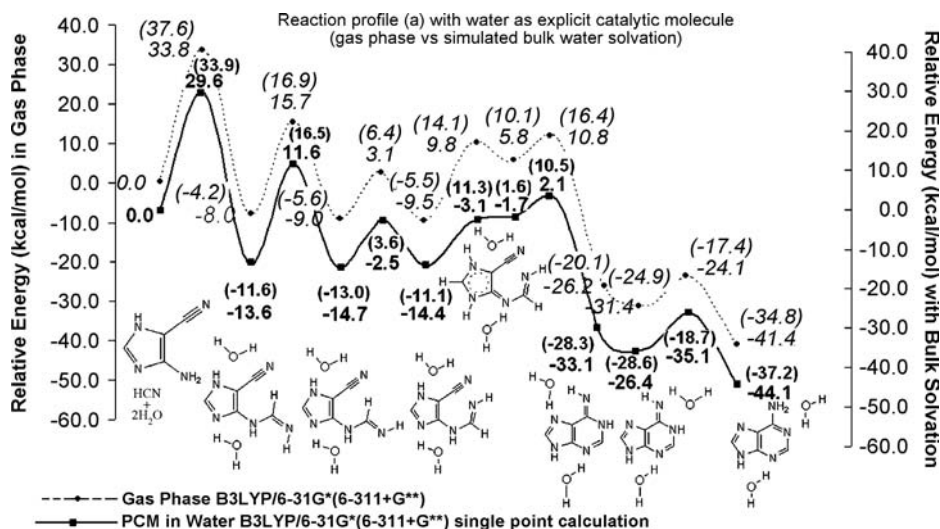


Figure 6.9 Reaction profiles for the formation of adenine starting from AICN(b) and HCN in gas phase and in the solvent phase via explicit water-catalyzed mechanism (two water molecules). The barrier heights are reported with both 6-31G* and at 6-311 + G** basis set (in parentheses). All the stable minima are shown. Comparison of reaction profiles in a vacuum and

in aqueous solution clearly shows that the transition states are stabilized through electrostatic effect of the solvent. The reaction seems easier in aqueous solution than in the gas phase: the first step is rate-determining in both cases. (Figure reproduced under Author Rights (3b) PNAS copyright [48].)

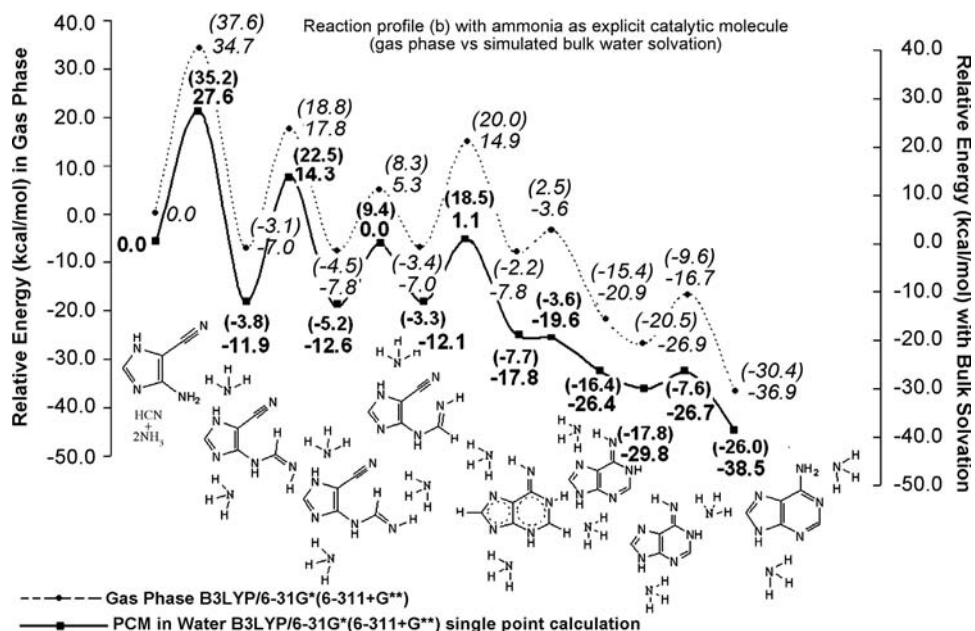


Figure 6.10 Reaction profiles for the formation of adenine starting from AICN(b) and HCN in the gas phase and with simulated bulk water solvation via explicit solvent-catalyzed mechanisms (two explicit NH_3 molecules). (Figure reproduced under Author Rights (3b) PNAS copyright [48].)

of bulk solvation are included in both plots. Except for the syn-anti hydrogen transfer and the C–N bond rotation in the second and third steps, all the stages require the catalytic participation of at least one H_2O or NH_3 molecule. One of the solvent molecules becomes positive participant of the reaction (proton transfer through a hydrogen bond) as a catalyst, while the other one is found to be hydrogen bonded to the stationary points. The proton relay across the six- and five-membered ring is an interesting mechanistic feature. The reaction profile depicted in Figure 6.9 includes the relative energies of each stationary point (with respect to AICN + HCN + $2\text{H}_2\text{O}$, the 1:1:2 reactant complex). The first step is rate determining (as stated above). The subsequent steps have lower reaction barriers.

The mechanistic features with the ammonia-catalyzed pathway as shown in Figure 6.10 are different to that of the water-catalyzed pathway. This can be attributed to the difference in lone pair and hydrogen bonding properties of ammonia and water. The first three steps, that is, addition of HCN to AICN catalyzed by ammonia, syn-anti hydrogen transfer and rotation around C–N, are identical in both water and ammonia catalyzed mechanism. However, concerted six-membered ring closure and 1-4 H transfer takes place before 1-3 H transfer in case of ammonia-catalyzed mechanism. Figure 6.11 shows the geometries of the stationary points for the ammonia catalyzed pathway.

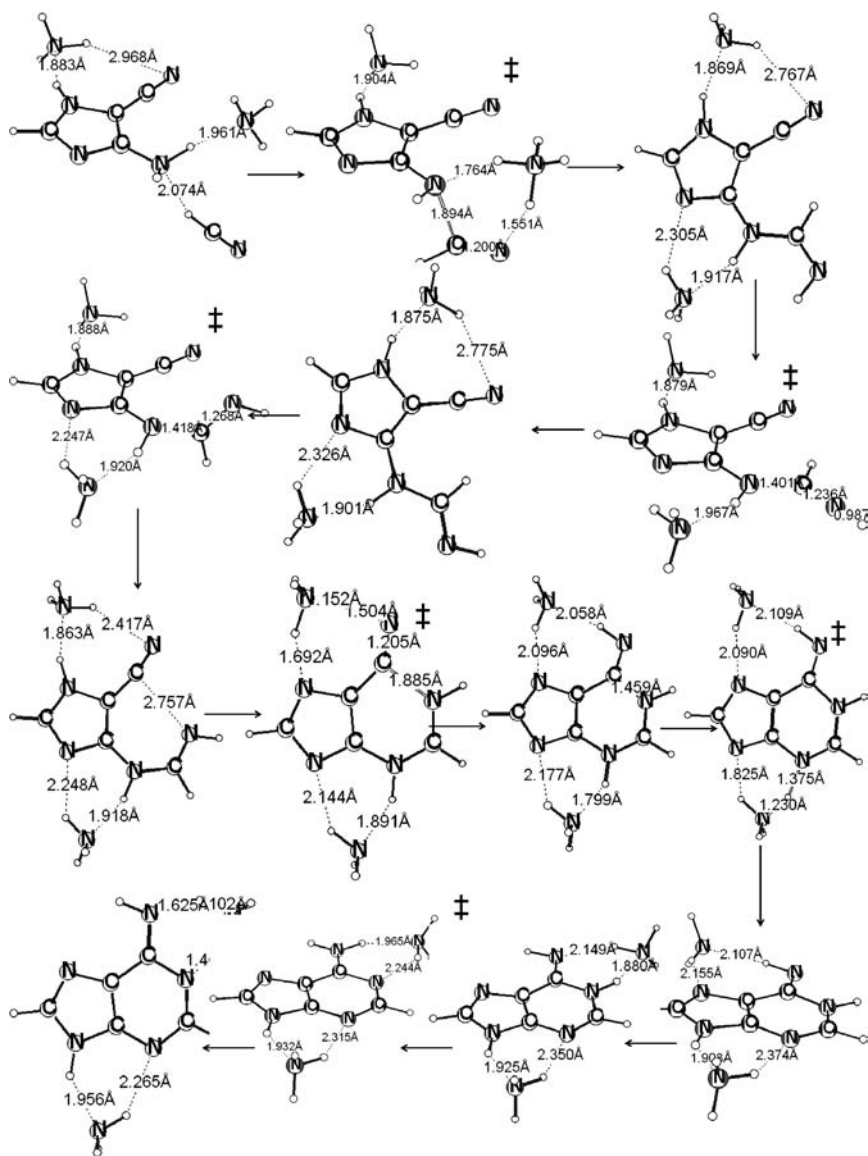


Figure 6.11 Stationary points for the ammonia (two explicit molecules) catalyzed pathway. Transition state structures are depicted by a ‡ sign.

To summarize, the reaction mechanism for adenine formation is mostly dominated by concerted C–N bond formation and 1,3 hydrogen shift (tautomerization). Specific water and ammonia molecules act as a catalyst for stabilizing the transition state and thus lower the barrier to tautomerization. The specific solvent molecule provides bridges that connect the donor and acceptor sites and thus relax the energy

required to bring these sites closer together prior to the proton transfer process. For the 1,3 hydrogen shift, the high classical barrier in the gas phase is due to the four-membered transition state and bending of NCN angle. The dihydrated system provides an optimal static condition and lowers the classical barrier for tautomerization.

We have reported a model study for “water (solvent) assisted proton transfer.” The “first model” is a neutral uncatalyzed mechanism in the gas phase; the second model is with explicit participation of a catalyst (H_2O or NH_3) in gas phase and the third model adds bulk solvation to simulate effects in aqueous solution. The results show that (i) explicit water/ammonia catalyzed mechanisms have lower reaction barriers than the uncatalyzed barriers and (ii) specific and bulk solvation contributions work together in facilitating the reaction. As there is no quantitative experimental data to match, a super-refined study is not mandatory. However, a more sophisticated treatment can be carried out eventually to model the medium effects. A QM/MM approach can be taken as described by Jorgensen and his coworker [55]. The energetics of the reacting systems can be described quantum mechanically with *ab initio* or DFT. The environment including solvent molecules can be represented using molecular mechanics and a sampling performed with Monte Carlo statistical mechanics.

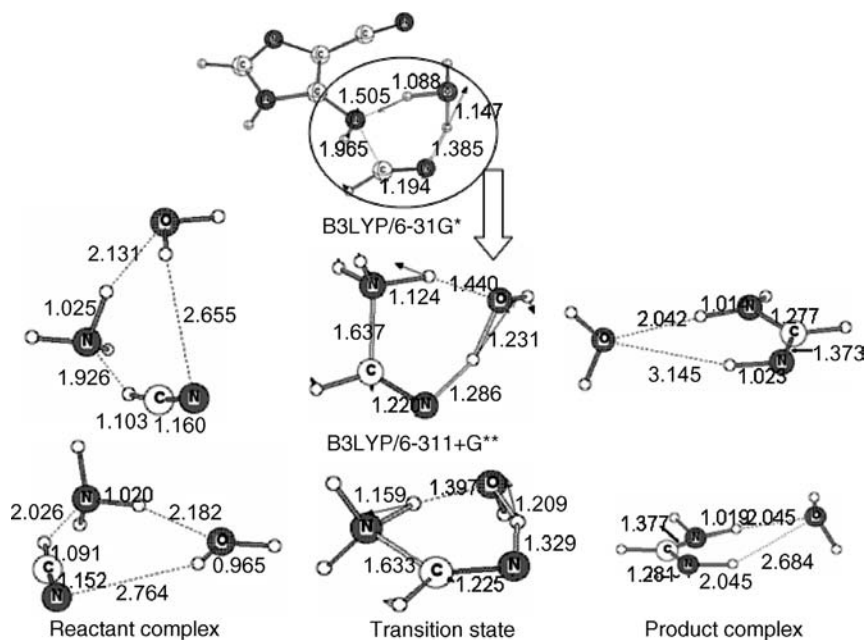
6.2.3.4 Validating the Methods Used for Computing Barrier Heights

Our computed reaction profile with low energy barriers reveals the feasibility of adenine formation from AICN under abiotic conditions. Are the reaction barriers reasonable? The gradient-corrected B3LYP was long regarded to be the best DFT method for structures and energies [56, 57]. However, B3LYP sometimes overestimates activation barriers and thereby underestimates the reaction barrier [56, 57]. We therefore recomputed the first step associated with highest barrier with MP2/6-311+G** and the gas phase reaction barrier turned out to be $43.9 \text{ kcal mol}^{-1}$ (with ZPVE), higher than that with B3LYP/6-311+G**. This is not surprising. Since MP2 is reported to give accurate geometries but overestimate barrier heights [58, 59]. Thus for standardization, we performed higher level CCSD (T) single-point computations on the simpler $\text{HCN} + \text{NH}_3 + \text{H}_2\text{O}$ system (Figure 6.12). This models the key $\text{AICN(b)} + \text{HCN} + \text{H}_2\text{O}$ rate-determining step, since it also involves the addition of an NH bond to HCN. The CCSD(T)/aug-cc-pVTZ single point reaction barrier is 3 kcal mol^{-1} higher than the B3LYP/6-311+G** result.

6.3

Conclusion

Our mechanistic investigation into the formation of adenine is a step towards a better understanding of the chemical origin of life. Our predictions are based on extensive computations of sequences of reaction steps along several possible mechanistic routes. We have deduced a plausible mechanism for formation of abiotic adenine



Level of theory/basis set	Reaction barrier(+ZPVE) (kcal mol ⁻¹)
B3LYP/6-31G*	31.3 (30.5)
B3LYP/6-311+G**	34.9 (34.6)
CCSD(T)/aug-ccpVDZ (single point starting from MP2/6-311+G** geometry)	38.4
CCSD(T)/aug-ccpVTZ (single point starting from MP2/6-311+G** geometry)	37.7

Figure 6.12 Geometries of reactant, product and transition state for the HCN + NH₃ + H₂O system (simulating the first HCN + RNH₂ addition step with the highest barrier) at the B3LYP level of theory with various basis sets are shown. Density functional theory (DFT) and CCSD(T) reaction barriers are compared.

starting from AICN (a tetramer of HCN) (Figure 6.13). To explain the existence of a relatively higher barrier for the key step, one has to keep in mind that, although adenine formation from five HCN is highly exothermic, it does require some activation. Rather drastic reaction conditions (e.g., heating up at 70 °C, or keeping a solution mixture for 25 years) were applied in abiotic synthesis of adenine. The reaction conditions provide us with a clue that there must be some steps associated with considerable reaction barriers. H₂O or NH₃ can act as a catalyst (as shown in our

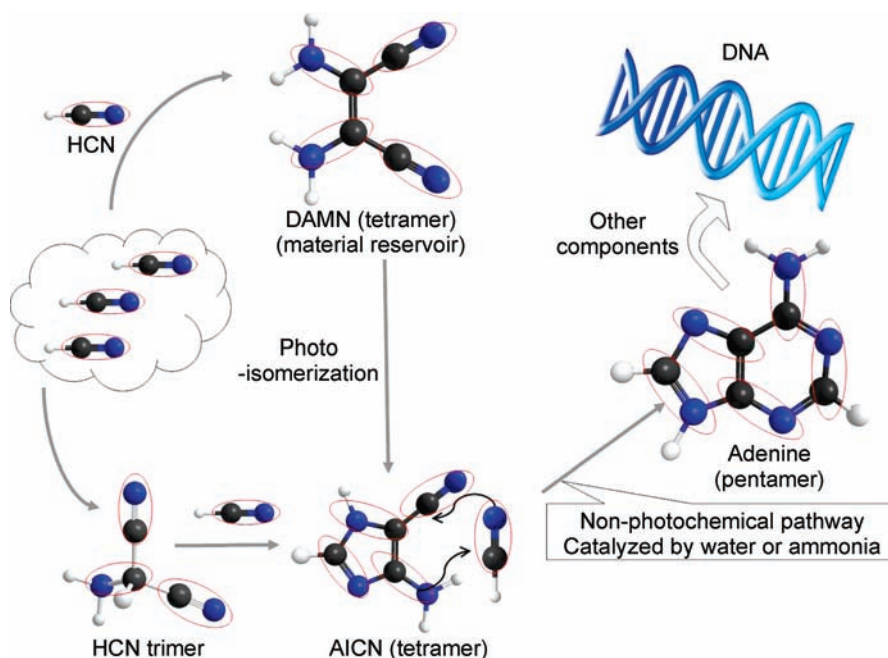


Figure 6.13 Summary of the step by step mechanism for the formation of adenine. The last crucial pentamer formation step is catalyzed by a specific solvent molecule (water/ammonia). CN building blocks are circled in red.

study) to lower the barriers considerably. Moreover, prebiotic conditions (bulk solvation, for instance) is implicitly considered in calculations using the quantum chemical techniques, which lowers the barrier of the rate-determining step further.

An alternate pathway based on the experimental isolation of 2- and 8-cyanoadenine or adenine-8-carboxamide as adenine precursor suggests a further complex mechanism involving hexamer and heptamers of HCN [60]. When we were refining our model over three years, another team published parallel research that also sheds light on the prebiotic origins of adenine, though from a different perspective [49]. Rainer Glaser and colleagues from the University of Missouri used a theoretical model to predict mechanisms for the production of adenine under photochemical conditions in space. Their study reveals that adenine formation from a HCN pentamer precursor is not viable thermochemically either by proton-catalyzed or by uncatalyzed cyclization, instead photoactivation of one of the cyclization paths may lead to the imino form of adenine. If, indeed, the production of adenine began in space and came to Earth on asteroids, this could possibly open the idea of life elsewhere to more serious avenues of study. Our study, however, offers clues to a more *earth-bound* solution to this important compound formed under prebiotic conditions on early earth.

The neutral water or ammonia catalyzed mechanism proposed in our study may well have been a major route for the formation of adenine on primitive earth.

However, detailed calculations on reactions of this kind are needed before a full picture emerges. Mechanisms modeling the formation of other remaining nucleic acid bases and biologically relevant molecules extraterrestrially, under more restricted conditions, are further challenges.

Acknowledgment

We are thankful to Professor Thomas Tidwell of University of Toronto for calling our attention to this problem and K. Najafian, our collaborator in the original work, for carrying out extensive computations on intermediates and transition states that might be involved in possible anionic and radical mediated pathways. This work was supported by National Science Foundation Grants CHE-0209857 and CHE-0716718 and the University of Georgia.

References

- 1 Miller, S.L. and Orgel, L.E. (eds) (1974) *The Origins of Life of Earth*, Prentice-Hall, Englewood Cliffs, NJ.
- 2 Schopf, J.W. (ed.) (1983) *Earth's Earliest Biosphere: Its Origin and Evolution*, Princeton University Press, Princeton, NJ.
- 3 Wöhler, F. (1828) *Annal. Phys. Chem. (Leipzig)*, **88**, 253–256.
- 4 Butlerow, A. (1861) *Compt. Redn. Acad. Sci.*, **53**, 145.
- 5 Glocker, G. and Lind, S. (1939) *The Electrochemistry of Gases and Other Dielectrics*, John Wiley & Sons, Inc., New York.
- 6 Löb, W. (1913) *Chem. Ber.*, **46**, 684.
- 7 Baudish, O. (1913) *Z. Angew. Chem.*, **26**, 612.
- 8 Miller, S.L. (1953) *Science*, **117**, 528.
- 9 Miller, S.L. and Urey, H.C. (1959) *Science*, **130**, 245.
- 10 Oró, J. (1961) *Nature*, **191**, 1193.
- 11 Oró, J. and Kimball, A.P. (1961) *Arch. Biochem. Biophys.*, **94**, 217.
- 12 Oró, J. and Kimball, A.P. (1962) *Arch. Biochem. Biophys.*, **96**, 293.
- 13 Ferris, J.P., Sanchez, R.A., and Orgel, L.E. (1968) *J. Mol. Biol.*, **33**, 693.
- 14 Oró, J. (1960) *Biochem. Biophys. Res. Commun.*, **2**, 407.
- 15 Levy, M., Miller, S.L., and Oró, J. (1999) *J. Mol. Evol.*, **49**, 165.
- 16 Voet, A.B. and Schwartz, A.W. (1982) *Origins Life*, **12**, 45.
- 17 Lowe, C.U., Rees, M.W., and Markham, F.R.S. (1963) *Nature*, **199**, 219.
- 18 Ferris, J.P. and Joshi, P.C. (1978) *Science*, **201**, 361.
- 19 Hulshof, J. and Ponnampereuma, C. (1976) *Origins Life*, **197**, 224.
- 20 McCord, T.B., Hansen, G.B., Fanale, F.P., Carlson, R.W., Matson, D.L., Johnson, T.V., Smythe, W.D., Crowley, J.K., Martin, P.D., Ocampo, A., Hibbitts, C.A., and Granahan, J.C. (1998) *Science*, **278**, 1242.
- 21 Oró, J. and Kamat, S.S. (1961) *Nature*, **190**, 442.
- 22 Matthews, C.N. and Moser, R.E. (1967) *Nature*, **215**, 1230.
- 23 Ferris, J.P., Wos, J.D., Nooner, D.W., and Oró, J. (1974) *J. Mol. Evol.*, **3**, 225.
- 24 Shen, C., Yang, L., Miller, S.L., and Oró, J. (1990) *J. Mol. Evol.*, **31**, 167.
- 25 Johnson, P., Cleaves, H.J., Dworkin, J.P., Glavin, D.P., Lazcano, A., and Bada, J.L. (2008) *Science*, **322**, 404.
- 26 Abelson, P.H. (1966) *Proc. Natl. Acad. Sci. U.S.A.*, **55**, 1365.
- 27 Hanel, R., Conrath, B., Flasar, F.M., Kunde, V., Maguire, W., Pearl, J., Pirraglia, J., Samuelson, R., Herath, L., Allison, M., Cruikshank, D., Gautier, D., Gierasch, P., Horn, L., Koppany, R., and Ponnampereuma, C. (1981) *Science*, **212**, 192.
- 28 Raulin, F. and Frere, J.C. (1989) *Br. Interplanetary Soc.*, **42**, 411.

- 29 Owen, T. (1982) *J. Mol. Evol.*, **18**, 150.
- 30 Levy, M., Miller, S., Brinton, K., and Bada, J. (2000) *Icarus*, **145**, 609–613.
- 31 Reynolds, R.T., Squyres, S.W., Colburn, D.S., and McKay, C.P. (1983) *Icarus*, **56**, 246.
- 32 Chyba, C.F. (1997) *Nature*, **385**, 201.
- 33 McCord, T.B., Carlson, R.W., Smythe, W.D., Hansen, G.B., Clark, R.N., Hibbitts, C.A., Fanale, F.P., Granahan, J.C., Segura, M., Matson, D.L., Johnson, T.V., and Martin, P.D. (1997) *Science*, **278**, 271.
- 34 Matthews, C.N. (1992) *Orig. Life Evol. Biosphere*, **21**, 421.
- 35 Irvine, W.M. (1999) *Space Sci. Rev.*, **90**, 203.
- 36 Huebner, W.F., Snyder, L.E., and Buhl, D. (1974) *Icarus*, **23**, 580.
- 37 Magee-Sauer, K., Mumma, M.J., DiSanti, M.A., Russo, N.D., and Rettig, T.W. (1999) *Icarus*, **142**, 498.
- 38 Hidayat, T., Marten, A., Bezard, B., Gautier, D., Owen, T., Matthews, H.E., and Paubert, G. (1998) *Icarus*, **126**, 170.
- 39 Shuman, R.F., Shearin, W.E., and Tull, R.J. (1979) *J. Org. Chem.*, **44**, 4532.
- 40 Voet, A.B. and Schwartz, A.W. (1983) *Bioorg. Chem.*, **12**, 8.
- 41 Ferris, J.P. and Orgel, L.E. (1966) *J. Am. Chem. Soc.*, **88**, 3829.
- 42 Ferris, J.P. and Orgel, L.E. (1965) *J. Am. Chem. Soc.*, **87**, 4976.
- 43 Ferris, J.P. and Orgel, L.E. (1966) *J. Am. Chem. Soc.*, **88**, 1074.
- 44 Wakamatsu, H., Yamada, Y., Saito, T., Kumashiro, I., and Takenishi, T. (1966) *J. Org. Chem.*, **31**, 2035.
- 45 Ferris, J.P., Joshi, P.C., Edelson, E.H., and Lawless, J.G. (1978) *J. Mol. Evol.*, **11**, 293.
- 46 Kikuchi, O., Watanabe, T., Satoh, Y., and Inadomi, Y. (2000) *J. Mol. Struct.*, **507**, 53.
- 47 Tureček, F. and Chen, X. (2005) *J. Am. Soc. Mass Spectrom.*, **16**, 1713–1726.
- 48 Roy, D., Najafian, K., and Schleyer, P.v.R. (2007) *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 17272.
- 49 Glaser, R., Hodgen, B., Farrelly, D., and McKee, E. (2007) *Astrobiology*, **7**, 455.
- 50 Frisch, M.J., Trucks, G.W., Schlegel, H.B. et al. (1998) Gaussian 98 (Revision A.9), Gaussian, Inc., Pittsburgh, PA.
- 51 Aida, M., Yamataka, H., and Dupuis, M. (1998) *Chem. Phys. Lett.*, **292**, 474.
- 52 Miertus, S., Scrocco, E., and Tomasi, J. (1981) *Chem. Phys.*, **55**, 117.
- 53 Yamabe, S., Tsuchida, N., and Miyajima, K. (2004) *J. Phys. Chem. A*, **108**, 2750.
- 54 Kiruba, G.S.M. and Wong, M.W. (2003) *J. Org. Chem.*, **68**, 2874.
- 55 Jorgensen, W.L. and Alexandrova, A.N. (2007) *J. Phys. Chem. B*, **111**, 720.
- 56 Barone, V. (1994) *Chem. Phys. Lett.*, **226**, 392.
- 57 Martell, J.M., Goddard, J.D., and Eriksson, L.A. (1997) *J. Phys. Chem. A*, **101**, 1927.
- 58 Wiberg, K.B. and Ochterski, J.W. (1997) *J. Comput. Chem.*, **18**, 108.
- 59 Gonzales-Garcia, N., Gonzales-Lafont, À., and Lluch, J.M. (2005) *J. Comput. Chem.*, **26**, 569.
- 60 Borquez, E., Cleaves, H.J., Lazcano, A., and Miller, S.L. (2005) *Origins Life*, **35**, 79.

7

Hydrogen Bonding and Proton Transfer in ionized DNA Base Pairs, Amino Acids and Peptides

Luis Rodríguez-Santiago, Marc Noguera, Joan Bertran, and Mariona Sodupe

7.1

Introduction

Radical cations exhibit a rich and varied chemistry [1]. On one hand, ionization creates a positive charge that introduces large electrostatic effects whereas, on the other hand, the deficit of one electron modifies, directly or indirectly, the strength of covalent bonds. Such effects result in unusual structures and reactivities, which are often significantly different from those of their neutral precursors. Some of the systems that are strongly affected by ionization are those that contain hydrogen bonds, since ionization induces important changes in the acidity and basicity of the centers involved in the hydrogen bond. In general, for intermolecular H-bonded systems, when ionization is produced on the proton donor molecule the hydrogen bond is strengthened and proton transfer reactions occur readily [2–4], in many cases spontaneously, leading to so-called distonic radical cations [5]. This is due to an increase of the acidity of the donor monomer. In contrast, if ionization is produced on the acceptor monomer the hydrogen bond is weakened and the proton transfer reaction becomes very unfavorable, in such a way that other rearrangements are frequently observed. Intramolecular H-bonded systems are more complex than intermolecular ones because ionization simultaneously increases the acidity and decreases the basicity of the donor and acceptors groups, respectively. Therefore, it is difficult to establish *a priori* how this interaction and subsequent reactivity will be affected by oxidation.

Hydrogen bond interactions play a crucial role in biology. They are responsible for the structure and stability of biological macromolecules such as DNA, RNA or proteins, and are involved in many molecular-recognition processes [6]. Analyzing the changes on hydrogen bonding and on proton transfer reactions on the basic constituents of biological macromolecules upon ionization is, therefore, very important as a first step to understanding the effects of ionizing radiation and oxidative damage on biological systems. In this chapter we review the contributions

of our group on the effects of ionization on proton transfer processes in hydrogen bonded systems of biological interest. First, we consider the intermolecular proton transfer reaction in ionized DNA base pairs. Second, we will discuss the effects of ionization on the conformational preferences of different amino acids, paying special attention to the role of the side chain. Finally, the changes induced on a peptide bond containing system and on the Ramachandran plots of model peptides are analyzed.

7.2

Methodological Aspects

Radical cations are open shell systems and, thus, they present specific methodological problems. Within the Hartree–Fock (HF) framework two strategies are possible: the unrestricted UHF and the restricted open shell ROHF. The UHF formalism has the advantage that it introduces spin polarization but the inconvenience that the UHF Slater determinant is no longer an eigenfunction of the S^2 operator, which leads to the so-called spin contamination. In contrast, the restricted open shell ROHF method has the advantage that it avoids spin contamination but at the cost of not including the spin polarization of the inner shells. Post Hartree–Fock methods correct the limitations of both formalisms. However, some caution should be taken if the UHF reference wavefunction is highly contaminated, particularly with methods such as UMP2, since in these cases the Møller–Plesset perturbation expansion converges slowly [7]. In these situations the restricted RCCSD(T) method, which extensively includes electron correlation and is based on a restricted reference wavefunction, is more appropriate. Unfortunately, this method is computationally very demanding and can only be applied to relatively small systems. For large systems other less costly methods are required. In this context, density functional theory (DFT) methods appear to be an attractive alternative. For some open shell systems, however, DFT methods with a small percentage of exact exchange have been shown to overstabilize structures with a too delocalized electron hole [8]. Therefore, it is always convenient to confirm the reliability of the chosen functional by performing RCCSD(T) calculations for model systems with a similar chemistry. This has been carried out for all systems considered in the present chapter. The results obtained indicate that in all situations, except those that exhibit three-electron bonds, the B3LYP functional [9, 10], with a 20% of exact exchange, performs reasonably well to describe the relative stability and proton transfer processes of these radical cations. For three-electron bond systems, for which the spin density is highly delocalized, a functional with a larger percentage of exact exchange such as the BHLYP [10, 11] or the more recently developed meta-hybrid MPWB1K functional [12] are more accurate. The results presented in the following sections correspond to either B3LYP or MPWB1K density functional methods, with basis sets of at least double-zeta plus diffuse and polarization functions. In some cases single point CCSD(T) calculations will also be reported.

7.3

Ionization of DNA Base Pairs

In pioneering work in 1963, Löwdin [13, 14] proposed that intermolecular proton transfer processes in DNA could alter the hydrogen-bonding patterns and ultimately lead, through base-mispairing, to spontaneous mutations. Two cases were distinguished. The first referred to pairs of equally charged bases. In this case, concerted double proton transfer reactions would take place in order not to create charge separation. The second referred to pairs with unequally charged bases generated through the introduction of a negative or a positive charge in one of the bases. Under these conditions the probability of single proton transfer reactions from the more positively charged moiety to the more negative charged one would be greatly increased. Since then many theoretical studies have been devoted to check Löwdin's hypothesis.

Initial studies [15–21] considered the double proton transfer process on the ground state of neutral pairs. However, because of the size of AT and GC base pairs, low computational levels of theory were used. In addition, these studies, using semiempirical [15–18] and *ab initio* methods [19–21], were performed using fixed geometries for the monomers during the proton transfer processes. Consequently, both single and double proton transfer reactions were found to be too unfavorable. More recent studies, in which geometries were fully optimized, found smaller energy barriers [22–24]. Much of the discussion focused on whether the mechanism of the double proton transfer reaction is concerted or stepwise, through a single proton transferred ion-pair like intermediate. These studies showed that electron correlation is essential to properly describe the topology of the potential energy surface (PES); the single proton transfer intermediate located at the HF level [25] being an artifact of the method. Our results [26, 27] with the B3LYP density functional method agree with this result. For neutral systems, only the product derived from the double proton transfer reaction is found as a minimum on the PES, its relative energy with respect to the reactant species at the B3LYP/6-311 + + G(d,p) level being 10.2 and 13.6 kcal mol⁻¹ (1 kcal = 4.184 kJ) for GC and AT, respectively, and the energy barriers 15.8 and 13.7 kcal mol⁻¹. Similar results have been obtained by other authors [28–31]. All studies agree that the double proton transfer reaction is more favorable than the single proton transfer because in the former the electroneutrality is maintained, whereas the single transfer process implies a charge separation. Consequently, the ion-pair complex resulting from the single proton transfer reaction is not stable in the gas phase although it may be significantly stabilized in a high dielectric medium (water) [32].

Less attention has been paid to proton transfer processes in base pairs with unequally charged bases. A particularly interesting case is that in which an extra positive charge arises from ionizing due to its connection to DNA oxidative damage and charge migration and conductivity in DNA. Indeed, interstrand proton transfer can constitute a very powerful stop to the migration of holes in the direction of the DNA helix via the stacked bases [33]. A few theoretical papers [34–39] have studied proton transfer reactions in DNA base pair radical cations. In this chapter we present the theoretical results obtained on ionized Watson–Crick base pairs by means of the B3LYP method with a 6-311 + + G(d,p) basis set.

Guanine-cytosine (GC) and adenine-thymine (AT) Watson–Crick base pairs are systems with multiple hydrogen bonds (three and two, respectively), with the interacting monomers acting both as proton donor and proton acceptor simultaneously. Such systems are more complex than ones with a single hydrogen bond because once the system is ionized many different proton transfer reactions can take place; that is, the proton transfer reaction can either be produced from any of the two monomers, and the existence of more than one hydrogen bond opens up the possibility of having not only single proton transfer reactions but also double proton transfer processes. Therefore, before examining the proton transfer reactions it is convenient to analyze the influence of ionization on the structure and interaction energies of DNA base pairs.

7.3.1

Equilibrium Geometries and Dimerization Energies

Figure 7.1 shows the optimized hydrogen bond distances of the neutral and cationic species of GC and AT base pairs as well as the dimerization energies. The H-bonds distances change dramatically upon ionization. This can be understood by considering the nature of ionized species. Natural population analysis indicates that

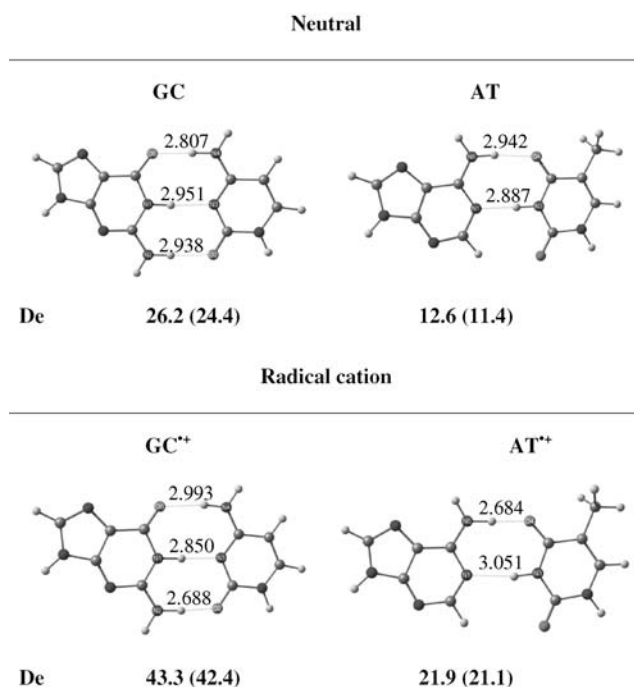


Figure 7.1 B3LYP/6-311++G(d,p) hydrogen bond distances (Å) and dimerization energies (kcal mol⁻¹) for neutral and ionized GC and AT systems. Counterpoise corrected values are given in parentheses.

ionization of GC and AT is mainly localized at the guanine and adenine monomers, respectively [35]. For GC, the charge on guanine is 0.85 and the spin density 1.0, whereas for AT the charge on adenine is 0.84 and the spin density 0.90. This was to be expected considering that guanine and adenine have a lower ionization potential than cytosine and thymine, respectively. Since guanine and adenine are the two monomers that lose the electron and, thus, become more acidic, those hydrogen bonds in which these two monomers act as proton donor become stronger in the ionized system. This implies a shortening of the distance between the two heavy atoms and a lengthening of the H–X bond involved. In contrast, those H-bonds in which guanine and adenine act as proton acceptor become weaker. Namely, the N_1-N_3 and N_2-O_2 H-bond distances of GC decrease after ionization while the O_6-N_4 distance increases. For AT, the N_6-O_4 is the bond that becomes shorter while the N_1-N_3 becomes longer.

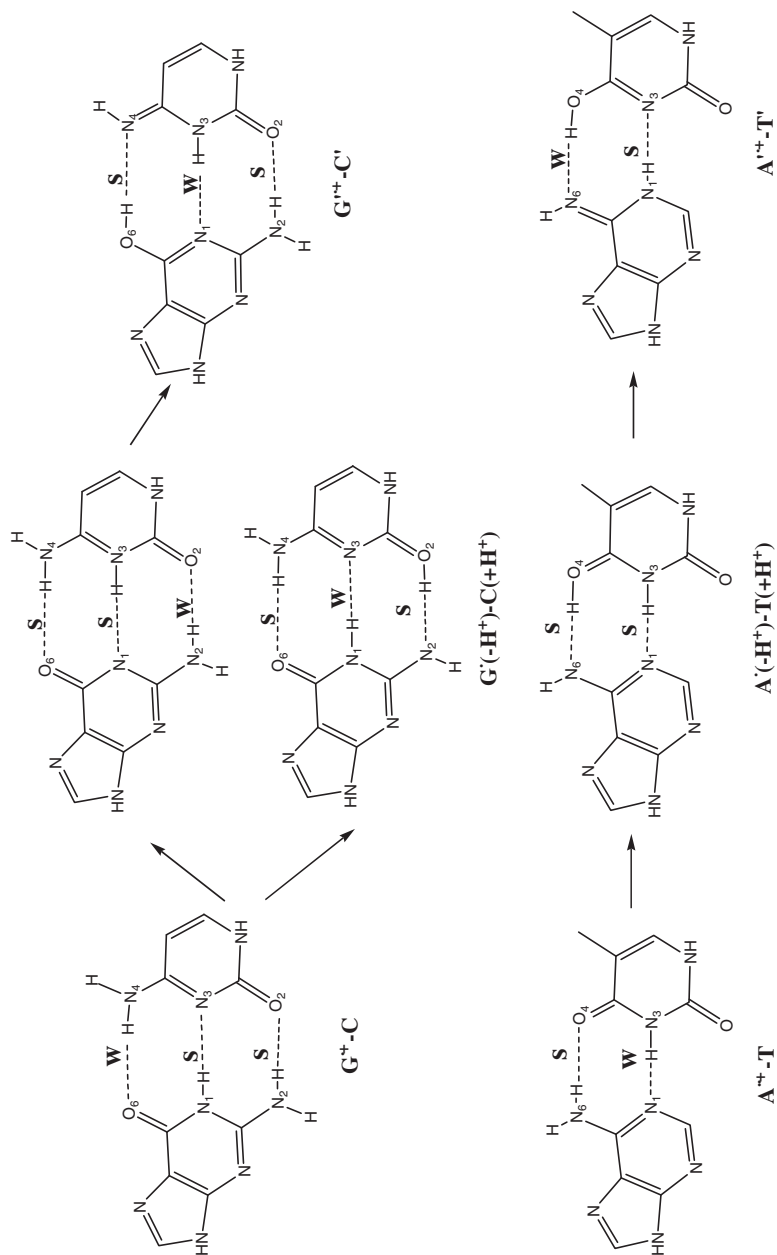
Figure 7.1 shows that ionization produces a significant increase of the binding energy of GC and AT due to the enhanced electrostatic interaction in the positively charged systems. Present values for interaction energies are somewhat smaller than those previously published [34, 35], particularly the one non-corrected for BSSE, because the basis set used in the present calculations is larger. For GC, the binding energy increases by about $17\text{--}18\text{ kcal mol}^{-1}$ and for AT by about $9\text{--}10\text{ kcal mol}^{-1}$. This increase is not due to an equal strengthening of the three hydrogen bonds of GC or the two hydrogen bonds of AC since, as mentioned, ionization strengthens or weakens the hydrogen bonds depending on whether the ionized monomer acts as proton donor or proton acceptor.

7.3.2

Single and Double Proton Transfer Reactions

Scheme 7.1 shows the single and double proton transfer processes considered [35]. The nature of the H-bonds in each structure has been indicated with letters s (strong) and w (weak).

As mentioned, ionized GC presents one weak and two neighboring strong H-bonds, that is, a w-s-s situation. Any of the two strong H-bonds can be involved in the first single proton transfer reaction, leading to a distonic radical cation [5], since the positive charge lies on protonated cytosine and the radical character on deprotonated guanine. The transfer from N_1 to N_3 leads to an s-s-w situation with two neighboring strong H-bonds, while transfer from N_2 to O_2 produces the alternated s-w-s pattern. In the first case, the two strong hydrogen bonds can benefit from the enhanced electrostatic interaction by decreasing simultaneously the N_1-N_3 and O_6-N_4 distances, while the N_2-O_2 distance increases to reduce repulsion. These geometry changes can take place through a relative movement of the two monomers that approaches the O_6-N_4 terminal bond and separates the other N_2-O_2 . Owing to the rigidity of the monomers the central H-bond distance does not decrease as much as the terminal one. The optimized H-bond distances agree with what is expected. Finally, notably, the resulting structure presents two neighboring strong hydrogen bonds, as in the reactant. In the second case, we have an s-w-s pattern, which does not



Scheme 7.1

allow a geometrical compromise that benefits strong short H-bonds, without introducing the central H-bond into a repulsive region. Consequently, this structure has not been localized as a minimum on the PES. The double proton transfer reaction leads to $G'^+ + C'$, which also has an alternate s-w-s situation. Thus, unsurprisingly, any attempt to localize this energy minimum collapsed to the single proton transferred structure.

For the ionized AT, we have one strong and one weak H-bond, s-w pattern, before any proton transfer is produced. After the single proton transfer, the positive charge moves to the protonated thymine and, thus, both hydrogen bonds become strong (s-s) and we obtain a distonic radical cation with the positive charge at the thymine moiety and the radical character at the adenine one. The double proton transfer reaction leads to a situation similar to the initial one, but with the two H-bonds reversed, w-s pattern. The optimized H-bond distances shown in Figure 7.2 confirm the expected changes. It is worth noting the particularly small distance of 2.573 Å, obtained for N_6-O_4 in the single proton transferred structure. This distance is much smaller than the sum of van der Waals radii typically observed in strong, short hydrogen bonds.

The relative energies of the single and double proton transferred structures with respect to the initial ionized structure are given in Figure 7.2. The energy of the non-transferred asymptote and the energies of the single and double proton transfer asymptotes have also been included. For ionized GC there is not a minimum corresponding to the double proton transfer reaction, while for ionized AT the double proton transfer reaction displays a shallow minimum, the transition state being close in energy.

Overall, these energy profiles indicate that two factors determine the stability of the different species: (i) the relative stability of the asymptotes from which they derive, which is related to the relative proton affinity of the centers involved in the proton transfer, and (ii) the strength of the interaction leading to the formation of the dimer. This second factor depends on the number and sequence of strong and weak hydrogen bonds formed. For example, for ionized GC, the non-proton transfer, the single proton transfer and the double proton transfer asymptotes lie more or less at the same energy (43.3, 44.6 and 44.9 kcal mol⁻¹, respectively). However, only the non-proton transferred and single proton transferred structures are stable. Furthermore, Figure 7.2 shows that the energy difference between the non-transferred and single transferred structures is the same as that found for the corresponding asymptotes, thus indicating that the strength of the H-bond interactions is very similar in the two structures. This is not surprising considering that both structures show a pattern with two neighboring strong H-bonds, whereas the double proton transferred structure presents an unfavorable s-w-s situation.

Similar arguments hold for the AT cationic system. The single proton transfer asymptote (39.6 kcal mol⁻¹) lies much higher in energy than the non-proton transfer one (22.0 kcal mol⁻¹). However, the derived proton transferred dimer is almost degenerate with the non-proton transferred one. This is because in the former we have one strong and one weak H-bond (s-w) while in the latter the two H-bonds are strong (s-s) and, as mentioned, one of them is very strong. Thus, the larger H-bond interaction compensates for the relative energy of the two asymptotes. The double

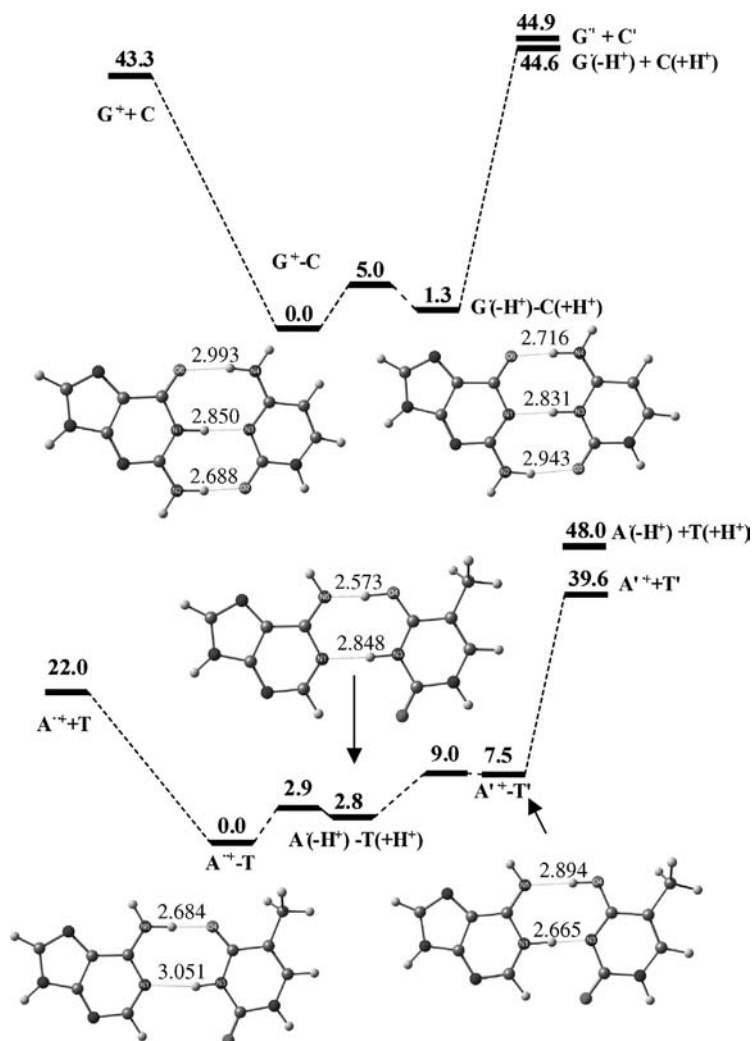


Figure 7.2 Energy profile (kcal mol⁻¹) for single and double proton transfer processes on GC^{•+} and AT^{•+} systems. All distances are given in Ångström.

proton transfer structure has a similar pattern, w-s, to the initial structure, s-w, but lies higher because the asymptote from which it derives is also less stable. The energy difference is, however, reduced from 17.6 to 7.5 kcal mol⁻¹, because the H-bonds in A^{•+}+T⁻ are slightly shorter than in the initial reactant and the hydrogen bond between nitrogen atoms is stronger than the one between nitrogen and oxygen atoms.

Both for GC and AT radical cations the single proton transfer reaction presents a low energy barrier (5 and 2.9 kcal mol⁻¹, respectively). In contrast, the double proton transfer process is not favorable. Notably, an experimental gas phase study using enhanced multiphoton ionization (REMPI) has found evidence for proton transfer in

GC pair [40]. In summary, the behavior of the radical cation species [35] is very different to that observed for the neutral base pairs, for which the single proton transfer reaction is found to be very unfavorable. This is due to the increased acidity of the ionized monomer and to the fact that the proton transfer does not imply a creation of charges but a transfer of a positive charge.

7.4

Ionization of Amino Acids

Amino acids are the basic building blocks of proteins and, thus, analyzing the structure and reactivity of their radical cations represents the first step to understanding the oxidative damage in proteins caused by ionizing radiation or oxidative agents. Moreover, the study of their structural properties and derived decomposition reactions is also important in the field of mass spectrometry since radical cations of some oligopeptides have been generated in the gas phase by collision-induced dissociation of $[\text{CuII}(\text{dien})\text{M}]^{2+}$ complex ions [41–43]. Their dissociation behavior is very rich and differs considerably from that of protonated peptides, which make them very attractive for peptide sequencing. Consequently, in recent years, the properties of different amino acid and derived radicals have attracted considerable attention, both from an experimental and theoretical point of view [41–60].

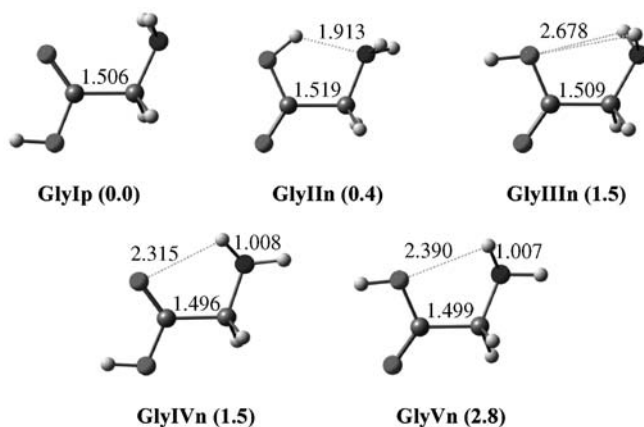
Amino acids present intramolecular hydrogen bonds that determine their conformational preferences. These hydrogen bonds can be largely modified by ionization, which leads to significant geometrical rearrangements. As will be shown, these rearrangements are highly dependent on the nature of the electron hole generated. Therefore, to establish the basic trends of the processes induced by ionization we first analyze the conformational behavior of neutral species and then, for each low-lying stable conformer, we explore the ionized species. We started with the simplest amino acids, glycine and alanine, and then extended our study to other seven amino acids of different nature (basic, acidic and aromatic) [55]. Notably, for the larger amino acids, a large number of conformations exist due to the presence of many single bond rotamers. In these cases, with a high conformational complexity, the structures to be calculated at the DFT level are chosen from a previous analysis using a Monte Carlo multiple minimum (MCMM) conformational search [61] with the MMFF94s force field [62]. Geometry optimizations have been performed using different functionals such as the hybrid B3LYP and hybrid-meta MPWB1K functionals with the 6-31++G(d,p) basis set. In addition, for some systems, calculations with larger basis sets and single-point calculations at the CCSD(T) level have also been carried out [46, 55].

7.4.1

Structural Features of Neutral and Radical Cation Amino Acids

Previous accurate theoretical studies have identified eight minimum energy conformers of neutral glycine [63]. Among them, five conformers present relative

Neutral



Radical cation

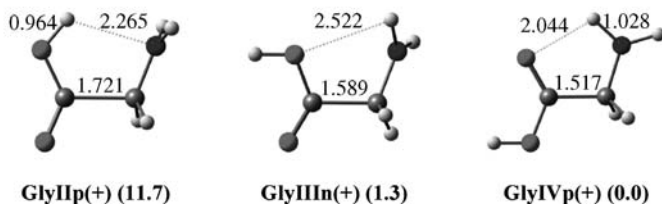


Figure 7.3 Optimized geometries and relative energies for the lowest-energy conformers of the neutral and radical cation species of glycine, at the MPWB1K/6-31 + + G(d,p) level of theory. Distances are in ångströms and energies in kcal mol⁻¹.

energies in a range smaller than 1000 cm⁻¹ (2.86 kcal mol⁻¹), with the relative energies of the three remaining conformers being larger than 4.7 kcal mol⁻¹ with respect to the ground state structure. Nevertheless, upon ionization of these neutral conformers, only three stable structures are found for the glycine radical cation. Figure 7.3 shows optimized geometries and relative energies. The notation used has been taken from Reference [63]. Structures designated with the letter p show C_s symmetry with planar heavy atom arrangement and those designated with n show C₁ symmetry. For the radical cation species we have added a (+) symbol.

First of all, it can be noted that structures GlyI(+) and GlyV(+) (n or p) are not found since ionization of GlyIp and GlyVn collapses to structure GlyIIIn(+). For all the considered conformers of glycine ionization is localized at the NH₂ group and, consequently, the hydrogen bonds that involve this group are modified; that is, the amino group becomes more planar, increases its acidity, and the intramolecular

hydrogen bonds in which the NH_2 acts as proton donor are strengthened. For this reason structure GlyIV(+) is largely stabilized, becoming the most stable structure for the radical cation. In contrast, structure GlyIIp(+) in which NH_2 acts as proton acceptor becomes the most unstable one due to the decrease of basicity of the amino group upon ionization [46, 55].

We have carried out a similar analysis for the neutral and radical cation conformers of the second simplest amino acid alanine (not shown in detail here) and the obtained results are very similar [55]: although for neutral alanine the number of conformers is even larger than for glycine [64], only structures of types II(+), III(+), and IV(+) are found to be minima on the PES upon ionization [55]. In all cases the ionization is also localized at the amino group, which becomes almost planar. Among these structures, the most stable one corresponds to the AlaIV(+) isomer shown in Figure 7.4, where the hydrogen bond has been strengthened with respect to the neutral form of this conformer, as in the case of glycine.

Figure 7.4 also shows the most stable structure of the radical cations of amino acids that contain side chains with acidic and basic groups such as Ser, Cys, Asp and Gln. In all cases ionization mainly takes place at the backbone of the amino acid, as in the case of Gly and Ala and, thus, the structure of type IV(+) is the lowest energy conformer. Notably, for Cys, Asp and Gln the most stable structure presents a two-center/three-electron hemibond interaction between the NH_2 group and another group of the side chain, SH for Cys and CO for Asp and Gln. For Ser, however, we have not been able to locate a conformation with such an interaction with the side chain, in agreement with the fact that the hydroxyl group prefers to establish hydrogen bond interactions than 2-center-3-electron hemibonds [2]. For structures that show the hemibond interac-

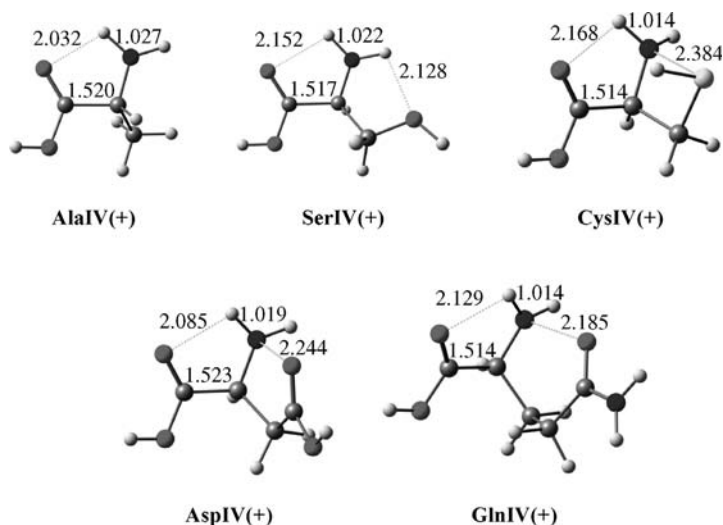


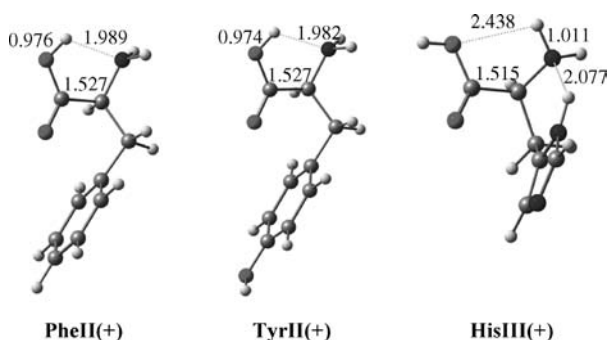
Figure 7.4 Optimized geometries for the lowest-energy conformers of Ala, Ser, Cys, Asp and Gln radical cations, at the MPWB1K/6-31G++(d,p) level of theory. Distances are in ångströms.

Table 7.1 MPWB1K/6-31 + + G(d,p) charge (spin density) from natural population analysis for the lowest energy conformer of Gly, Ala, Ser, Cys, Asp, Gln, Phe, Tyr and His radical cations.

Structure	NH ₂	COOH	R	CH
GlyIVp(+)	0.64 (0.90)	0.12 (0.00)	0.34 (0.06)	−0.10 (0.04)
AlaIV(+)	0.62 (0.88)	0.11 (0.00)	0.17 (0.08)	0.10 (0.04)
SerIV(+)	0.63 (0.87)	0.12 (0.01)	0.16 (0.06)	0.10 (0.06)
CysIV(+)	0.19 (0.40)	0.08 (0.00)	0.62 (0.60)	0.10 (0.00)
AspIV(+)	0.51 (0.75)	0.10 (0.00)	0.27 (0.24)	0.11 (0.01)
GlnIV(+)	0.46 (0.69)	0.08 (0.00)	0.36 (0.31)	0.10 (0.00)
PheII(+)	−0.06 (0.03)	0.23 (0.20)	0.72 (0.72)	0.11 (0.03)
TyrII(+)	−0.07 (0.02)	0.14 (0.11)	0.83 (0.84)	0.10 (0.03)
HisIII(+)	−0.10 (0.00)	0.08 (0.03)	0.89 (0.96)	0.12 (0.01)

tion, natural population analysis indicates that the spin density is delocalized between the two interacting groups (Table 7.1).

Figure 7.5 shows the low-lying structure of Phe, Tyr and His radical cations; that is, for amino acids that contain an aromatic side chain. In these cases ionization mainly takes place at the side chain, in agreement with the spin density values shown in Table 7.1 and the nature of the open shell orbital. In fact, Figure 7.6 shows the open shell orbital of GlyIVp(+), which is mainly centered at the NH₂ group of the backbone, and the open shell orbital of HisIII(+), mainly centered at the side chain. As a consequence of the different ionization mechanisms for these aromatic amino acids, for which ionization of the side-chain prevails over ionization of the NH₂ group, the structure of type IV(+) is no longer the most stable. In fact, the most stable structure for His is a distorted III(+) structure in which the NH group of the imidazole ring forms a hydrogen bond with the NH₂ group. This imidazole NH group is more acidic due to the ionization of the side-chain. In contrast, for Phe and Tyr, structures derived from II(+) become the ground state structures, which shows the importance of the side chain nature in the effects of ionization.

**Figure 7.5** Optimized geometries for the lowest-energy conformers of Phe, Tyr and His radical cation, at the MPWB1K/6-31 + + G(d,p) level of theory. Distances are in ångströms.

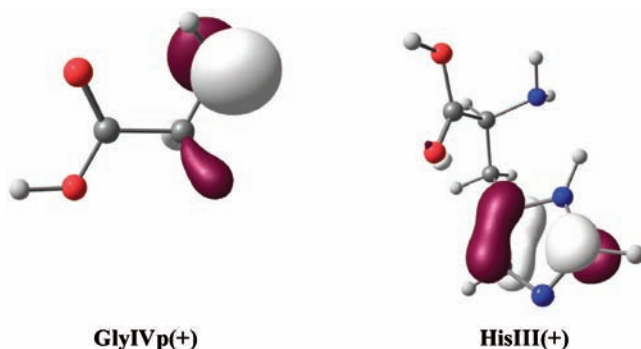
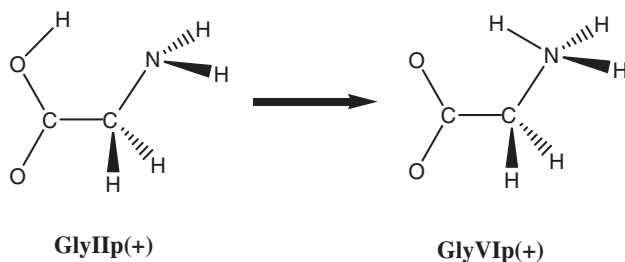


Figure 7.6 Single occupied molecular orbital of the lowest conformer of Gly and His radical cations.

7.4.2

Intramolecular Proton-Transfer Processes

As mentioned above, ionization leads to changes in the intramolecular hydrogen bonds of amino acids that influence the intramolecular proton transfer reactions. A highly important intramolecular proton transfer process in amino acids is the one that connects the neutral with the zwitterionic form. Consequently, we have studied in detail the influence of ionization on this proton transfer (Scheme 7.2) for the simplest amino acid glycine [46]. We considered the structure GlyIIp(+), which, although is not the most stable conformer, is the one involved in this process.



Scheme 7.2

For the neutral system in gas phase this intramolecular proton transfer reaction does not take place because the zwitterionic form of glycine is not stable. However, ionization favors this reaction [46], with the proton transferred structure $[\text{NH}_3\text{CH}_2\text{COO}]^{\bullet+}$, GlyVIp(+), becoming similar in energy to that of the non-transferred one $[\text{NH}_2\text{CH}_2\text{COOH}]^{\bullet+}$, GlyIIp(+) (Figure 7.7). As mentioned, ionization is mainly located at the amino group. Consequently, ionization of GlyIIIn decreases its basicity and the hydrogen bond becomes less favorable, which produces an important increase of the hydrogen bond distance (Figure 7.3). After the hydrogen is transferred to the nitrogen atom, structure GlyVIp(+) in Figure 7.7, the hydrogen bond is strengthened due to an important electronic reorganization. The radical

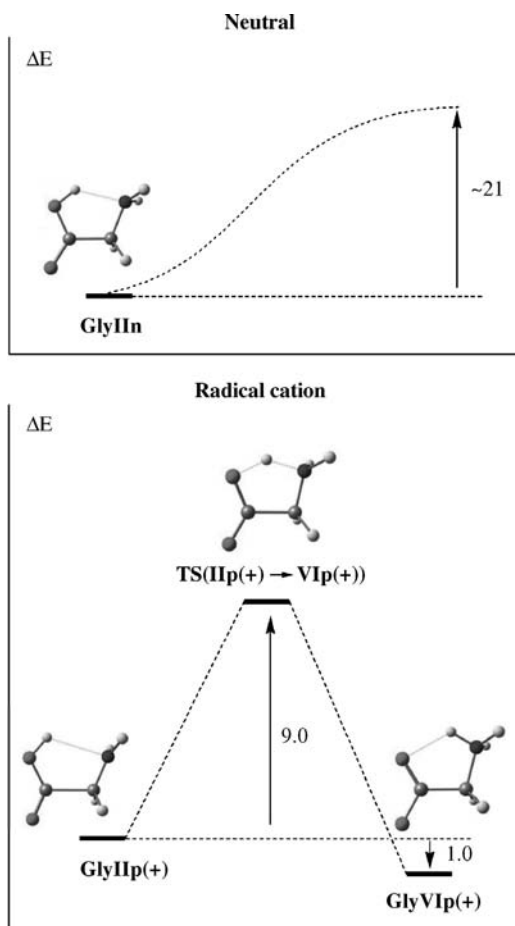


Figure 7.7 Energy profile (kcal mol^{-1}) of the proton transfer process in the neutral and radical cation species of Gly computed at the CCSD(T)/6-311 + + G(3df,2p) level of calculation – with single point calculations computed as described in Reference [46].

character moves completely to the O atom not involved in the hydrogen bond (the spin density over this O atom is 0.88), since this allows the other oxygen to participate in the hydrogen bond with two electrons, thereby strengthening the hydrogen bond.

Other isomerizations involving H-atom transfer reactions can be envisaged arising from structures other than GlyIIp(+) [65, 66]. In fact, structure GlyVIp(+) is the lowest energy conformer of $[\text{NH}_2\text{CH}_2\text{COOH}]^{*+}$ and can undergo isomerization through H-atom transfer, leading to the diol $[\text{NH}_2\text{CHC}(\text{OH})_2]^{*+}$ species, which is the lowest energy isomer of the glycine radical cation PES (Figure 7.8). This large stability is because its structure corresponds to an α -carbon centered radical with both a π donor (NH_2) and a π acceptor, $\text{C}(\text{OH})_2^+$, substituents. Such radicals show an extra stability due to the so-called captodative effect [67]. The isomerization reaction can evolve through two different mechanisms: (i) a single step mechanism consisting

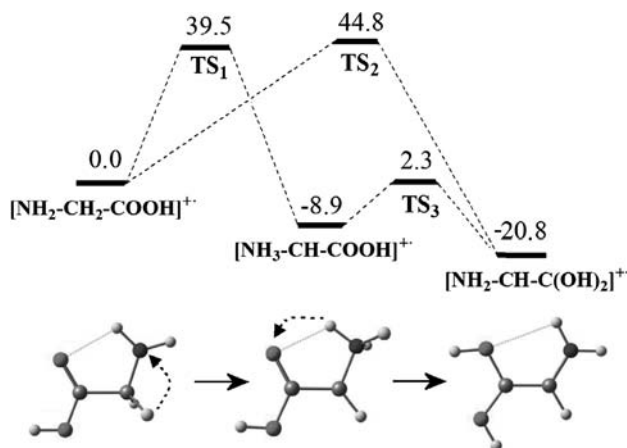


Figure 7.8 Energy profile (kcal mol^{-1}) of the isomerization processes of $[\text{NH}_2\text{CH}_2\text{COOH}]^{\bullet+}$ (structure GlyIVp(+)), computed at the CCSD(T)/6-31++G(d,p) level of calculation. See Reference [65] for details.

of 1,3 H-atom transfer from the CH_2 group to the carbonyl oxygen or (ii) a two-step mechanism, the first step being 1,2 H-atom transfer from the CH_2 group to the amino group, leading to $[\text{NH}_3\text{-CH-COOH}]^{\bullet+}$, and the second step consisting of proton transfer from the NH_3^+ group to the carbonyl oxygen. Natural population analysis seems to indicate that the $[\text{NH}_2\text{CH}_2\text{COOH}]^{\bullet+} \rightarrow [\text{NH}_3\text{CHCOOH}]^{\bullet+}$ process corresponds to a hydrogen atom transfer given that in the reactant the spin is mainly located at the N atom (0.91) whereas in the products the unpaired electron mainly lies at the carbon atom (0.89) [65]. In contrast the $[\text{NH}_3\text{CHCOOH}]^{\bullet+} \rightarrow [\text{NH}_2\text{CHC(OH)}_2]^{\bullet+}$ process can be viewed as a proton transfer reaction. The $[\text{NH}_2\text{CH}_2\text{COOH}]^{\bullet+} \rightarrow [\text{NH}_2\text{CHC(OH)}_2]^{\bullet+}$ process is more complex since the radical character in the reactant lies on the amino group, which is not involved in the 1,3 transfer whereas in the product the spin is delocalized all over the molecule.

Both isomerization mechanisms show high energy barriers because they imply large geometry distortions and electronic rearrangements that transform a nitrogen-centered radical ($[\text{NH}_2\text{CH}_2\text{COOH}]^{\bullet+}$) into two different centered carbon radicals ($[\text{NH}_3\text{CHCOOH}]^{\bullet+}$ and $[\text{NH}_2\text{CHC(OH)}_2]^{\bullet+}$) [65]. However, the presence of a water molecule reduces dramatically the energy barriers of these processes [66], with the most favorable one now being the one-step mechanism with an overall energy barrier of about 7 kcal mol^{-1} . The water molecule acts as a proton transport catalyst but its role is somewhat different in both mechanisms, changing the nature of the isomerizations. As a consequence, for the water-catalyzed system the direct process is energetically more favorable than the two-step one [66].

For amino acids with large enough side chains containing basic sites, isomerization to the most stable diol species can take place through a similar mechanism in which the role of the water molecule is carried out by the side chain [68]. In particular, for the glutamine radical cation ($[\text{NH}_2\text{CHRCOOH}]^{\bullet+}$, $\text{R}=\text{CH}_2\text{CH}_2\text{CONH}_2$) this

process consists of three steps: (i) proton transfer from the C_α to the side chain carbonyl oxygen to form $[NH_2C(CH_2CH_2COHNH_2)COOH]^{\bullet+}$, (ii) rotation of the protonated side chain to form a $OH \cdots OC$ hydrogen bond with the $COOH$ group and (iii) proton transfer from the protonated side chain carbonyl oxygen to $COOH$ to form the final product $([NH_2CRC(OH)_2]^{\bullet+})$ [68]. The overall energy barrier in this self-catalyzed process is about 9 kcal mol^{-1} , which is similar to that found for glycine radical cation in the presence of a water molecule. Thus, this reaction can be considered as an example of a gas-phase, proton-transport, self-catalyzed reaction.

7.5

Ionization of Peptides

Peptides, like amino acids, usually present intramolecular hydrogen bonds that are crucial to understanding their structure and reactivity. Similarly to what has been discussed in the previous section for amino acids, these intramolecular hydrogen bonds can be largely modified upon ionization, inducing important conformational changes.

7.5.1

Ionization of *N*-Glycylglycine

The simplest peptide obtained from the condensation reaction of two amino acid molecules is *N*-glycylglycine ($NH_2CH_2CONHCH_2COOH$). This system provides a reasonable model to study the behavior of the ends of the peptide chains, where the $-(R)CH-$ group is flanked by only one peptide bond. Our main goal in this study is to analyze how ionization influences (i) the intramolecular hydrogen bond strength, (ii) the conformational stability and (iii) the peptide bond.

N-Glycylglycine ($NH_2CH_2CONHCH_2COOH$) embodies three functional groups: $-COOH$, $-NH_2$ and the amide $-CONH-$. Given that the respective ionization energies of $HCOOH$, NH_3 and NH_2COH are 11.33, 10.02 and 10.16 eV [69], ionization of *N*-glycylglycine is expected to take place mainly at the terminal $-NH_2$ group and amide bond. However, since hydrogen-bond interactions modify the energy cost of removing an electron (i.e., proton acceptor groups increase their ionization energy and proton donor groups decrease it) an exhaustive conformational study is important. Thus, as carried out for amino acids, the conformational preferences of neutral *N*-glycylglycine was first explored by carrying out a Monte Carlo multiple minimum conformational search [61] and then reoptimizing the low-lying structures (within an energy window of 10 kcal mol^{-1}) at the DFT level of theory. Radical cation structures were obtained after ionization and reoptimization of the 28 neutral structures obtained. Reference [70] gives detailed information for all structures.

As expected, for neutral *N*-glycylglycine the trans structures are generally more stable than cis ones. The former present relative energies within the $0-7.5 \text{ kcal mol}^{-1}$ range, whereas cis structures lie between 5 and $12.5 \text{ kcal mol}^{-1}$. Removing an

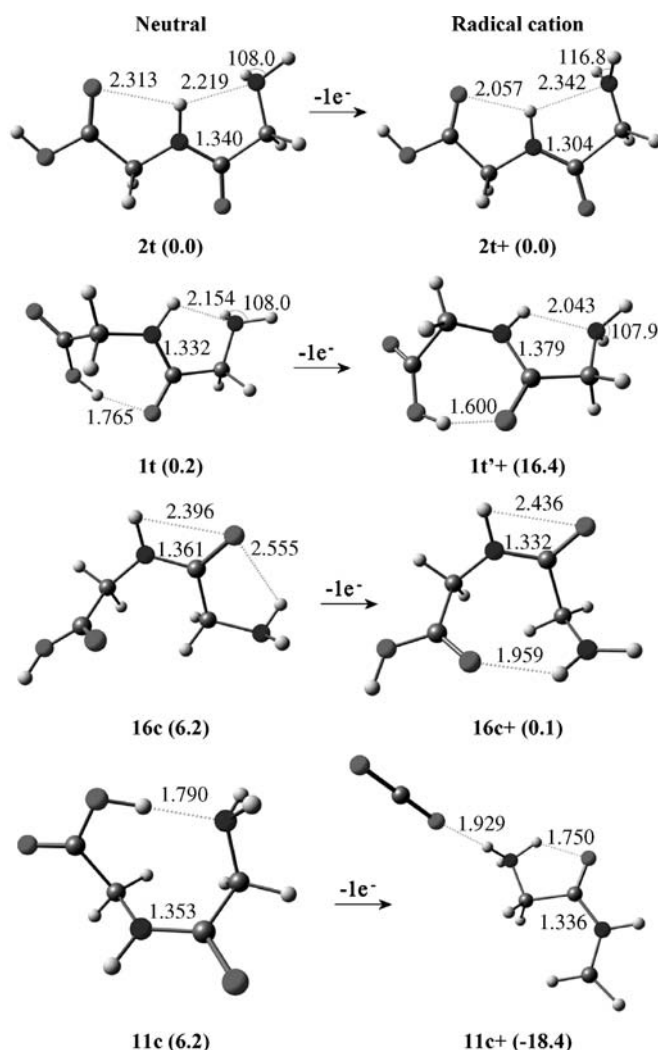


Figure 7.9 Optimized geometries of some representative structures of neutral and ionized *N*-glycylglycine conformations at the MPWB1K/6-31 + G(d,p) level of theory. Distances are in ångströms and energies in kcal mol⁻¹.

electron from neutral *N*-glycylglycine induces significant structural changes that vary according to the starting conformation. Figure 7.9 shows these changes for the most representative conformations. We have included two trans conformations (structures **1t** and **2t**) that are the lowest conformations found for neutral *N*-glycylglycine, in agreement with previous studies [71, 72], and two cis conformations (**11c** and **16c**), all of them presenting significant differences in the rearrangements induced by ionization. For example, for **2t**, ionization is mainly located at the terminal $-NH_2$ group and amide bond, as expected considering that these functional groups are the

ones that present a lower ionization energy (see above). Thus, the initially pyramidalized $-\text{NH}_2$ group becomes more planar in the radical cation species and increases its acidity, which favors intramolecular hydrogen bonds in which this group acts as proton donor. This is particularly remarkable for cis structures such as **16c** + , which can easily fold to form a $-\text{NH}_2^{\bullet+} \cdots \text{OC}$ hydrogen bond. As a consequence of this stabilizing interaction, this ionized cis structure **16c** + becomes almost degenerate with the trans isomer **2t** + . For structures **1t** and **11c**, however, in which the $-\text{COOH}$ and NH_2 groups simultaneously act as proton donor and proton acceptor, respectively, vertical ionization is displaced to the carboxylic group. Thus, the acidity of this group is increased and, for instance, for **1t** the $-\text{COOH} \cdots \text{OC}_{\text{amide}}$ hydrogen bond is significantly strengthened. For **11c**, with a $\text{COOH} \cdots \text{NH}_2$ H-bond, the increase of acidity of COOH and the basicity of NH_2 are high enough to produce a spontaneous proton transfer process that is accompanied by a cleavage of the C–C bond, leading to CO_2 elimination. The structure obtained (**11c** +), $[\text{NH}_3\text{CH}_2\text{CONHCH}_2]^{\bullet+} \cdots [\text{CO}_2]$, is much more stable than any of the conformations explored for the *N*-glycylglycine radical cation. According to these results ionization of *N*-glycylglycine will easily decompose, eliminating CO_2 , in agreement with experiments on the photooxidation of glycylglycine [73], which show a very fast CO_2 elimination and transformation into $[\text{NH}_3\text{CH}_2\text{CONHCH}_2]^{\bullet+}$ distonic radical cations.

Finally, an important aspect to analyze is how ionization modifies the peptide bond. As a general trend, results show that ionization induces a strengthening of the peptide bond. Except for a few structures the shortening of the C–N distance is about 0.03–0.06 Å. This is in agreement with the fact that the reaction energy associated to the condensation reaction of two glycines to form *N*-glycylglycine is more negative for the ionized system ($-13.9 \text{ kcal mol}^{-1}$) than for the neutral one ($-6.7 \text{ kcal mol}^{-1}$) [70]. The strengthening of the peptide bond can be understood considering that ionization takes place essentially at the terminal NH_2 and the O associated to the peptide bond in almost all conformations, which stabilizes the $-\text{HN}^+=\text{CO}^{\bullet}$ resonance form. For structures with a $-\text{OOH} \cdots \text{OC}_{\text{amide}}$ hydrogen bond in the neutral species the shortening of the peptide bond is much smaller or even increases (**1t** +) upon ionization. This is related to the spin distribution of the ionized system, which shows that there is almost no contribution of $-\text{OC}_{\text{amide}}^{\bullet}$. The shortening of the peptide bond is accompanied by a lengthening of the C– CH_2NH_2 distance, which suggests an easy cleavage of this bond after ionization. This cleavage is confirmed by mass spectroscopy experiments that show that the most intense peak is located at m/z 30 and corresponds to the $\text{CH}_2\text{NH}_2^{\bullet+}$ fragment [69].

7.5.2

Influence of Ionization on the Ramachandran Maps of Model Peptides

Backbone conformational features of peptides play an essential role in the secondary structure of proteins. A first step to understanding the intrinsic backbone properties of peptides is to analyze the local conformational landscape of small peptide models through the Ramachandran map [74]; that is, by using the two major conformational

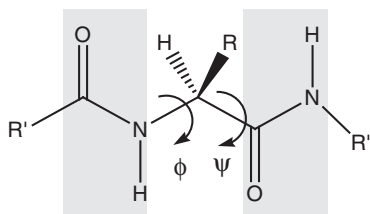


Figure 7.10 The two major conformational variables, ϕ and ψ .

variables, ϕ and ψ (Figure 7.10) the $E = E(\phi, \psi)$ potential energy surface can be built to characterize the stable backbone conformers.

In the last two decades, many Ramachandran maps have been reported for peptide models of glycine and alanine using molecular mechanics, semiempirical methods and *ab initio* and DFT calculations [75–81]. Other studies have analyzed the interplay between the intrinsic backbone properties and the side chain backbone interactions [82–84]. However, no previous studies have analyzed the influence of oxidation on these Ramachandran maps, an important aspect considering that this is a common process in the oxidative damage of proteins that may largely influence the intrinsic backbone conformational properties of peptides. This section aims, thus, to compare the topological features of the Ramachandran surfaces of neutral and ionized systems. Two model peptides were chosen: formyl-glycine amide (HCO-Gly-NH₂) and *N*-formyl-alanine amide (HCO-Ala-NH₂) [70]. Since the main features are similar in both systems we only present the results obtained for HCO-Gly-NH₂. The $E = E(\phi, \psi)$ -type PESs were built at 10° intervals for each conformational variables, ϕ and ψ , optimizing the rest of geometry parameters at the B3LYP/6-31G++G(d,p) level of theory. Direct optimizations, starting from each minima located in the different grids, were also carried using the recently developed meta-hybrid functional MPWB1K and with the larger 6-311++G(2df,2pd) basis set as well as CCSD(T)/6-311++G(2df,2pd) single point calculations at the B3LYP geometries.

Figure 7.11 displays the Ramachandran surfaces for neutral and ionized HCO-Gly-NH₂, as well as the structure of the minima obtained from direct optimization. Since HCO-Gly-NH₂ has a symmetric PES, each stationary point located in the positive part of ϕ is also located in the negative part of ϕ simply by changing the sign of the corresponding ψ . For the neutral system there are three distinguished minima, γ_L , β_L and δ_L . Relative energies at the CCSD(T) level indicate that γ_L is the most stable structure, although β_L lies only 0.8 kcal mol⁻¹ above and δ_L 2.5 kcal mol⁻¹.

The Ramachandran surface of the ionized system shows important differences as compared to the neutral one. The most remarkable is that a new very stable conformer, $\epsilon_L^{\bullet+}$, appears on the PES and that the (ϕ, ψ) values for this conformer correspond to a highly disfavored zone of the neutral PES. On the other hand, the ionized γ_L structure ($\gamma_L^{\bullet+}$) becomes much more stable than the ionized β_L one ($\beta_L^{\bullet+}$). Moreover, the minimum corresponding to the δ_L structure in neutral system disappears whereas another $\gamma_L^{\bullet+}$ conformer, renamed, $\gamma_L^{\bullet+}_{\text{dist}}$ is located at (−10°, −40°). Nevertheless, the relative energy for this new $\gamma_L^{\bullet+}_{\text{dist}}$ conformation is quite high compared to the absolute minimum.

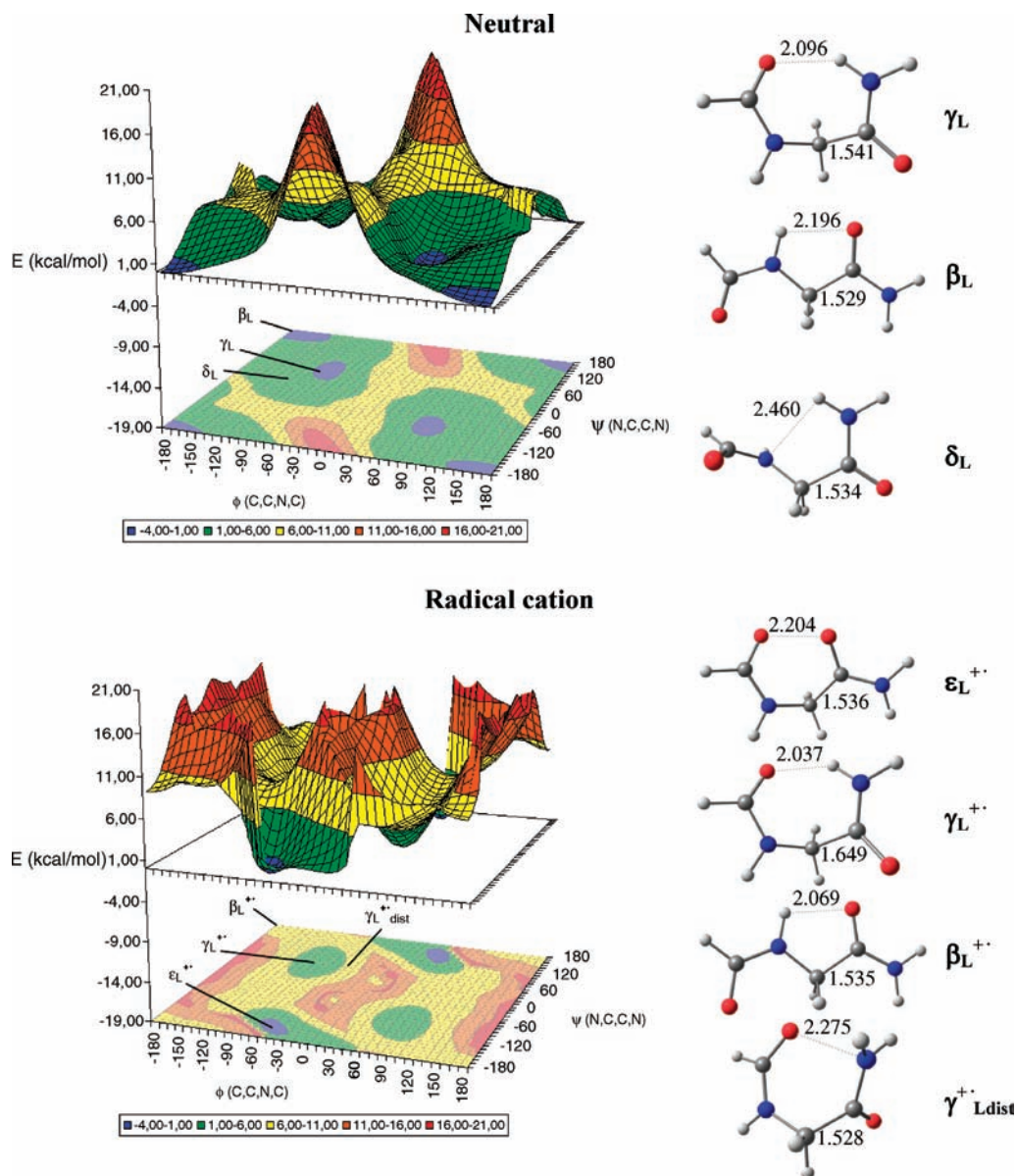


Figure 7.11 Ramachandran maps for neutral and radical cation glycine dipeptide at the B3LYP/6-31++G(d,p) level of theory. Distances are in ångströms and energies in kcal mol⁻¹.

The stability order of these conformers at the CCSD(T) level is: ϵ_L^{*+} (0.0) $\sim$ γ_L^{*+} (0.1) $<$ γ_L^{*+} dist (7.7) $<$ β_L^{*+} (14.5), with relative energies in parenthesis in kcal mol⁻¹. This stability cannot be explained only by means of the changes in hydrogen bond interactions. Other factors such as delocalization phenomena and the possi-

bility of forming hemibonded structures, situations that are typical in open-shell systems, need to be considered. For instance, structure $\beta_L^{\bullet+}$ shows a significant strengthening of the $\text{NH} \cdots \text{OC}$ hydrogen bond because the electron hole is mainly located at the NH group that is acting as proton donor, thereby increasing its acidity. However, despite the strengthening of the hydrogen bond, this $\beta_L^{\bullet+}$ structure is less stable than $\gamma_L^{\bullet+}$ and $\epsilon_L^{\bullet+}$. Thus, other factors need to be taken into account to explain the relative stability of these configurations. One important factor is the large delocalization of the unpaired electron in structure $\gamma_L^{\bullet+}$, for which population analysis indicates that, although the spin density mainly lies at the internal CO (0.64), the CH_2 group acquires a significant radical character (0.14). Consequently, the $\text{C}_\alpha\text{--C}$ bond length increases significantly (about 0.1 Å). In contrast, the great stability of $\epsilon_L^{\bullet+}$ is explained by the formation of a 2-center-3-electron hemibond between the two carbonyl oxygens. This is in agreement with the spin density values (0.45 and 0.46 in each carbonyl group) and with the nature of the open shell orbital, which indicate that the electron hole is delocalized in both oxygen atoms. This is an important issue since these species, in which carbonyl groups of adjacent peptides approach, can be involved in the charge conductivity of peptides [85–88]. Finally, notably, another 2-center-3-electron hemibond between O and N is produced in structure $\gamma_L^{\bullet+}_{\text{dist}}$, although this interaction is less effective than for $\epsilon_L^{\bullet+}$ because it is produced between different kind of atoms [2, 3].

Overall, we can conclude that ionization has a substantial effect on the conformational preferences of a peptide backbone. In general, ionization mainly takes place at the carbonyl groups of the peptide bonds due to the larger stability of the distonic form. Hemibond 2-center-3-electron interactions can then be formed between the two carbonyl groups, leading to very stable $\epsilon_L^{\bullet+}$ structures that could play a crucial role in the electron transfer channel in peptide bond units. However, if the ionized conformer has the terminal carbonyl group involved in a hydrogen bond, as for structures γ , ionization is displaced to the internal carbonyl and the spin density is delocalized towards the adjacent groups, significantly weakening the $\text{C}_\alpha\text{--C}$ bond.

7.6

Conclusions

Noncovalent interactions, and in particular hydrogen bonding, play a fundamental role in biology. They are responsible for the structure and stability of biological macromolecules such as DNA, RNA or proteins and are involved in many molecular-recognition processes as well as in enzymatic reactions. This interaction is largely altered by ionization since removing an electron from a system significantly modifies the basicity and the acidity of the centers involved in the hydrogen bond. In this chapter we have provided the fundamental chemistry that is behind this process in biological molecules such as DNA base pairs, amino acids and peptides, mainly focusing on work performed in our laboratory.

A fundamental point in intermolecular hydrogen bonds is that ionization of the proton donor increases its acidity. Thus, the hydrogen bond interaction is strength-

ened and the proton transfer reaction occurs readily, leading to distonic radical cations. For DNA base pairs, an exhaustive study of different proton transfer processes have allowed us to draw the following conclusions: (i) purines are the monomers that are ionized, which strengthens those hydrogen bonds in which they act as proton donor and weakens those in which they act as proton acceptor; (ii) the relative stability of the different species is determined by the number and sequence of strong and weak hydrogen bonds, the most stable species being those with neighboring strong hydrogen bonds; and (iii) the single proton transfer reactions N_1-N_3 for $GC^{\bullet+}$ and N_6-O_4 for $AT^{\bullet+}$ become much more favorable than for the neutral system because they do not imply a charge separation but a transfer of a positive charge.

Ionization of intramolecular hydrogen bond systems also increases the acidity of the proton donor, but at the same time it decreases the basicity of the proton acceptor. Therefore, it is not clear a priori which would be the effects of ionization on this interaction. Results obtained for amino acids indicate that, in general, ionization is mainly localized at the backbone $-NH_2$ group, thereby strengthening the intramolecular hydrogen bond in which this group acts as proton donor. Aromatic amino acids, however, do not follow this trend because ionization mainly takes place at the side chain. Moreover, results for glycine indicate that the most stable isomer glycine radical cation corresponds to a diol species, the energy barrier for the isomerization reaction being quite high in energy due to important geometrical distortions and large electronic reorganization. This barrier, however, can be significantly reduced by a solvent molecule that acts as a proton transport catalyst, or in the case of amino acids with a long side chain with basic sites, such as glutamine, by the side chain itself in what is called proton transport, self-catalyzed reaction.

For *N*-glycylglycine peptide the nature of the electron hole created is largely dependent on the initial conformation of the neutral system and, thus, an exhaustive prior study on its conformational preferences is essential. In most cases, ionization takes place at the terminal $-NH_2$ and $-CO$ of the amide bond, which produces a strengthening of the peptide bond and the formation of new $-NH_2 \cdots OC_{amide}$ and $-NH_2 \cdots OCOH$ hydrogen bonds. However, if $-NH_2$ and $-CO_{amide}$ act as proton acceptor in the neutral conformation, ionization is displaced at the carboxylic group, which due to its increase of acidity can lead to spontaneous proton-transfer reactions. In fact, the most favorable process observed corresponds to the only low-lying conformation that presents an $OH \cdots NH_2$ hydrogen bond, which after a spontaneous proton transfer process leads to CO_2 elimination.

Finally, the study of Ramachandran plots of model peptides indicates that ionization can cause drastic changes in the shape of the PES in such a way that highly disallowed regions in neutral PES become low-energy regions in the radical cation surface. The structures localized in these regions, $\epsilon_L^{\bullet+}$ and $\epsilon_D^{\bullet+}$, are highly stabilized due to the formation of 2-center-3 electron interactions between the two carbonyl oxygens. This is an important issue because these species, in which the carbonyl groups of adjacent peptides approach, could be involved in the charge conductivity of peptides.

Overall, present studies provide the fundamental trends on the intrinsic chemistry (in absence of solvent) of these radical cations, which we expect can provide valuable

information in different fields. For instance, in mass spectrometry experiments they can help interpret the fragmentation patterns observed for amino acid and peptide radical cations. Moreover, analysis of the electronic structure of these species can help in understanding the mechanism of charge migration in DNA or peptides. Finally, they are the first step for understanding oxidative damage in biological systems.

Acknowledgments

Financial support from MCYT and DURSI, through the CTQ2005-08797-C02-02/BQU and SGR2005-00244 projects, is gratefully acknowledged. S. Simon and A. Gil are also gratefully acknowledged for many contributions to this work.

References

- Gebicki, J. and Bally, T. (1997) *Acc. Chem. Res.*, **30**, 477.
- Sodupe, M., Oliva, A., and Bertran, J. (1994) *J. Am. Chem. Soc.*, **116**, 8249.
- Sodupe, M., Oliva, A., and Bertran, J. (1995) *J. Am. Chem. Soc.*, **117**, 8416.
- Sodupe, M., Oliva, A., and Bertran, J. (1997) *J. Phys. Chem. A*, **101**, 9142.
- Yates, B.F., Bouma, W.J., and Radom, L. (1984) *J. Am. Chem. Soc.*, **106**, 5805.
- Cerny, J. and Hobza, P. (2007) *Phys. Chem. Chem. Phys.*, **9**, 5291.
- Handy, N.C., Knowles, P.J., and Somasundram, K. (1985) *Theor. Chim. Acta*, **68**, 87.
- Sodupe, M., Bertran, J., Rodriguez-Santiago, L., and Baerends, E.J. (1999) *J. Phys. Chem. A*, **103**, 166.
- Becke, A.D. (1993) *J. Chem. Phys.*, **98**, 5648.
- Lee, C., Yang, W., and Parr, R.G. (1988) *Phys. Rev. B*, **37**, 785.
- Becke, A.D. (1993) *J. Chem. Phys.*, **98**, 1372.
- Zhao, Y. and Truhlar, D.G. (2004) *J. Phys. Chem. A*, **108**, 6908.
- Löwdin, P.O. (1965) *Adv. Quantum Chem.*, **2**, 213.
- Löwdin, P.O. (1963) *Rev. Mod. Phys.*, **35**, 724.
- Lunell, S. and Sperber, G. (1967) *J. Chem. Phys.*, **46**, 2119.
- Rein, R. and Harris, F.E. (1964) *J. Chem. Phys.*, **41**, 3393.
- Scheiner, S. and Kern, C.W. (1978) *Chem. Phys. Lett.*, **57**, 331.
- Scheiner, S. and Kern, C.W. (1979) *J. Am. Chem. Soc.*, **101**, 4081.
- Clementi, E., Mehl, J., and Vonnies, W. (1971) *J. Chem. Phys.*, **54**, 508.
- Kong, Y.S., Jhon, M.S., and Lowdin, P.O. (1987) *Int. J. Quantum Chem.*, **32**, 189.
- Clementi, E. (1972) *Proc. Natl. Acad. Sci. U.S.A.*, **69**, 2942.
- Florian, J., Hroudá, V., and Hobza, P. (1994) *J. Am. Chem. Soc.*, **116**, 1457.
- Florian, J. and Leszczynski, J. (1996) *J. Am. Chem. Soc.*, **118**, 3010.
- Hroudá, V., Florian, J., and Hobza, P. (1993) *J. Phys. Chem.*, **97**, 1542.
- Kryachko, E.S. (2002) *Int. J. Quantum Chem.*, **90**, 910.
- Noguera, M., Sodupe, M., and Bertran, J. (2004) *Theor. Chem. Acc.*, **112**, 318.
- Noguera, M., Sodupe, M., and Bertran, J. (2007) *Theor. Chem. Acc.*, **118**, 113.
- Gorb, L., Podolyan, Y., Dziekonski, P., Sokalski, W.A., and Leszczynski, J. (2004) *J. Am. Chem. Soc.*, **126**, 10119.
- Herrera, B. and Toro-Labbe, A. (2007) *J. Phys. Chem. A*, **111**, 5921.
- Shimizu, N., Kawano, S., and Tachikawa, M. (2005) *J. Mol. Struct.*, **735**, 243.
- Villani, G. (2005) *Chem. Phys.*, **316**, 1.
- Hayashi, T. and Mukamel, S. (2004) *Isr. J. Chem.*, **44**, 185.
- Steenken, S. (1997) *Biol. Chem.*, **378**, 1293.
- Bertran, J., Noguera, M., and Sodupe, M. (2003) *Fundamental World Quantum Chem.*, eds. E.J. Brändas and E.S. Kryachko, **2**, 557, Kluwer Academic Publishers.

- 35 Bertran, J., Oliva, A., Rodriguez-Santiago, L., and Sodupe, M. (1998) *J. Am. Chem. Soc.*, **120**, 8159.
- 36 Colson, A.O., Besler, B., and Sevilla, M.D. (1992) *J. Phys. Chem.*, **96**, 9787.
- 37 Hutter, M. and Clark, T. (1996) *J. Am. Chem. Soc.*, **118**, 7574.
- 38 Li, X.F., Cai, Z.L., and Sevilla, M.D. (2001) *J. Phys. Chem. B*, **105**, 10115.
- 39 Li, X.F., Cai, Z.L., and Sevilla, M.D. (2002) *J. Phys. Chem. A*, **106**, 9345.
- 40 Nir, E., Kleinermaans, K., and de Vries, M.S. (2000) *Nature*, **408**, 949.
- 41 Bagheri-Majdi, E., Ke, Y., Orlova, G., Chu, I.K., Hopkinson, A.C., and Siu, K.W.M. (2004) *J. Phys. Chem. B*, **108**, 11170.
- 42 Chu, I.K., Rodriguez, C.F., Lau, T.C., Hopkinson, A.C., and Siu, K.W.M. (2000) *J. Phys. Chem. B*, **104**, 3393.
- 43 Ke, Y., Verkerk, U.H., Shek, P.Y.I., Hopkinson, A.C., and Siu, K.W.M. (2006) *J. Phys. Chem. B*, **110**, 8517.
- 44 Chu, I.K., Zhao, J., Xu, M., Siu, S.O., Hopkinson, A.C., and Siu, K.W.M. (2008) *J. Am. Chem. Soc.*, **130**, 7862.
- 45 Ke, Y., Zhao, J., Verkerk, U.H., Hopkinson, A.C., and Siu, K.W.M. (2007) *J. Phys. Chem. B*, **111**, 14318.
- 46 Rodriguez-Santiago, L., Sodupe, M., Oliva, A., and Bertran, J. (2000) *J. Phys. Chem. A*, **104**, 1256.
- 47 Seymour, J.L. and Turecek, F. (2002) *J. Mass Spectrom.*, **37**, 533.
- 48 Siu, C.K., Ke, Y., Guo, Y., Hopkinson, A.C., and Siu, K.W.M. (2008) *Phys. Chem. Chem. Phys.*, **10**, 5908.
- 49 Steill, J., Zhao, J.F., Siu, C.K., Ke, Y.Y., Verkerk, U.H., Oomens, J., Dunbar, R.C., Hopkinson, A.C., and Siu, K.W.M. (2008) *Angew. Chem. Int. Ed.*, **47**, 9666.
- 50 Sutherland, K.N., Mineau, P.C., and Orlova, G. (2007) *J. Phys. Chem. A*, **111**, 7906.
- 51 Turecek, F. and Carpenter, F.H. (1999) *J. Chem. Soc., Perkin Trans. 2*, 2315.
- 52 Turecek, F., Carpenter, F.H., Polce, M.J., and Wesdemiotis, C. (1999) *J. Am. Chem. Soc.*, **121**, 7955.
- 53 Zhao, J., Siu, K.W.M., and Hopkinson, A.C. (2008) *Phys. Chem. Chem. Phys.*, **10**, 281.
- 54 Siu, C.K., Ke, Y., Orlova, G., Hopkinson, A.C., and Siu, K.W.M. (2008) *J. Am. Soc. Mass Spectrom.*, **19**, 1799.
- 55 Gil, A., Simon, S., Rodriguez-Santiago, L., Bertran, J., and Sodupe, M. (2007) *J. Chem. Theor. Comput.*, **3**, 2210.
- 56 Bonifaci, E., M., Stefanic, I., Hug, G.L., Armstrong, D.A., and Asmus, K.-D. (1998) *J. Am. Chem. Soc.*, **120**, 9930.
- 57 Huang, Y. and Kentamaa, H. (2005) *J. Am. Chem. Soc.*, **127**, 7952.
- 58 Lu, H.-F., Li, F.-Y., and Lin, S.H. (2004) *J. Phys. Chem. A*, **108**, 9233.
- 59 Rauk, A., Yu, D., and Armstrong, D.A. (1998) *J. Am. Chem. Soc.*, **120**, 8848.
- 60 Rega, N., Cossi, M., and Barone, V. (1998) *J. Am. Chem. Soc.*, **120**, 5723.
- 61 Chang, G., Guida, W.C., and Still, W.C. (1989) *J. Am. Chem. Soc.*, **111**, 4379.
- 62 Halgren, T.A. (1999) *J. Comput. Chem.*, **20**, 720.
- 63 Csaszar, A.G. (1992) *J. Am. Chem. Soc.*, **114**, 9568.
- 64 Csaszar, A.G. (1996) *J. Phys. Chem.*, **100**, 3541.
- 65 Simon, S., Sodupe, M., and Bertran, J. (2002) *J. Phys. Chem. A*, **106**, 5697.
- 66 Simon, S., Sodupe, M., and Bertran, J. (2004) *Theor. Chem. Acc.*, **111**, 217.
- 67 Easton, C.J. (1997) *Chem. Rev.*, **97**, 53.
- 68 Gil, A., Simon, S., Sodupe, M., and Bertran, J. (2007) *Theor. Chem. Acc.*, **118**, 589.
- 69 Linstrom, P.J. and Mallard, W.G. (June 2005) NIST chemistry WebBook, NIST Standard Reference Database Number 69, (<http://webbook.nist.gov>).
- 70 Gil, A., Bertran, J., and Sodupe, M. (2006) *J. Chem. Phys.*, **124**, 154306/1.
- 71 Shoeib, T., Rodriguez, C.F., Michael Siu, K.W., and Hopkinson, A.C. (2001) *Phys. Chem. Chem. Phys.*, **3**, 853.
- 72 Wong, C.H.S., Ma, N.L., and Tsang, C.W. (2002) *Chem.-Eur. J.*, **8**, 4909.
- 73 Tarabek, P., Bonifacic, M., and Beckert, D. (2004) *J. Phys. Chem. A*, **108**, 3467.
- 74 Ramachandran, G.N., Ramakrishnan, C., and Sasisekharan, V. (1963) *J. Mol. Biol.*, **7**, 95.
- 75 Head-Gordon, T., Head-Gordon, M., Frisch, M.J., Brooks, C.L. III, and Pople, J.A. (1991) *J. Am. Chem. Soc.*, **113**, 5989.
- 76 Baldoni, H.A., Zamarbide, G.N., Enriz, R.D., Jauregui, E.A., Farkas, O., Perczel, A., Salpietro, S.J., and Csizmadia, I.G. (2000) *THEOCHEM*, **500**, 97.
- 77 Vargas, R., Garza, J., Hay, B.P., and Dixon, D.A. (2002) *J. Phys. Chem. A*, **106**, 3213.

- 78 Perczel, A., Farkas, O., Jakli, I., Topol, I.A., and Csizmadia, I.G. (2003) *J. Comput. Chem.*, **24**, 1026.
- 79 Oldziej, S., Kozłowska, U., Liwo, A., and Scheraga, H.A. (2003) *J. Phys. Chem. A*, **107**, 8035.
- 80 Solov'yov, I.A., Yakubovitch, A.V., Solov'yov, A.V. and Greiner, W. (2005) *Los Alamos Natl. Lab., Prepr. Arch., Phys.*, 1.
- 81 McAllister, M.A., Perczel, A., Csaszar, P., and Csizmadia, I.G. (1993) *THEOCHEM*, **107**, 181.
- 82 Sahai, M.A., Fejer, S.N., Viskolcz, B., Pai, E.F., and Csizmadia, I.G. (2006) *J. Phys. Chem. A*, **110**, 11527.
- 83 Chin, W., Piuze, F., Dimicoli, I., and Mons, M. (2006) *Phys. Chem. Chem. Phys.*, **8**, 1033.
- 84 Sahai, M.A., Kehoe, T.A.K., Koo, J.C.P., Setiadi, D.H., Chass, G.A., Viskolcz, B., Penke, B., Pai, E.F., and Csizmadia, I.G. (2005) *J. Phys. Chem. A*, **109**, 2660.
- 85 Chen, X.H., Zhang, L., Wang, Z.P., Li, J.L., Wang, W., and Bu, Y.X. (2008) *J. Phys. Chem. B*, **112**, 14302.
- 86 Schlag, E.W., Sheu, S.Y., Yang, D.Y., Selzle, H.L., and Lin, S.H. (2000) *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 1068.
- 87 Schlag, E.W., Sheu, S.-Y., Yang, D.-Y., Selzle, H.L., and Lin, S.H. (2000) *J. Phys. Chem. B*, **104**, 7790.
- 88 Schlag, E.W., Yang, D.Y., Sheu, S.Y., Selzle, H.L., Lin, S.H., and Rentzepis, P.M. (2000) *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 9849.

8

To Nano-Biochemistry: Picture of the Interactions of DNA with Gold

Eugene S. Kryachko

8.1

Introductory Nanoscience Background

Nano-small is different [1].

Nano-tiny is beautiful [2].

We are living in the remarkable time of the nano-revolution [4–13] and of the renaissance of gold in science and technology that has brought together researchers from many fields, such as physics, chemistry, material science, electronics, biophysics, biochemistry, biology and medicine, to create, develop and use the structures, devices and systems of the extreme tininess, of the size of about 0.1–100 nm (1 nm is a billionth, 10^{-9} , of a metre or about one 25-millionth of an inch.). This is far smaller than the world of our everyday objects, which are described by Newton's laws of motion, but, conversely, is bigger than an atom or simple molecules like water for instance – the objects that definitively obey the laws of quantum mechanics [3]. This is the world in nano-dimensions, the world of nanosize particles or nanoparticles (NPs) – objects a few nanometres to a few hundred in size – of nanostructures, of nanostructured materials – the *nanoworld* that has recently become one of the largest areas of chemistry, physics, material science, biology and medicine with myriads of applications in catalysis, biophysics and the health sciences.

On the other hand, it is certainly true that many biological molecules, such, primarily, as deoxyribonucleic acid (DNA), ribonucleic acid (RNA), proteins, viruses and biomembranes, are also of a nanometre size – for example, the diameter of the DNA double helix is about 2 nm, its short structural repeat (helical pitch) is about 3.4–3.6 nm, and its “stiffness” with a persistence length (a measure of stiffness) is around 50 nm; many viruses have dimensions of ~ 10 nm, many bacteria are about 100 nm long – and have recently been well recognized as able to naturally integrate with nanoparticles in nano-constructions.

According to M. Roco from the US National Nanotechnology Institute, we may envisage four generations of nanotechnology. The first one, which we are

now witnessing, is the era of passive nanostructures that are engineered to perform a single task.¹⁾ The second is the era of “active” nanostructures for multitasking.¹⁾ The first generation in particular is where nanoparticles of gold met with DNA, which in turn was a major cause of the renaissance of gold. Why does gold matter?

8.1.1

Gold in Nanodimensions

“Zum Golde drängt, am Golde hängt doch alles.”²⁾

J. W. v. Goethe, ‘Faust’

Gold was always considered as the noblest atom in general and, in particular, as the noblest among the seven “coinage” metals – gold copper, iron, mercury, lead, silver and tin (see, for example, Reference [14] and references therein).

What is the origin of the *nobleness* of gold? Primarily, it is merely historical: gold was found in some river sands in its native metallic form and attracted man’s attention due to its color and luster, and its resilience to tarnish and corrosion. The ancients related the coinage metals to certain gods and to certain stellar objects, and to the weekdays as well. Gold was naturally linked to the Sun due to its bright yellow color and was associated with Sunday. It has been known since at least 7000 BC. The fact that gold is not subject to rust, verdigris or emanation, it steadily resists the corrosive action of salt and vinegar was apparently the reason of why gold was used as a medium of exchange for nearly three thousand years. The first gold coins appeared in Lydia in 700 BC.

An atom of gold (Au) has the atomic number 79 in the periodic table of the elements, with an atomic mass of 196.96654. Its ground-state electronic configuration is (Xe)4f¹⁴5d¹⁰6s¹. Gold is the least reactive of the coinage metals. In its bulk form, gold is essentially inert [9, 15]. The rather unique properties of the gold atom are mostly dictated by strong relativistic effects [16, 17].^{3),4)}

1) See in this regard the novel *Prey* by Michael Crichton.

2) The English translation: “Toward gold throng all, to gold cling, all.”

3) For instance, the relativistic effect on the first ionization potential or ionization energy of Au, which is experimentally measured at $IE_1^{\text{exp}}(\text{Au}) = 9.22567 \text{ eV}$ [17f], causes its increase by 1.068 eV compared to the non-relativistic computed value (cf. with that of 0.008 and 0.212 eV for the other coinage metals Cu and Ag, respectively) [17g]. The trend of the EA is similar: for the above series, it is incremented by 0.007, 0.20, and 1.07 eV, respectively [17g]. Note that the experimental value of the EA of Au is equal to $2.927 \pm 0.050 \text{ eV}$ [17h]. Thus, from the point of view of the donor–acceptor features of the gold atom, relativistic effects increase the sum of its ionization potential and electron affinity (electronegativity)

by 2.14 eV, that is, by more than 25% with respect to the non-relativistic value [17i]. In contrast, the relativistic effect on dipole polarizability of Au is negative, -40.0 au , and results in the relativistic dipole polarizability of Au of 36.1 au (notice that the polarizability of gold atom evaluated in Reference [17j] at the PW91/LANL2DZ computational level is equal to 37 au, which agrees fairly well with the early higher-level calculation yielding 39 au [17k]). Within the MO picture, the relativistic effects significantly shrink the size of the 6s valence orbital of the gold atom and extends its 5d subvalence shell [16,17l–m].

4) One of the striking manifestations of the strong relativistic effects is a quite unusual shape of gold clusters Au_N ; when N becomes greater than nine, clusters of gold turn are preferentially three-dimensional (3D) [18].

Within the context of nanotechnology, gold is not used widely as a bulk metal, rather as nanoparticles (colloids), which have known for about 1000 years from their use in ancient stained glass technology to obtain unusual optical properties. The idea of the nanometre scale originated from the following assumption. When the dimensions of the particles are shrunk from the bulk to a scale of the order of the Fermi wavelength of the electron (~ 0.7 nm), that is, to a nanometre size, their behavior dominantly obeys the principles of quantum mechanics [19, 20] and in particular exhibits discrete, quantum-confined electronic transitions [21].

Therefore, new properties were expected to emerge that may be entirely different from those of the bulk and sometimes completely unexpected. This remarkable expectation was first demonstrated by Haruta and coworkers who discovered a surprisingly high catalytic activity of nano-sized gold particles supported on metal oxides for CO oxidation at low temperature [22].

Since then, Au catalysts have attracted growing interest that has been primarily focused on the elucidation of the nature and mechanisms of the extraordinary catalytic activity of nanogold [15b,23]. The activity is due to the so-called quantum “size effect” that modifies the electronic structure and increases the ratio of atoms located on the surface relative to the total number of atoms composing the nanoparticles (NPs) [24]. Indeed, gold NP's of 2–3 nm bridge the “material gap” [25] and exhibit novel shape-dependent electronic, optical and magnetic properties.

In nanodimensions, gold can be of a ruby-red, purple or even blue color depending on the nanoparticle size [15d,26]. In the nanoparticle regime, gold drastically changes its catalytic activity compared to its inert bulk form. For example, oxide-supported gold clusters, particularly gold dispersed on various metal oxides as well as nanosized islands on titania oxide, demonstrate an enhanced chemical reactivity [15b,c,22, 27] that make them very well suited for use as chemical catalysts in many reactions like the combustion of hydrocarbons [28], reduction of nitrogen oxide [29], propylene epoxidation [30, 31] and low-temperature oxidation of carbon monoxide [15b,c, 22, 27, 32–35].

Carbon monoxide oxidation encompasses numerous practical applications. In these, the activity of the gold NPs critically depends on the nanoparticle size, the nature of the support and the detailed synthetic procedure [15d, 27, 35, 36]. STM experiments by Valden *et al.* [15b, 34] demonstrated that the efficiency of CO oxidation on TiO_2 -supported 2D Au islands depends on the island thickness, which is interpreted in terms of a quantum size effect. As shown by Goodman and Chen [23a], the bilayer gold catalyst exhibits an activity for CO oxidation more than ten times larger than the monolayer and the reaction proceeds about 50 times faster on the bilayer gold catalyst than on Au/ TiO_2 catalysts. On the other hand, Freund and coworkers [37] have reported the first experimental evidence that thin islands of gold have the same CO adsorption behavior as large gold NPs and extended gold surfaces. This finding implies that the gold reactivity arises from the presence of highly non- or low-coordinated gold atoms (see also Reference [38]).

To summarize, the experimental studies emphasize that quantum-size effect, lower coordination [37, 39, 40], as well as the charge state, shape [38, 41] and support

interface [42] effects play important roles in the exceptional activity of gold NPs. This has attracted a theoretical interest and motivated further research aimed at providing insights into the molecular origins of such phenomena [15b,c, 16, 37, 40, 43–59], focusing particularly on the exceptional catalytic properties of small gold aggregates [22a,b]. First-principles calculations demonstrate a clear correlation between particle size and chemical activity, which can be explained in terms of an enhanced density of low coordination sites with decreasing cluster size [40, 43]. It is suggested that the Au–Au coordination number has a larger effect on the reactivity of the Au particles than the electronic structure, support or strain and, thus, the presence of low-coordinated gold atoms is the major factor that determines the catalytic activity of the Au NPs [38].

Other studies [15c, 44–46] have also demonstrated that the oxide provides an excess charge to the Au cluster, which is important for the ability to bind and activate O₂. The most reactive sites [44] occur on gold-oxide support interfaces, where the precise interface structure depends on the cluster size and geometry. Investigations on size-selected small gold clusters, Au_{2≤n≤20}, soft-landed on a well-characterized metal oxide support [specifically, a MgO(001) surface with and without oxygen vacancies or F centers [47, 57]], have revealed that gold octamers bound to F centers of the magnesia surface are the smallest known gold heterogeneous catalysts that can oxidize CO into CO₂ at temperatures as low as 140 K.

Together with the exceptional chemical reactivity of nano-sized gold clusters, their three-dimensional shapes, which are largely dictated by the strong relativistic effects, can either be space-filled, compact or admit the existence of some voids, emptiness – so-called cages. The latter resemble the famous buckyball fullerene C₆₀ and bigger fullerenes [60] – which is why they are called “golden fullerenes” [61]. Among them there are few remarkable examples: the cage Au₅₅ (~1.4 nm in diameter) and two lower-energy cages Au₂₀ about 0.8 nm in diameter – the latter are displayed in Figure 8.1 [61s]. The former, space-filled clusters, can be, for instance, represented by the magic gold cluster Au₂₀(T_d) with a unique tetrahedral shape (Figure 8.2), with all atoms on the surface, and a large HOMO–LUMO gap that even slightly exceeds that of the buckyball fullerene C₆₀ [61r, 62].

8.1.2

Gold and DNA: Meeting Points in Nanodimensions

“The nucleic-acid ‘system’ that operates in terrestrial life is optimised (through evolution) chemistry incarnate. Why not use it . . . to allow human beings to sculpt something new, perhaps beautiful, perhaps useful, certainly unnatural”

[Roald Hoffmann, ‘DNA as Clay’, *Am. Sci.* **1994**, 82, 308].

In nanotechnology, there exist two basic types of nanosize constructions: “top-down,” where microscopic manipulations of small numbers of atoms or molecules

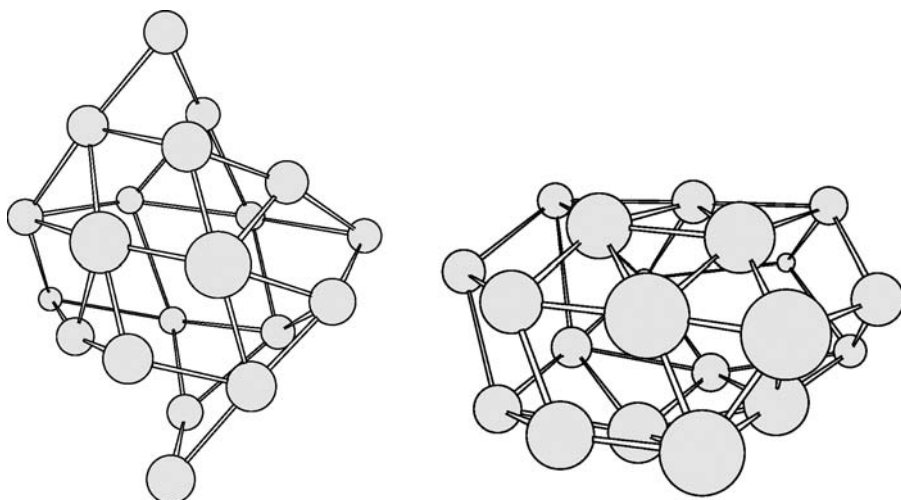


Figure 8.1 Two lower-energy Au₂₀ “fullerene-type” cages.

fashion elegant patterns, and “bottom-up,” where many molecules self-assemble in parallel steps, as a function of their molecular recognition properties.

The integration of molecular systems with the unusual properties of gold nanoparticles has been the focus of numerous concerted efforts directed towards

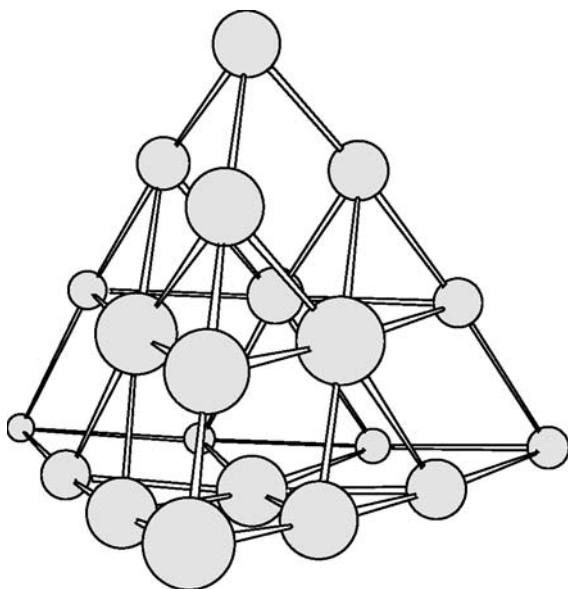


Figure 8.2 The magic gold cluster Au₂₀(T_d).

controlling and improving the properties of a wide range of electro-, photo- or bioactive molecules [63, 64]. The twofold intent is to understand the interactions between metal NPs and the active systems and to design new functional materials that owe their properties to the proximity of a molecule to the metal surface.

The DNA molecule possesses many appealing features for use in nanotechnology – they were partially mentioned in Section 8.1 [65]. As a chemically based assembly system, DNA will be a key player in “bottom-up” nanotechnology.

Utilization of the tremendous recognition properties and functionality of nucleic acids, peptides and proteins, as building blocks in the “bottom-up” self-assembly of nanometre-scaled functional devices, and DNA and peptides as templates in biohybrid complexes has led, in the past few years, to a new research discipline, descriptively termed as nanobiotechnology (see References [66–91] and references therein). Examples of nanobiotechnology are abundant:

- the amine group of peptides that assembles silver and gold cations and caps the growing nanoparticle surface [74];
- adsorption of single-stranded DNA on gold surfaces and its stabilization and assembly of Au NPs of different sizes [67–69];
- metallization of DNA [72] that relies on anchoring the oligonucleotides to a Au surface via thiol-group linkers; these mercapto-group mediators account for some key structural features of biomolecule–metallic nanoparticle complexes [73, 75].

Applications of nanobiotechnology are numerous and in particular include the following:

- organization of metal and semiconductor nanoclusters [68c];
- numerous bioanalytical techniques [78];
- biomolecular electronics [79] and nanomechanical devices, particularly those relying on the DNA [80]:
 - devices from DNA molecules whose functionality is based on conformational changes induced by the binding of intercalators [81];
 - metals, for example, Co^{3+} -ion-dependent B- to Z-DNA transformation [82]; Mg^{2+} -ion-dependent DNA supercoiling [83]; binding of metal ions, such as Cu^{2+} , Zn^{2+} and Ag^{+} , to nucleobases, deoxyribose or phosphodiester backbone, thus providing potential applications of the metallized DNA to nanomaterials and biosensors;
 - pH-dependent formation of intramolecular cytosin quartet structures [84];
 - intermolecular hybridization with so-called fueling oligonucleotides [85]; here, a given DNA conformation is changed upon hybridization with an effector oligomer, which, in turn, can be removed from the complex by hybridization with a second oligomer.

The latter also includes the phenomenon of aggregation of DNA-functionalized gold NPs induced by hybridization of target DNA that does not crosslink the nanoparticles [86], and the discovery and development of polyvalent gold nanoparticle oligonucleotide conjugates DNA-Au NPs [67a,f,69c,d,87]. Interactions between metals and

proteins, polysaccharides and nucleic acids are important since they can be essential for several natural and industrial phenomena [88]. These range from interactions of highly specific metal cofactors with particular proteins [89a] to biosorption of heavy metals by polysaccharide hydrogels [89b]. The unique features of DNA have been exploited in the development of novel materials, especially in the areas of medicine and nanotechnology. Classical research concerning antitumor drugs has focused on the interactions of platinum or ruthenium-containing compounds with the major or minor grooves of polynucleotides [89c–f]. There is tremendous interest in the use of DNA in nanotechnology as a positioning template for the immobilization of metal nanoclusters with view to future applications in the construction of nanoelectronic devices [89g–k]. As an example of DNA–Au NPs interaction, Figures 8.3 and 8.4 report the interaction of Au_{13} and Au_{55} nanoclusters with the major grooves of B-DNA [88]. Finally, DNA-guided assemblies of metal nanoparticles have been reported [90], as

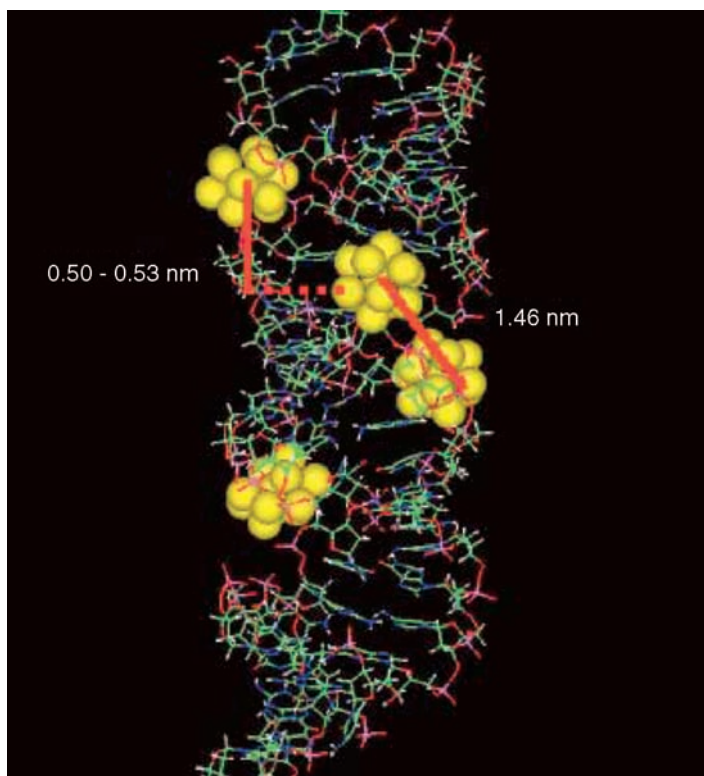


Figure 8.3 Energy minimum complex of A-DNA with Au_{13} clusters located along the phosphate backbone of the DNA major grooves. Its binding energy amounts to 22 kcal mol^{-1} with B-DNA and to 28 kcal mol^{-1} with A-DNA. The

distances between two Au_{13} clusters are indicated by red lines. (Reproduced from Figure 6 of Reference [88] with the permission from the Wiley-VCH Verlag GmbH & Co.)

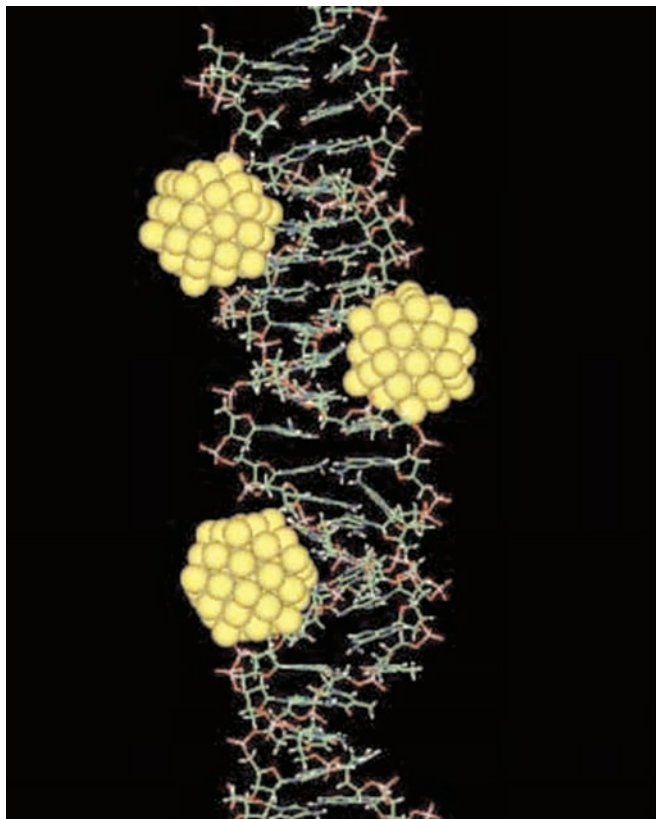


Figure 8.4 Energy minimum complex of the B-DNA with Au_{55} clusters arranged along the phosphate backbone of the DNA major grooves. Its binding energy amounts to $-286 \text{ kcal mol}^{-1}$ with B-DNA and to $+75 \text{ kcal mol}^{-1}$ with A-DNA. (Reproduced from Figure 5 of Reference [88] with permission from the Wiley-VCH Verlag GmbH.)

have DNA–gold nanotubes [91]: hybrid DNA sensors consisting of a gold atom (Au) deposited on two types of single-walled carbon nanotubes.

The ability to build “bottom-up” nanosize biomolecular constructions and to manipulate nanoscopic matter precisely is critical for the development of active nanosystems. To do this we have to know the molecular interactions that govern the formation of these nanosystems, and the DNA–gold interactions in particular. However, the understanding of the mechanism of the bonding between gold nanoparticles and DNA and of the factors that control its efficiency is still rather limited. The assembling of thiolated DNA films is based on specific linker–surface and nonspecific strand–surface interactions. The latter, still not well understood, affect the kinetics of the assembly process as well as the oligonucleotides coverage and orientation within the DNA film. In molecular electronics, the use

of thiol-containing molecules covalently attached to two gold electrodes has raised the question of what is actually the role played by the interface in their resistance. This question remains a subject of debate because of the large disparity existing so far between different experiments [92]. However, the fact [93] that only the two ends of long λ -DNA molecules are fixed on a gold surface via the anchor Au–S bonds often results in mid-segmentation of the DNA chain that easily breaks under rinsing and drying. That is why it is of a great current interest to investigate the direct DNA–gold bonding interaction. What experiments tell us about it?

8.2

DNA–Gold Bonding Patterns: Some Experimental Facts

The experimental picture of the bonding between the DNA bases and gold is as follows:

- 1) The DNA bases adenine (A), thymine (T), guanine (G) and cytosine (C) interact with Au surfaces and small Au NPs in a strong non-specific and sequence-dependent manner [94–96]. The relative binding affinities of these nucleobases for adsorption on polycrystalline Au films decrease in the order [95]:

$$A > C \geq G > T \quad (8.1)$$

Some experiments [95c,e] have demonstrated the occurrence of the dissociation of double-stranded DNA (dsDNA) by 5-nm Au NPs. It is suggested that this is caused by the strong non-specific interaction between the oligonucleotide nitrogen bases and the surface of small Au NPs since the interaction is strong enough to disrupt the hydrogen bonds formed between complementary oligonucleotides, separating the hybridized DNA into single strands, resulting in the stabilization of the small Au NPs. The following two sites of binding of adenine to gold surfaces were usually considered: the N₆ exocyclic amino group [95b,96a] and the N₇ atom [95b]. Nevertheless, the precise geometry of the stable A–Au complex or complexes between A and even the gold atom remained unknown [96b] at that time (see also Reference [96c] for molecular dynamics and [96d] for the density functional theory simulations):

- 2) The heats of desorption, ΔH_{des} , of the DNA bases from Au thin films are [94a,c]:
 - **Adenine:** $\Delta H_{\text{des}}(\text{A}) = 31.3 \pm 0.7 \text{ kcal mol}^{-1}$ (1 kcal = 4.184 kJ) (temperature-programmed desorption, TDP) and $30.8 \pm 1.0 \text{ kcal mol}^{-1}$ (IR)
 - **Guanine:** $\Delta H_{\text{des}}(\text{G}) = 34.9 \pm 0.5 \text{ kcal mol}^{-1}$ (TDP) and $34.4 \pm 0.5 \text{ kcal mol}^{-1}$ (IR)
 - **Thymine:** $\Delta H_{\text{des}}(\text{T}) = 26.5 \pm 0.5 \text{ kcal mol}^{-1}$ (TDP) and $26.3 \pm 0.5 \text{ kcal mol}^{-1}$ (IR)
 - **Cytosine:** $\Delta H_{\text{des}}(\text{C}) = 30.6 \pm 1.0 \text{ kcal mol}^{-1}$ (TDP) and $31.1 \pm 1.2 \text{ kcal mol}^{-1}$ (IR).

That is:

$$\Delta H_{\text{des}}(\text{G}) > \Delta H_{\text{des}}(\text{A}) = \Delta H_{\text{des}}(\text{C}) > \Delta H_{\text{des}}(\text{T}) \quad (8.2)$$

Below, we characterize in detail the properties of these DNA base–Au bonding patterns to identify the factors controlling their formation, with special emphasis on the effects of the gold cluster size and of the coordination of the Au atom, choosing the gold atom Au and the triangular gold cluster Au₃ as a simple catalytic models of Au particles [97].

8.3

Adenine–Gold Interaction

We begin by probing the binding sites of DNA bases to the gold atom Au in the neutral charge state $Z = 0$ – that is, for so-called *gold affinity*.

8.3.1

Adenine–Au and Adenine–Au₃ Bonding Patterns

The potential energy surface (PES) of the A–Au interaction, obtained in Reference [98a] at the computational level B3LYP/RECP (gold) [11] $\cup$ 6-311 + + G(d,p) (DNA base), consists of three stable and one metastable conformers that reflect the gold binding sites. These are depicted in Figure 8.5, where the data in brackets refer to the PES of A·Au₃ calculated at the B3LYP/RECP (gold) $\cup$ 6-31 + G(d) (DNA base) computational level⁵⁾ and the natural order of stability of conformers – I⁰ is more stable or equal to II⁰, which is more stable or equal to III⁰, which is more stable or equal to IV⁰, that is, $E_{\text{b}}^{\text{ZPE}}(\text{I}^0) \leq E_{\text{b}}^{\text{ZPE}}(\text{II}^0) \leq E_{\text{b}}^{\text{ZPE}}(\text{III}^0) \leq E_{\text{b}}^{\text{ZPE}}(\text{IV}^0)$ – in the $Z = 0$ charge state is suggested. ZPE is the usual abbreviation of the zero-point energy.

The above complexes, consisting of a single gold atom and adenine, are characterized by extremely low binding energies that fall within the interval of 2.2–2.5 kcal mol^{−1}. If, instead of the gold atom, a bare triangular gold cluster (Au₃) is bonded to adenine, the binding energies increase by an order of magnitude – see also Table 8.1. For both gold clusters, Au₁ and Au₃, the ring nitrogen atoms N₁, N₃ and N₇ of adenine are the preferential sites to anchor Au₃. The resultant complexes, shown in Figure 8.6 (Table 8.2), are planar and characterized by the binding energies $E_{\text{b}}[\text{A} \cdot \text{Au}_3(\text{N}_{i=1,3,7})]$ 22.6, 24.4 and 22.3 kcal mol^{−1} (Table 8.1), respectively. The latter is about ten times larger $E_{\text{b}}[\text{A} \cdot \text{Au}_1(\text{N}_{i=1,3,7})]$ [98b].

The significant strengthening of the A·Au₃ binding relative to the A·Au₁ results from two bonding patterns. The primary and dominant one is the anchor Au–N

5) The higher computational level B3LYP/RECP (gold) $\cup$ 6-31 + + G(d,p) (DNA base) only slightly modifies the binding energies. For example, $E_{\text{b}}^{\text{ZPE}}[\text{A} \cdot \text{Au}_3(\text{N}_3)]$ varies from −24.4 kcal mol^{−1} for the 6-31 + G(d) basis set on A to −24.0 for the 6-31 + + G(d,p) one.

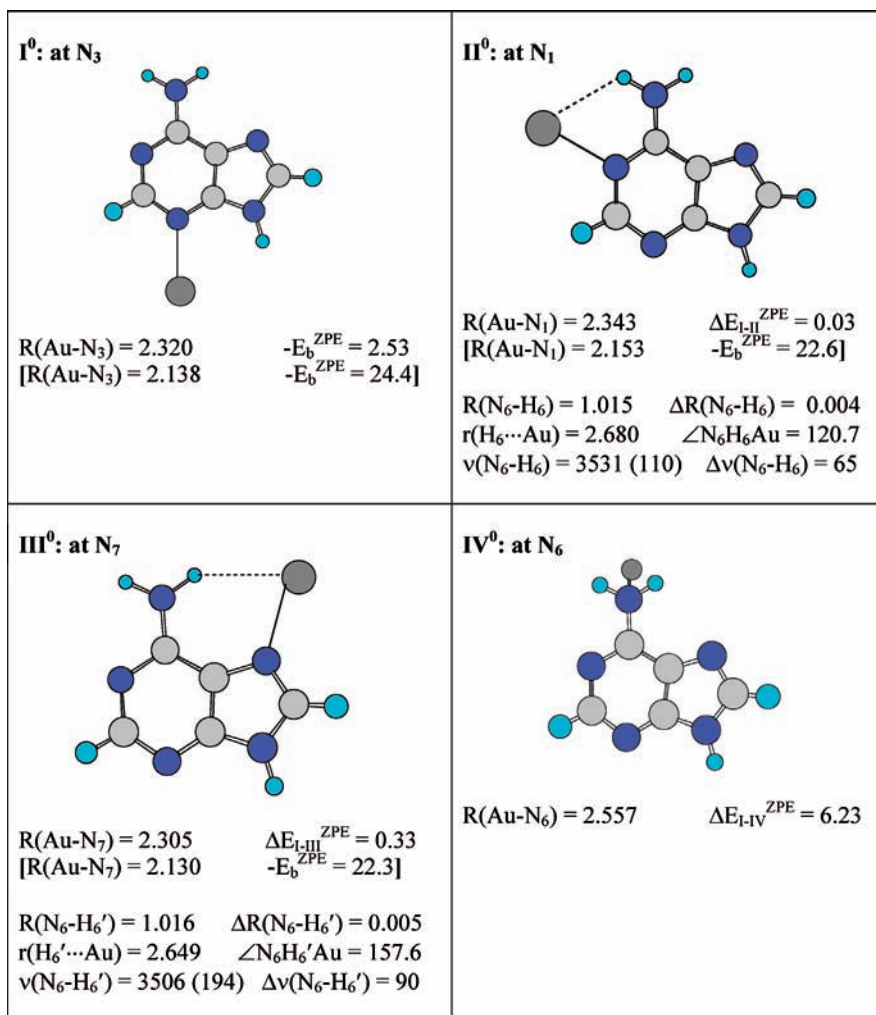


Figure 8.5 Three stable and one metastable conformers that reflect the gold binding sites of the A–Au interaction obtained at the computational level B3LYP/RECP (gold) $\cup$ 6-311 + + G(d,p) (DNA base) [98a].

bonding, which is considerably stronger in $\text{A}\cdot\text{Au}_3(\text{N}_{1,3,7})$ than in $\text{A}\cdot\text{Au}_1(\text{N}_{1,3,7})$ – as is readily seen upon comparison of their lengths, viz., 2.130 Å [$\text{A}\cdot\text{Au}_3(\text{N}_7)$], 2.138 Å [$\text{A}\cdot\text{Au}_3(\text{N}_3)$] and 2.153 Å [$\text{A}\cdot\text{Au}_3(\text{N}_1)$] vs. 2.343 Å [$\text{A}\cdot\text{Au}_1(\text{N}_7)$], 2.320 Å [$\text{A}\cdot\text{Au}_1(\text{N}_3)$] and 2.305 Å [$\text{A}\cdot\text{Au}_3(\text{N}_1)$]. Since the complexes $\text{A}\cdot\text{Au}_3(\text{N}_7)$ and $\text{A}\cdot\text{Au}_3(\text{N}_1)$, which exhibit the shortest and longest anchor bonds, with a difference of ≈ 0.02 Å, are quasi isoenergetic, within the margin of $0.3 \text{ kcal mol}^{-1}$, it is clear that the complexation of $\text{A}\cdot\text{Au}_3(\text{N}_{1,3,7})$ is not only determined by the anchoring. It is also determined by the N–H $\cup$ Au interaction or contact that is established between the N–H group of

Table 8.1 Basic features of the A·Au₃ complexes calculated at the B3LYP/RECP (Au) ∪ 6-31 + G(d) (A) computational level.^{a)}

Complex	−E _b (kcal mol ^{−1})	−ΔH _f (kcal mol ^{−1})	Anchor bond (Å)	ΔR(N–H) (Å)	r(H⋯Au) (Å)	∠N–H⋯Au (°)	−Δν(N–H) (cm ^{−1})	R _{IR}	δσ _{iso} (ppm)	δσ _{an} (ppm)
A·Au ₃ (N ₁)	22.6	22.7	2.153	0.009	2.836	175.2	153	5.6	−2.4	10.3
A·Au ₃ (N ₃)	24.4 24.0 ^{b)}	24.0 23.8 ^{b)}	2.138 2.137 ^{b)}	0.014 0.014 ^{b)}	2.698 2.691 ^{b)}	160.8 161.0 ^{b)}	252 270 ^{b)}	8.7 8.3 ^{b)}	−2.4	13.0
A·Au ₃ (N ₆)	9.9	9.7	2.243							
A·Au ₃ (N ₇)	22.3	21.9	2.130	0.007	2.816	165.1	116	9.0	−2.2	14.0
AH ₁ ⁺ ·Au ₃ (N ₃)	10.6	10.1	2.212	0.028	2.437	165.5	542	10.5		
AH ₆ [−] ·Au ₃ (N ₃)	45.5	45.0	2.091	0.006	3.106	156.0	89	9.4		

a) Few H⋯Au bond lengths exceed the sum of van der Waals radii (equal to 2.86 Å). The binding energy, E_b, and the enthalpy of formation, ΔH_f, are defined with respect to the infinitely separated monomers; Δν(N–H) is and taken relative to the monomer; R_{IR} is the ratio of the IR activities of the corresponding N–H stretches in the hydrogen bonds in the bases or in the base pairs; δσ_{iso} and δσ_{an} are the NMR shifts taken with respect to the corresponding monomers. Extremal values in each column of data are shown in bold.

b) Computational level B3LYP/RECP (Au) ∪ 6-31 + G(d,p) (A).

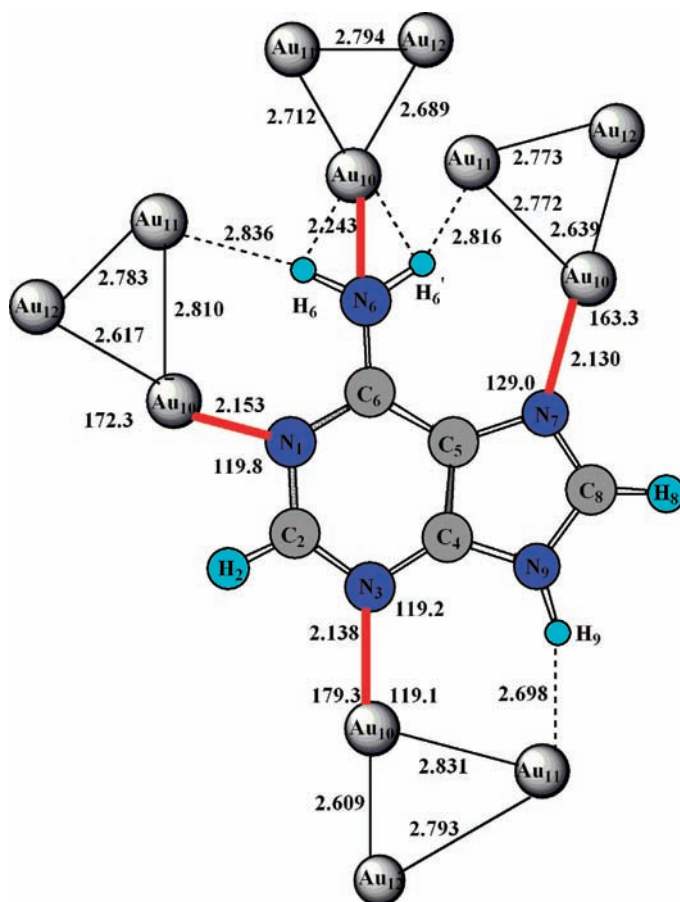


Figure 8.6 Three possible planar –N₁, N₃, N₇– and one non-planar –N₆– binding sites of the gold cluster Au₃ to adenine [98b]. Also shown is the NH₂ anchored complex A·Au₃(N₆). For each complex, the anchor bond is drawn as a thick red line and the nonconventional hydrogen bond as a dotted line. The stability ordering of the complexes is (see also Table 8.1): A·Au₃(N₃) > A·Au₃(N₁) > A·Au₃(N₇) > A·Au₃(N₆). Bond lengths are given in Å and bond angles in degrees. The B3LYP/RECP (gold) ∪ 6-31 + G(d) (DNA base) computational level is invoked. (Reproduced from Figure 1 of Reference [98b] with permission from the American Chemical Society.)

adenine and the unanchored atom of gold [98b]. What is the nature of this N–H ∪ Au contact? It is elucidated in the next subsection.

8.3.2

P propensity of Gold to Act as Nonconventional Proton Acceptor

Below, we provide the definitive computational evidence that the N₆–H₆ ∪ Au₁₁, N₉–H₉ ∪ Au₁₁ and N₆–H₆' ∪ Au₁₁ bonds, formed correspondingly in the complexes

Table 8.2 Bond lengths of the complexes $A \cdot Au_3(N_{1,3,7})$ and of some related protonated and deprotonated species; the B3LYP/RECP (Au) $\cup$ 6-31 + G(d) (A) computational level is invoked.

Bond	A (Å)	A·Au ₃ (N ₁) (Å)	A·Au ₃ (N ₃) (Å)	A·Au ₃ ^{ch} (N ₃) (Å)	A·Au ₃ (N ₇) (Å)	AH ₁ ⁺ (Å)	AH ₁ ⁺ ·Au ₃ (N ₃) (Å)	AH ₆ [−] (Å)	AH ₆ ·Au ₃ (N ₃) (Å)
r(N ₁ -C ₂)	1.344	1.364	1.329	1.330	1.339	1.385	1.375	1.323	1.304
r(N ₁ -C ₆)	1.345	1.364	1.352	1.351	1.350	1.372	1.377	1.401	1.410
r(C ₂ -N ₃)	1.337	1.321	1.349	1.349	1.339	1.299	1.309	1.357	1.374
r(N ₃ -C ₄)	1.339	1.344	1.353	1.349	1.335	1.351	1.364	1.349	1.365
r(C ₄ -C ₅)	1.400	1.395	1.397	1.396	1.398	1.402	1.404	1.394	1.386
r(C ₄ -N ₉)	1.378	1.374	1.366	1.368	1.383	1.363	1.353	1.385	1.373
r(C ₅ -C ₆)	1.411	1.414	1.412	1.412	1.421	1.406	1.404	1.458	1.458
r(C ₆ -N ₆)	1.357	1.339	1.347	1.347	1.344	1.332	1.331	1.308	1.298
r(C ₅ -N ₇)	1.386	1.382	1.382	1.383	1.394	1.372	1.369	1.390	1.387
r(N ₇ -C ₈)	1.312	1.313	1.314	1.312	1.322	1.314	1.318	1.314	1.314
r(C ₈ -N ₉)	1.381	1.381	1.380	1.383	1.363	1.383	1.380	1.384	1.382
r(N ₉ -H ₉)	1.011	1.012	1.025	1.014	1.012	1.015	1.043	1.010	1.016

A-Au₃(N_{i=1,3,7}), obey all the criteria of the conventional hydrogen bond and conclude that they can be therefore treated as so-called “nonconventional” hydrogen bonds.

8.3.2.1 Pause: A Short Excursion to Hydrogen Bonding Theory

The hydrogen bonding interaction [99–109] is a well recognized and widely studied phenomenon that manifests in the formation of a so-called conventional hydrogen ($\text{H} \cdots \text{H}$) bond. According to Pimentel and McClellan [100], “a hydrogen bond is said to exist when (i) there is evidence of a bond and (ii) there is evidence that this bond sterically involves a hydrogen atom already bonded to another atom.” This definition assumes that a conventional hydrogen bond is at least a three-party interaction. One party is a proton donor atom or molecule X. It donates the hydrogen atom H, a second party, bonded to X at the bond length R (X–H), to the third party which is a proton acceptor group Y. The latter, while interacting with X–H, yields the complex X–H $\cdots$ Y hydrogen bond, which is therefore attractive.

Geometrically, the X–H $\cdots$ Y bond is characterized by the bond length R (X–H), the H-bond separation $r(\text{H} \cdots \text{Y})$ and the bond angle $\angle \text{XHY}$. By definition, the H-bond X–H $\cdots$ Y is formed if the following conditions are satisfied [99–109]:

- 1) There exists a clear evidence of the bond formation – this might be, for example, the appearance of the H-bond stretching mode $\nu_o(\text{X} \cdots \text{Y})$.
- 2) There exists a clear evidence that this bond specifically involves a hydrogen atom (hydron) bonded or bridged to Y predominantly along the bond direction X–H (see in particular Reference [108]).
- 3) The X–H bond elongates relative to that in the monomer, that is:

$$\Delta R(\text{X-H}) R_{\text{complex}}(\text{X-H}) - R_{\text{monomer}}(\text{X-H}) > 0 \quad (8.3)$$

- 4) The H-bond separation $r(\text{H} \cdots \text{Y})$, defined as the distance between the bridging proton and the proton acceptor Y, is shorter than the sum of van der Waals radii of H and Y, that is, shorter than the so-called van der Waals cutoff (see in particular References [104, 105, 109] and also Reference [26] $\in$ [108b]):

$$r(\text{H} \cdots \text{Y}) < w_{\text{H}} + w_{\text{Y}} \quad (8.4)$$

where w_{Z} is the van der Waals radius of Z (Z = H, Y). Note that w_{H} varies and usually takes the value of 1.20 Å (see, for example, Reference [110a]) or 1.10 Å [110b]; $w_{\text{Au}} = 1.66$ Å. The distance $r(\text{X} \cdots \text{Y})$ between the proton donor X and the proton acceptor Y is often referred to as the H-bond length.

- 5) The stretching vibrational mode $\nu(\text{X-H})$ undergoes a redshift with respect to that of the isolated X–H group, that is:

$$\Delta \nu(\text{X-H}) \nu_{\text{complex}}(\text{X-H}) - \nu_{\text{monomer}}(\text{X-H}) < 0, \quad (8.5)$$

and its IR intensity significantly increases.

- 6) The proton nuclear magnetic resonance (^1H NMR) chemical shift in the $\text{X}-\text{H}\cdots\text{Y}$ hydrogen bond is shifted downfield compared to the monomer.

Conditions 3–6 can also be treated as some indirect justification of the validity of condition 2.

Throughout a hydrogen bridge, a hydrogen bond connects together X and Y whose electronegativities must be larger than that of hydrogen. Hence, X and Y can, in particular, be chosen as the following atoms: F (3.98), N (3.04), O (3.44), C (2.55), P (2.19), S (2.58), Cl (3.16), Se (2.55), Br (2.96) and I (2.66) where the Pauling electronegativity is given in parentheses.

The above definition of the attractive hydrogen bond interaction is rather general and allows the unification of many types of interaction under the “hydrogen bonding” category, thus considerably extending its conventional manifold, either its X- or Y-submanifolds, or both. What about gold – the “cornerstone,” in some sense, of nanoscience, and in nano-biochemistry in particular? Can gold belong to the Y-submanifold? Or put in the other words: Are the gold atom or clusters of gold prone to play, while interacting with conventional proton donors such as the O–H and N–H groups, a role of a proton acceptor and hence to participate in the formation of nonconventional hydrogen bonds?⁶⁾

8.3.2.2 Proof that $\text{N}-\text{H} \cup \text{Au} \equiv \text{N}-\text{H}\cdots\text{Au}$ in $\text{A}-\text{Au}_3(\text{N}_{i=1,3,7})$

The title proof is rather trivial. Consider the $\text{N}-\text{H} \cup \text{Au}$ contacts that are present in the complexes $\text{A}-\text{Au}_3(\text{N}_{i=1,3,7})$. Obviously, they are similar to conventional weak hydrogen bonds since they obey, according to Table 8.1, all the necessary prerequisites 1–6 of conventional hydrogen bonds gathered in the previous subsection. Therefore, they can be treated as nonconventional hydrogen ones, by analogy with other nonconventional hydrogen bonds with transition metals.

- 6) The suggestion that gold can in principle be a potential candidate for a nonconventional proton acceptor was made and computationally proved for the first time in Reference [111b], which reported a strong computational evidence of the propensity of a triangular gold cluster to behave as a proton acceptor with the O–H group of formic acid and the N–H one of formamide. Since this work, the existence of the $\text{X}-\text{H}\cdots\text{Au}_n$ nonconventional hydrogen bond has been computationally predicted for a wide variety of molecules in different charge states $Z = 0, \pm 1$, ranging from $\text{Au}_n(\text{HF})_m$ [111c] to $[\text{Au}_n^Z(\text{H}_2\text{O})_m]^Z$ [111a] to $[\text{Au}_n(\text{NH}_3)_m]^Z$ [111e–g] complexes including the smallest nanosize 20-gold cluster $\text{Au}_{20}^Z(T_d)$. The hydrogen acceptor propensity of the gold atom and some of its clusters has been experimentally detected

for the complexes $[\text{Au}(\text{H}_2\text{O})]^- - \text{Ar}_n$ [Schneider, H.; Boese, A. D.; Weber, J. M. *J. Chem. Phys.* **2005**, *123*, 084307], $[\text{Au}(\text{H}_2\text{O})_{n=1,2}]^-$ [Zheng, W.; Li, X.; Eustis, S.; Grubisic, A.; Thomas, O.; de Clercq, H.; Bowen, K. *Chem. Phys. Lett.* **2007**, *444*, 232], the crown compound $[\text{Rb}([18]\text{crown-6})(\text{NH}_3)_3]\text{Au}-\text{NH}_3$ [Nuss, H.; Jansen, M. *Angew. Chem. Int. Ed.* **2006**, *45*, 4369; Nuss, H.; Jansen, M. *Z. Naturforsch. Sect. B, J. Chem. Sci.* **2006**, *61*, 1205], for the complexes of small gold clusters with acetone [Shafai, G. S.; Shetty, S.; Krishnamurthy, S.; Kanhere, D. G. *J. Chem. Phys.* **2007**, *126*, 014704] and with amino acids [Pakiari, A. H.; Jamshidi, Z. *J. Phys. Chem. A* **2007**, *111*, 4391] and for a gold(III) antitumor complex [Shi, P.; Jiang, Q.; Lin, J.; Zhao, Y.; Lin, L.; Guo, Z. *J. Inorg. Biochem.* **2006**, *100*, 939].

8.3.2.3 Nonconventional Hydrogen Bonds N–H···Au in A·Au₃ (N_{i=1,3,7})

In the title bonds, the nitrogen atoms, N₆ or N₉, of adenine act as the conventional proton donors whereas the unanchored gold atom Au₁₁, with its lone-pair-like 5d_{±2} and 6s orbitals, behaves as a nonconventional proton acceptor [111]. According to Table 8.1, A·Au₃(N₃) has the strongest nonconventional H-bond in the series A·Au₃(N_{i=1,3,7}): the N₉–H₉···Au₁₁ hydrogen bond exhibits the smallest H-bond distance $r(\text{H}_9 \cdots \text{Au}_{11}) = 2.698 \text{ \AA}$ and the largest redshift of the $\nu(\text{N}_9\text{--H}_9)$ stretch,⁷⁾ equal to 252 cm^{-1} . In addition, the isotropic chemical shift of the bridging proton H₉ is changed by -2.4 ppm , a value close to the NMR shift of the bridging proton of the water dimer (see Table 8.2 in Reference [111a]). The strong character of the N₉–H₉···Au₁₁ hydrogen bond originates from a smaller deprotonation energy (or enthalpy) (DPE) of the proton donating N₉–H₉ group of adenine as compared to the N₆–H₆ and N₆–H₆', viz., $\text{DPE}(\text{N}_9\text{--H}_9) = 336.8 \text{ kcal mol}^{-1} < \text{DPE}(\text{N}_6\text{--H}_6') = 355.2 \text{ kcal mol}^{-1} < \text{DPE}(\text{N}_6\text{--H}_6) = 355.8 \text{ kcal mol}^{-1}$ [113].

The role of nonconventional hydrogen bonding is to enhance the stabilization of the A·Au₃ complexes. Cooperatively with the anchoring, the nonconventional hydrogen bonding substantially contributes to the redistribution of the electron charge within the entire interacting system. This effect can be understood in terms of the NBO analysis of the most stable complex A·Au₃(N₃) shown in Figure 8.6. The Natural Bond Orbital (NBO) analysis demonstrates that the net Natural Population Atomic (NPA) charge on the nitrogen atom, N₉, involved in the formation of the nonconventional H-bond N₉–H₉···Au₁₁ decreases from -0.5866 to -0.5781 |e| [$\Delta q^{\text{NPA}}(\text{N}_9) = -8.5 \text{ m|e|}$].

Since A·Au₃(N₃) is an open-shell many-electron system, this net change of the NPA charge on the atom N₉ is decomposed into the spin components, $\Delta q_{\uparrow}^{\text{NPA}}(\text{N}_9) = +1.9 \text{ m|e|}$ and $\Delta q_{\downarrow}^{\text{NPA}}(\text{N}_9) = -10.4 \text{ m|e|}$. The occupancy of the bonding molecular orbital $\sigma(\text{N}_9\text{--H}_9)$ decreases by 0.8 m|e| (↑) and by 1.4 m|e| (↓). This implies a flow of the electron density from $\sigma(\text{N}_9\text{--H}_9)$, largely of the s- and p-character ($\Delta s = 1.2\text{--}1.3\%$ and $\Delta p = -1.2$ to -1.3%), mostly to the lone pair of the acceptor gold atom Au₁₁ (through the HOMO and LUMO of the triangular gold cluster^{111a}) and to the antibonding MO $\sigma^*(\text{N}_9\text{--H}_9)$. The occupancy of the latter MO increases by 20.2 m|e| (↑) and 1.6 m|e| (↓) and shows a dominant s-character, $\Delta s(\text{N}_9, \uparrow) = +4.5\%$, $\Delta p(\text{N}_9, \uparrow) = -4.6\%$, $\Delta d(\text{N}_9, \uparrow) = +0.1\%$, $\Delta s(\text{N}_9, \downarrow) = +1.2\%$ and $\Delta p(\text{N}_9, \downarrow) = -1.2\%$. The occupancy of the spin-up lone-pair MO on the acceptor gold atom Au₁₁ increases by 0.152 |e| (partly due to the rearrangement of the electron density within gold cluster resulting from the formation of the anchor Au₁₀–N₃ bond) and that of spin-down antibonding lone-pair MO by 24.7 m|e| with $\Delta s(\text{Au}_{11}, \downarrow) = -2.4\%$, $\Delta p(\text{Au}_{11}, \downarrow) = 0.8\%$ and $\Delta d(\text{Au}_{11}, \downarrow) = +1.6\%$. Lastly, the net natural population charge on the bridging proton H₉ increases by only 1.0 m|e| , that is, in terms of spin components, $\Delta q_{\uparrow}(\text{H}_9) = -8.0 \text{ m|e|}$ and $\Delta q_{\downarrow}(\text{H}_9) = +9.0 \text{ m|e|}$. Summarizing, the anchoring bond in the cyclic complex A·Au₃(N_{i=1,3,7}) is a prerequisite for the formation of the

7) Some selected B3LYP/6-31 + G(d) frequencies of adenine [112]: the coupled stretching vibrational modes $\nu(\text{N}_6\text{--H}_6)$ and $\nu(\text{N}_6\text{--H}_6')$ are centered at 3595 cm^{-1} (87 km mol^{-1}) and 3722 cm^{-1} (53 km mol^{-1}); $\nu(\text{N}_9\text{--H}_9) = 3644 \text{ cm}^{-1}$ (78 km mol^{-1}).

nonconventional $N-H \cdots Au_{11}$ hydrogen bond. In turn, the nonconventional H-bond reinforces the anchor bond by a cooperative “back donation” mechanism.

8.3.3

Complex A·Au₃(N₆)

Possible anchoring sites of gold to adenine are not exhausted by the ring atoms N₁, N₃ and N₇. The N₆ atom of the amino group also enables anchoring to Au₃, yielding thus the complex A·Au₃(N₆). The latter's equilibrium geometry is also displayed in Figure 8.6. The anchoring Au₁₀–N₆ bond of A·Au₃(N₆) is ≈ 0.1 Å longer than in A·Au₃(N_{i=1,3,7}) and is equal to 2.243 Å. This difference in the anchoring bond length is the main reason for the weaker stability of A·Au₃(N₆) relative to A·Au₃(N₃), by 14.5 kcal mol^{−1}. Notice that while the Au₁₀–N₆ bond is formed it causes a lengthening of the adjacent C₆–N₆ bond by ~ 0.06 Å.

There exist two basic features that distinguish A·Au₃(N₆) and A·Au₃(N_{i=1,3,7}) from each other: (i) the former does not have a nonconventional H-bond; (ii) the anchoring of Au₃ at N₆ results in their mutual non-planar coordination, which is characterized by the dihedral angle $\angle H_6 H_6' Au_{11} Au_{12} = 82.7^\circ$ and the bond angle $\angle C_6 N_6 Au_{10} = 114.4^\circ$. This coordination resembles the one that occurs under the adsorption of adenine on a Cu(110) surface, as observed experimentally [95b] (see in particular Figure 7 therein and Reference [94a] for a further discussion). It implies that the N₆ atom of the amino group of A partially adopts a sp³ character – note in this respect that $\angle H_6 H_6' N_6 = 110.5^\circ$, $\angle H_6 N_6 Au_{11} = 104.7^\circ$ and $\angle H_6' N_6 Au_{11} = 105.0^\circ$. The Au₁₀–N₆ stretching mode is predicted at 276 cm^{−1}. This value can be compared with the 376 cm^{−1} experimentally detected for the A·Cu(110) system [95b].

8.3.4

Interaction between Adenine and Chain Au₃ Cluster

It is worth ending this section by pointing out why the triangular structure for the three-gold cluster Au₃ was chosen above. As well known, actually Au₃ admits two stable conformers – the chain, Au₃^{ch}, and the triangular, Au₃^{tr} ($\equiv$ Au₃ according to the present convention).⁸⁾ The choice of Au₃^{tr} to model the interaction of the three-gold cluster with DNA is merely motivated by its higher reactivity, which leads to a stronger binding to the nucleobases. For instance, the complex A·Au₃^{ch}(N₃) (Figure 8.7) has a binding energy $E_b[A \cdot Au_3^{ch}(N_3)] = 17.0$ kcal mol^{−1}, that is 7.4 kcal mol^{−1} smaller $E_b[A \cdot Au_3(N_3)]$. Moreover, the complex A·Au₃^{ch}(N₃) is

8) The chain structure Au₃^{ch} is characterized by an electronic energy of -407.911124 hartree and a ZPE of 0.43 kcal mol^{−1}. Its bond lengths $r(Au_1-Au_2) = r(Au_1-Au_3) = 2.619$ Å and the bond angle $\angle Au_2 Au_1 Au_3 = 115.2^\circ$. The chain structure is the most stable conformer of Au₃ that lies below the triangular one by 2.4 kcal

mol^{−1}, after ZPE, that is consistent with the value of 2.3 kcal mol^{−1} [18g]. However, this value falls within the range of the so-called density functional margin approximately equal to 4 kcal mol^{−1} [18d,18p,14b]. Throughout this chapter, Au₃ is identified as the triangular gold cluster.

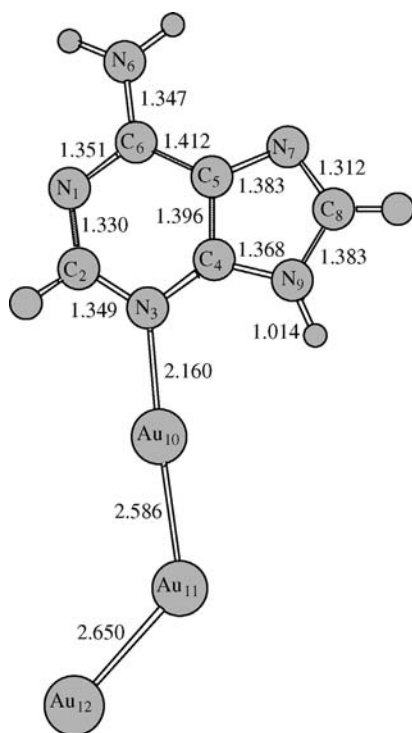


Figure 8.7 Complex $A \cdot Au_3^{ch}(N_3)$; bond lengths are given in Å.

structurally open due to an absence of the $N-H \cdots Au$ H-bond and is thus less interesting in the present context.

8.4

Guanine–Gold Interaction

The PES of the G–Au interaction consists of the four conformers shown in Figure 8.8.

Six stable conformers lie on the low-energy portion of the PES of the $G \cdot Au_3$ displayed in Figure 8.9.⁹⁾ Four of them, $G \cdot Au_3(N_3; N_2)$,¹⁰⁾ $G \cdot Au_3(N_3; N_9)$, $G \cdot Au_3(O_6; N_1)$ and $G \cdot Au_3(N_7)$, are planar and most stable, whereas $G \cdot Au_3(O_6; N_7)$ and $G \cdot Au_3(N_2)$ are non-planar and less strongly bonded (Table 8.3). The ring N_3 atom of G is the most favorable binding site to anchor Au_3 that adopts two quasi-isoenergetic conformers: Au_3 leans either to the amino group, yielding the complex $G \cdot Au_3(N_3; N_2)$ with $E_b[G \cdot Au_3(N_3;$

9) This PES also includes the weak planar complexes $G \cdot Au_3^{ch}(N_3)$, $G \cdot Au_3^{ch}(O_6)$ and $G \cdot Au_3^{ch}(N_7)$. The latter, for example, has a binding energy of $2.3 \text{ kcal mol}^{-1}$.

10) Convention: the word “side” is hereafter omitted and the side site is separated by a semicolon from the anchoring site, for example, $G \cdot Au_3(N_3; N_2 \text{ side}) \equiv G \cdot Au_3(N_3; N_2)$.

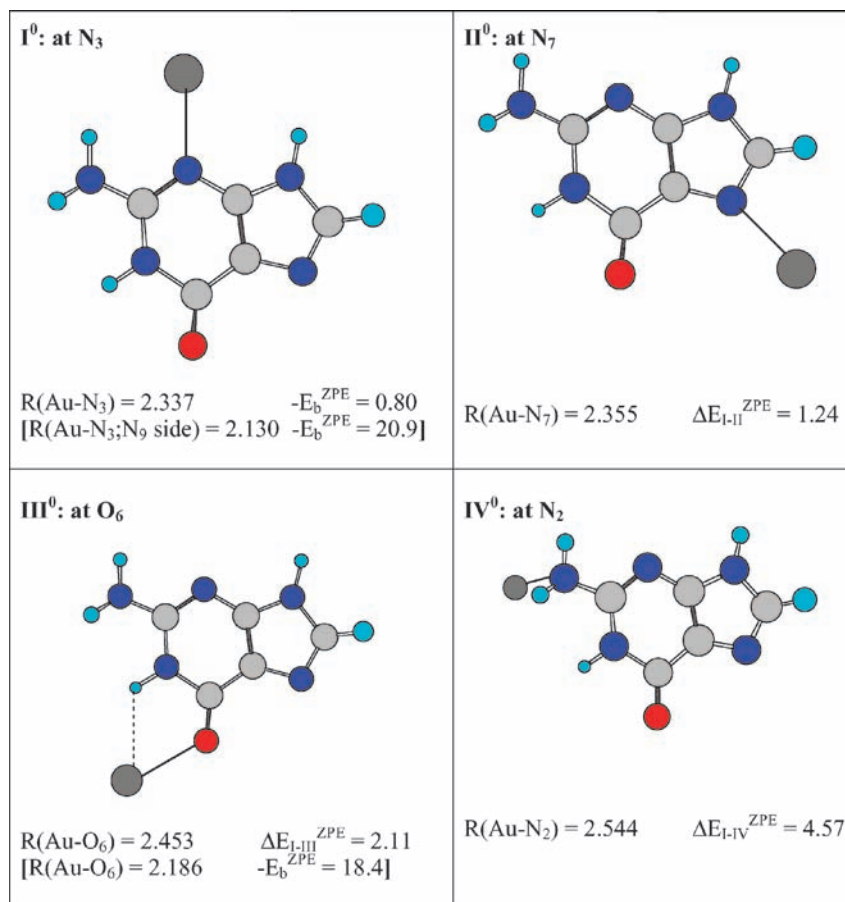


Figure 8.8 The four conformers that comprise the PES of the G–Au interaction.

$\text{N}_2]$ = 20.7 kcal mol^{−1}, or to the $\text{N}_9\text{--H}_9$, giving rise to the complex $\text{G}\cdot\text{Au}_3(\text{N}_3; \text{N}_9)$ characterized by $E_b[\text{G}\cdot\text{Au}_3(\text{N}_3; \text{N}_9)] = 20.9$ kcal mol^{−1}. The latter exceeds $E_b[\text{G}\cdot\text{Au}(\text{N}_3)]$ by a factor of approximately 26. Both complexes, $\text{G}\cdot\text{Au}_3(\text{N}_3; \text{N}_2)$ and $\text{G}\cdot\text{Au}_3(\text{N}_3; \text{N}_9)$, possess the anchoring $\text{Au}_{10}\text{--N}_3$ bonds of practically equal length, 2.146–2.147 Å, the same as in the complex $\text{G}\cdot\text{Au}_3(\text{N}_7)$ whose binding energy $E_b[\text{G}\cdot\text{Au}_3(\text{N}_7)]$ is 19.7 kcal mol^{−1}, that is, ≈ 1 kcal mol^{−1} lower than that of $E_b[\text{G}\cdot\text{Au}_3(\text{N}_3; \text{N}_2)]$ and $E_b[\text{G}\cdot\text{Au}_3(\text{N}_3; \text{N}_9)]$. In contrast to $\text{G}\cdot\text{Au}_3(\text{N}_7)$, $\text{G}\cdot\text{Au}_3(\text{N}_3; \text{N}_2)$ and $\text{G}\cdot\text{Au}_3(\text{N}_3; \text{N}_9)$ possess, however, the nonconventional nearly linear $\text{N}\text{--H}\cdots\text{Au}_{11}$ hydrogen bonds. As shown in Table 8.3, the $\text{N}_9\text{--H}_9\cdots\text{Au}_{11}$ one of $\text{G}\cdot\text{Au}_3(\text{N}_3; \text{N}_9)$ appears to be a bit stronger due to the negative difference in the DPEs of the $\text{N}_9\text{--H}_9$ and $\text{N}_2\text{--H}_2'$ groups [113]: $\text{DPE}(\text{N}_9\text{--H}_9) - \text{DPE}(\text{N}_2\text{--H}_2') = 336.4 - 343.0 = -6.6$ kcal mol^{−1}.

The $\text{Au}\text{--O}_6$ anchoring gives rise to two complexes, $\text{G}\cdot\text{Au}_3(\text{O}_6; \text{N}_1)$ and $\text{G}\cdot\text{Au}_3(\text{O}_6; \text{N}_7)$, and resembles the analogous bond of the complexes $\text{T}\cdot\text{Au}_3(\text{O}_{i=2,4}; \text{N}_3)$ (Section 8.5), slightly shortened by 0.01–0.03 Å in $\text{G}\cdot\text{Au}_3(\text{O}_6; \text{N}_1)$ though. $E_b[\text{G}\cdot\text{Au}_3(\text{O}_6;$

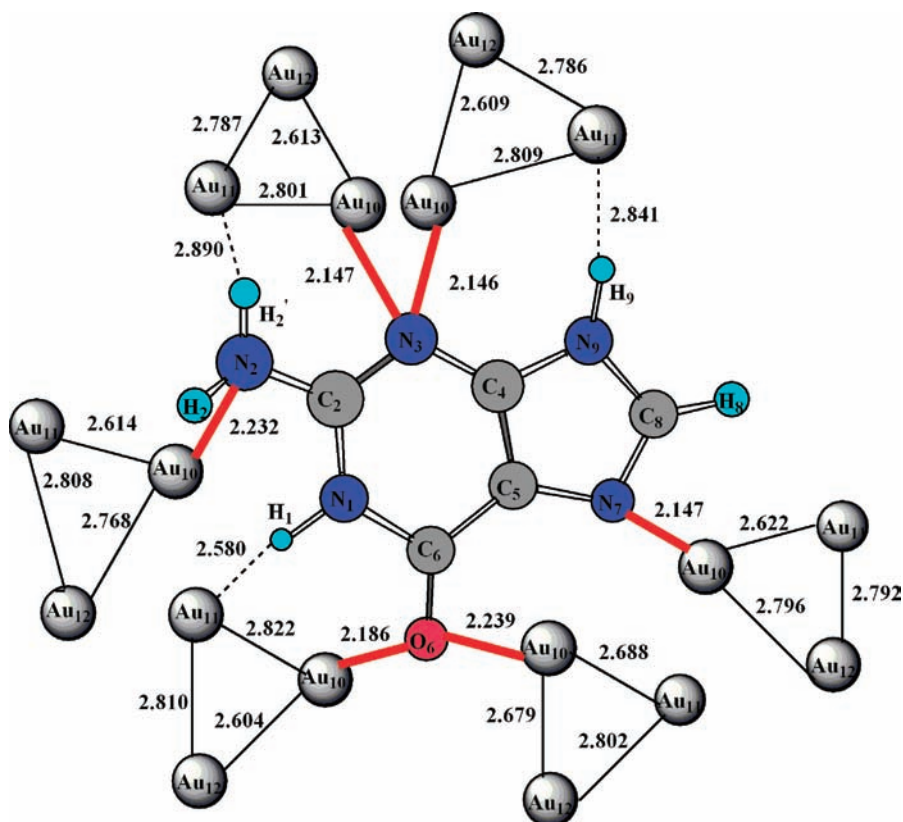


Figure 8.9 Six possible, planar— $\text{N}_3(\text{N}_2), \text{N}_3(\text{N}_9), \text{O}_6(\text{N}_1), \text{O}_6(\text{N}_7), \text{N}_7$ —and N_2 —, whereas $\text{G} \cdot \text{Au}_3(\text{N}_2)$ is non-planar and together with $\text{G} \cdot \text{Au}_3(\text{O}_6; \text{N}_7)$ they are less strongly bonded binding sites of the gold cluster Au_3 to guanine. For each complex, the anchor bond is drawn as a thick (red) line and the nonconventional H-bond as a dotted line. The anchoring in N_2 is to the amino group. The stability ordering of the complexes is (see also Table 8.3): $\text{G} \cdot \text{Au}_3(\text{N}_3; \text{N}_9 \text{ side}) > \text{G} \cdot \text{Au}_3(\text{N}_3; \text{N}_2 \text{ side}) > \text{G} \cdot \text{Au}_3(\text{N}_7) > \text{G} \cdot \text{Au}_3(\text{O}_6; \text{N}_1 \text{ side}) > \text{G} \cdot \text{Au}_3(\text{O}_6; \text{N}_7 \text{ side}) > \text{G} \cdot \text{Au}_3(\text{N}_2)$. The bond lengths are given in Å and bond angles in degrees. The B3LYP/RECP (gold) $\cup 6\text{-}31 + \text{G(d)}$ (DNA base) computational level is invoked. (Reproduced from Figure 3 of Reference [98b] with permission from the American Chemical Society.)

N_1] amounts to $18.4 \text{ kcal mol}^{-1}$, whereas that of $\text{G} \cdot \text{Au}_3(\text{O}_6; \text{N}_7)$ is only $7.9 \text{ kcal mol}^{-1}$ lower. This difference is partially due to the nonconventional hydrogen bond in the former that reinforces the anchoring and causes it to contract by 0.054 Å . In comparison with the Au–N anchoring, the Au–O one is weaker, as reflected by their bond lengths (Table 8.4), and, thus, the $\text{G} \cdot \text{Au}_3$ complexes having an Au–N anchor bond are more stable than those with an Au–O one. The formation of the anchoring bond between N_3 and N_7 of G and gold NPs was pointed out in Reference [95h,95i].

The anchoring of Au_3 at the amino group of the guanine molecule yields the non-planar and less stable complex $\text{G} \cdot \text{Au}_3(\text{N}_2)$ with the bond angle $\angle \text{C}_2\text{N}_2\text{Au}_{10} = 116.7$ and with $E_b[\text{G} \cdot \text{Au}_3(\text{N}_2)] = 9.1 \text{ kcal mol}^{-1}$ (Table 8.3). The formation of the $\text{Au}_{10}\text{—N}_2$

Table 8.3 Basic features of the G·Au₃ complexes calculated at the B3LYP/RECP (Au) ∪ 6-31 + G(d) (A) computational level.^{a)}

Complex	−E _b (kcal mol ^{−1})	−ΔH _f (kcal mol ^{−1})	Anchor bond (Å)	ΔR(N–H) (Å)	r(H···Au) (Å)	∠N–H···Au (°)	−Δν(N–H) (cm ^{−1})	R _{IR}	δσ _{iso} (ppm)	δσ _{an} (ppm)
G·Au ₃ (N ₃ ;N ₂)	20.7	20.1	2.147	0.009	2.890	176.1	115	9.0	−2.5	10.2
G·Au ₃ (N ₃ ;N ₉)	20.9	20.3	2.146	0.010	2.841	161.8	181	6.0	−1.8	11.7
G·Au ₃ (O ₆ ;N ₁)	18.4	17.9	2.186	0.015	2.580	173.1	302	15.0	−3.2	18.7
	18.4 ^(b)	17.9 ^(b)	2.185 ^(b)	0.016 ^(b)	2.568 ^(b)	173.6 ^(b)	324 ^(b)	13.5 ^(b)	−4.0 ^(b)	20.4 ^(b)
G·Au ₃ (O ₆ ;N ₇)	10.5	9.8	2.239							
G·Au ₃ (N ₇)	19.7	19.1	2.147							
G·Au ₃ (N ₂)	9.1	8.8	2.232							
GH ₆ ⁺ ·Au ₃ (N ₃ ;N ₉)	10.8	10.4	2.199	0.024	2.516	164.5	449	7.8		
GH ₁ [−] ·Au ₃ (N ₃ ;N ₉)	42.8	42.3	2.100	0.005	3.185	156.2	75	7.9		
GH ₂ [−] ·Au ₃ (N ₃ ;N ₉)	43.3	42.8	2.100	0.007	2.995	160.4	113	9.6		

- a) Few H···Au bond lengths exceed the sum of van der Waals radii of 2.86 Å (see condition 4 in Section 8.3.2.1). The binding energy, E_b, and the enthalpy of formation, −ΔH_f, are defined with respect to the infinitely separated monomers; Δν(N–H) is taken relative to the monomer; R_{IR} is the ratio of the IR activities of the corresponding N–H stretches in the H-bonds in the bases or in the base pairs; δσ_{iso} and δσ_{an} are the NMR shifts taken with respect to the corresponding monomers. Extremal values in each column of data are shown in bold. Some selected vibrational modes of guanine: the coupled stretching vibrational modes ν(N₂–H₂) and ν(N₂–H₂') are centered at 3562 cm^{−1} (46 km mol^{−1}) and 3668 cm^{−1} (36 km mol^{−1}); ν(N₁–H₁) = 3580 cm^{−1} (44 km mol^{−1}) and ν(N₉–H₉) = 3640 cm^{−1} (68 km mol^{−1}).
- b) Computational level B3LYP/RECP (Au) ∪ 6-31 + G(d,p) (A).

Table 8.4 Bond lengths of the complexes $G \cdot Au_3$ and of some related protonated and deprotonated species; the B3LYP/RECP (Au) $\cup$ 6-31 + G(d) (A) computational level is invoked.

Bond	G (Å)	$G \cdot Au_3$ ($N_3; N_9$) (Å)	$G \cdot Au_3$ ($N_3; N_2$) (Å)	$G \cdot Au_3$ ($O_6; N_1$) (Å)	$G \cdot Au_3$ (N_7) (Å)	GH_6^+ (Å)	$GH_6^+ \cdot Au_3$ ($N_3; N_9$) (Å)	GH_1^- (Å)	$GH_1^- \cdot Au_3$ ($N_3; N_9$) (Å)	GH_2^- (Å)	$GH_2^- \cdot Au_3$ ($N_3; N_9$) (Å)
$r(N_1-H_1)$	1.015	1.015	1.015	1.030	1.015	1.017	1.018			1.013	1.013
$r(N_1-C_2)$	1.372	1.362	1.365	1.378	1.373	1.384	1.391	1.330	1.316	1.415	1.405
$r(N_1-C_6)$	1.438	1.448	1.446	1.397	1.434	1.369	1.372	1.397	1.403	1.403	1.406
$r(C_2-N_2)$	1.379	1.360	1.352	1.358	1.370	1.337	1.343	1.410	1.391	1.308	1.294
$r(C_2-N_3)$	1.312	1.330	1.335	1.319	1.316	1.325	1.337	1.352	1.373	1.372	1.391
$r(N_2-H_2)$	1.013	1.016	1.021	1.010	1.012	1.013	1.018	1.014	1.014	1.022	1.022
$r(N_2-H_2')$	1.013	1.011	1.010	1.014	1.012	1.010	1.012	1.014	1.012		
$r(N_3-C_4)$	1.359	1.374	1.370	1.349	1.353	1.334	1.350	1.346	1.364	1.334	1.359
$r(C_4-C_5)$	1.396	1.393	1.392	1.404	1.391	1.417	1.414	1.396	1.389	1.409	1.396
$r(C_5-C_6)$	1.440	1.439	1.439	1.421	1.441	1.379	1.379	1.460	1.461	1.439	1.443
$r(C_6-O_6)$	1.221	1.217	1.217	1.258	1.218	1.318	1.315	1.236	1.244	1.241	1.233
$r(C_4-N_9)$	1.370	1.362	1.364	1.369	1.375	1.361	1.350	1.382	1.371	1.385	1.370
$r(C_5-N_7)$	1.382	1.380	1.381	1.383	1.385	1.381	1.379	1.392	1.389	1.390	1.387
$r(N_7-C_8)$	1.309	1.310	1.308	1.307	1.319	1.306	1.308	1.310	1.310	1.308	1.308
$r(C_8-N_9)$	1.385	1.384	1.387	1.388	1.370	1.394	1.393	1.388	1.386	1.391	1.389
$r(N_9-H_9)$	1.011	1.021	1.014	1.011	1.012	1.014	1.038	1.010	1.015	1.010	1.017

bond weakens the N_2-H_2 and N_2-H_2' ones, which is why their symmetric and asymmetric stretching vibrational modes are downshifted by 204 and 16 cm^{-1} , respectively.

8.5

Thymine–Gold Interactions

The three conformers shown in Figure 8.10 lie on the PES of the T–Au interaction; Table 8.5 gives the basic features of the T·Au₃ complexes.

Unlike adenine, thymine binds the triangular Au₃ cluster via anchoring at the carbonyl bonds. Three conformers, T·Au₃(O₂; N₁), T·Au₃(O₂; N₃) and T·Au₃(O₄), lie at

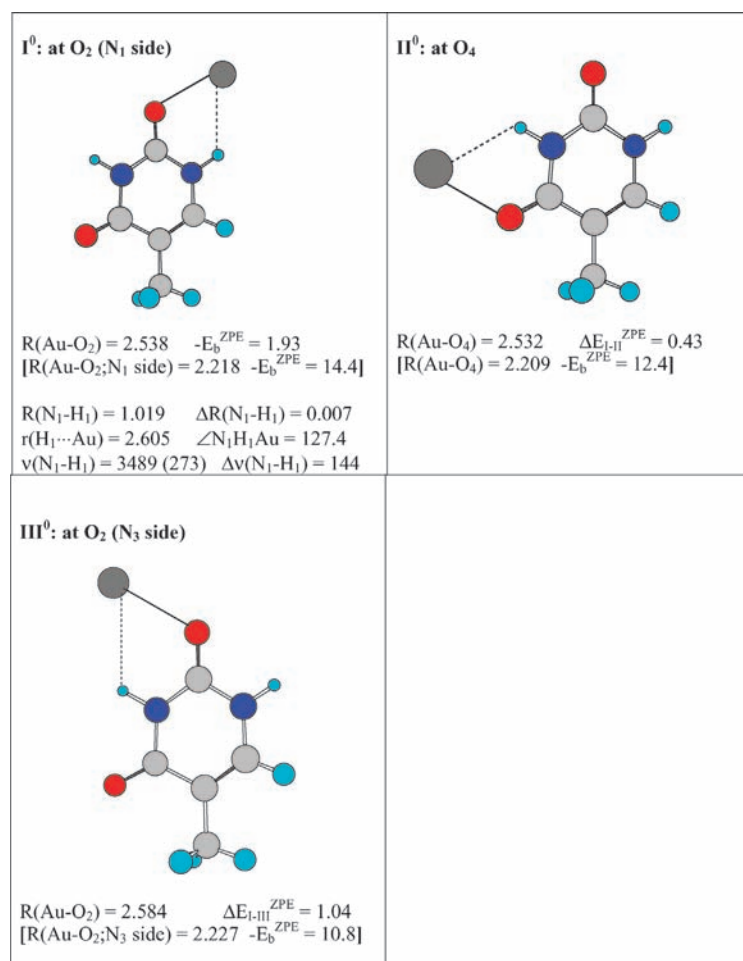


Figure 8.10 The three conformers that lie on the PES of the T–Au interaction.

Table 8.5 Basic features of the T·Au₃ complexes calculated at the B3LYP/RECP (Au) ∪ 6-31 + G(d) (A) computational level.^{a)}

Complex	$-E_b$ (kcal mol ⁻¹)	$-\Delta H_f$ (kcal mol ⁻¹)	Anchor bond (Å)	$\Delta R(N-H)$ (Å)	$r(H \cdots Au)$ (Å)	$\angle N-H \cdots Au$ (°)	$-\Delta \nu(N-H)$ (cm ⁻¹)	R_{IR}	$\delta\sigma_{iso}$ (ppm)	$\delta\sigma_{an}$ (ppm)
T·Au ₃ (O ₂ ;N ₁)	14.4	13.9	2.218	0.017	2.608	178.8	324	11.0	-2.9	16.6
T·Au ₃ (O ₂ ;N ₃)	10.8	10.3	2.227	0.011	2.913	171.8	199	9.0	-1.9	13.9
T·Au ₃ (O ₄)	12.4	11.9	2.209	0.013	2.883	174.4	224	9.0	-2.2	14.1
TH ₄ ⁺ ·Au ₃ (O ₂ ;N ₁)	10.5	9.0	2.365	0.048	2.260	178.0	861	16.9		
TH ₃ ⁻ ·Au ₃ (O ₂ ;N ₁)	37.5	37.1	2.111	0.006	3.137	173.5	103	11.9		

a) Few H ··· Au bond lengths exceed the sum of van der Waals radii of 2.86 Å (see condition 4 in Section 8.3.2.1). The binding energy, E_b , and the enthalpy of formation, $-\Delta H_f$, are defined with respect to the infinitely separated monomers; $\Delta \nu(N-H)$ is taken relative to the monomer; R_{IR} is the ratio of the IR activities of the corresponding N–H stretches in the H-bonds in the bases or in the base pairs; $\delta\sigma_{iso}$ and $\delta\sigma_{an}$ are the NMR shifts taken with respect to the corresponding monomers. Extremal values in each column of data are shown in bold.

the bottom of the potential energy surface of the T-Au₃. They are displayed in Figure 8.11.[19] The anchoring Au₁₀-N_{i=1,3,7} bonds of the complexes A-Au₃(N_{i=1,3,7}) are ≈ 0.1 Å shorter than the Au₇-O_{i=2,4} ones of the above complexes of T-Au₃. These longer bond lengths indicate that the T-Au₃ complexes are less strongly bonded than the A-Au₃ ones, which is confirmed by $E_b[\text{T-Au}_3(\text{O}_2; \text{N}_1)] = 14.4$, $E_b[\text{T-Au}_3(\text{O}_2; \text{N}_3)] = 10.8$ and $E_b[\text{T-Au}_3(\text{O}_4)] = 12.4$ kcal mol⁻¹. These complexes, T-Au₃(O₂; N₁), T-Au₃(O₂; N₃) and T-Au₃(O₄), are also partially stabilized by the nonconventional N-H...Au hydrogen bonding. Among them, T-Au₃(O₂; N₁) has the strongest nonconventional N₁-H₁...Au₈ hydrogen bond. By the six properties that define

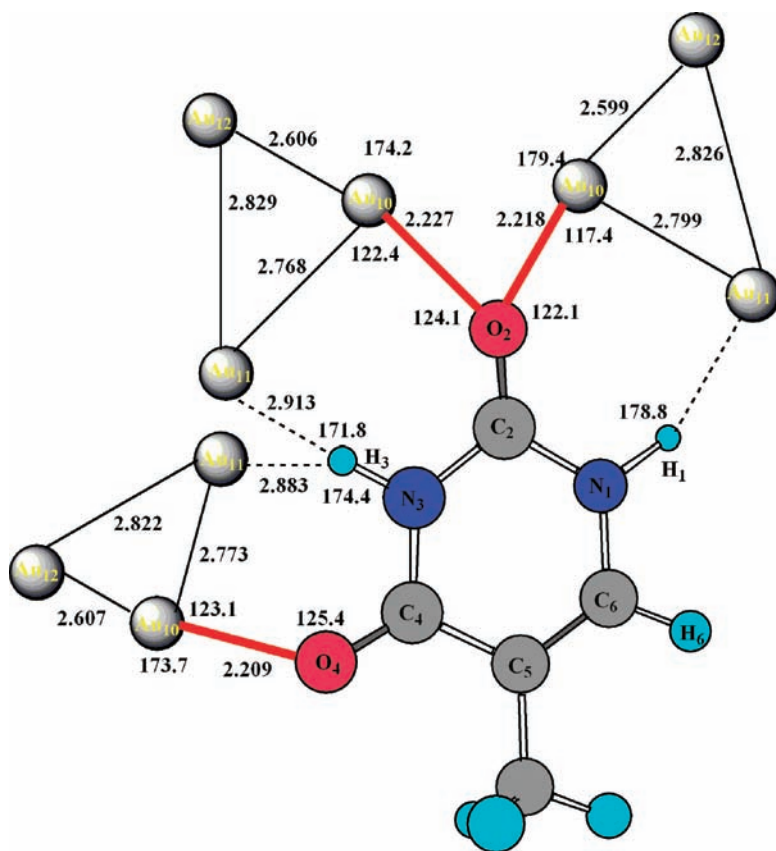


Figure 8.11 Three possible planar –O₂(N₁), O₂(N₃), O₄ – binding sites of the gold cluster Au₃ to thymine. For each complex, the anchor bond is drawn as a thick red line and the nonconventional H-bond as a dotted line. The stability ordering of the complexes is (see also Table 8.5): T·Au₃O₂;

$\text{N}_1) > \text{T} \cdot \text{Au}_3(\text{O}_4) > \text{T} \cdot \text{Au}_3(\text{O}_2, \text{N}_3)$. Bond lengths are given in Å and bond angles in degrees. The B3LYP/RECP (gold) $\cup$ 6-31 + G(d) (DNA base) computational level is invoked.
(Reproduced from Figure 2 of Reference [98b] with permission from the American Chemical Society.)

hydrogen bonding (Section 8.3.2.1) this H-bond is also stronger than those of the complexes $A \cdot Au_3(N_{i=1,3,7})$ despite a weaker anchoring. This can be seen by comparing the redshift $\Delta\nu(N_1-H_1) = 324 \text{ cm}^{-1}$ in $T \cdot Au_3(O_2; N_1)^{11)}$ to $\Delta\nu(N_9-H_9) = 252 \text{ cm}^{-1}$ in $A \cdot Au_3(N_3)$. The relatively stronger H-bonding of $T \cdot Au_3(O_2; N_1)$ is explained by a smaller DPE of the N_1-H_1 bond of thymine compared to that of the N_9-H_9 of adenine: $DPE(N_1-H_1; T) = 334.2 \text{ kcal mol}^{-1} < DPE(N_9-H_9; A) = 336.8 \text{ kcal mol}^{-1}$ [113]. In contrast, the inequality [113] $DPE(N_9-H_9; A) = 336.8 \text{ kcal mol}^{-1}$ $DPE(N_3-H_3; T) = 346.6 \text{ kcal mol}^{-1}$ is a reason of a higher stability of $T \cdot Au_3(O_2; N_1)$ over $T \cdot Au_3(O_2; N_3)$, both anchored to the same $C_2=O_2$ bond of thymine. The latter also indicates a stronger character of the nonconventional hydrogen bonding of $A \cdot Au_3(N_3)$ with respect to that of $T \cdot Au_3(O_2; N_3)$ (Tables 8.1 and 8.6) since the N_9-H_9 group of A is a better proton donor than the N_3-H_3 one of T.

A net strengthening of the stretching vibrational modes $\nu(C_2=O_2)$ and $\nu(C_4=O_4)$ in the studied $T \cdot Au_3$ complexes is a firm indicator of the coordination of thymine to gold. When Au_3 anchors T at O_2 , the $\nu(C_2=O_2)$ downshifts by 97 (N_1) or 87 cm^{-1} (N_3) and its IR activity is enhanced by factor of 1.5–1.6. The $\nu(C_4=O_4)$ undergoes a small blue-shift by 16 and 22 cm^{-1} , respectively. When Au_3 anchors T at O_4 , the $\nu(C_4=O_4)$ is redshifted by 99 cm^{-1} (its IR activity reduces by 25 km mol^{-1}) whereas the frequency of $\nu(C_2=O_2)$ increases by 25 cm^{-1} and its IR activity decreases by 197 km mol^{-1} . The tendency of the $\nu(C=O)$ stretches to downshift under the thymine–gold hybridization is in agreement with the experimental observations [114].

Table 8.6 Bond lengths of the complexes $T \cdot Au_3$ and of some related protonated and deprotonated species; the B3LYP/RECP (Au) $\cup$ 6-31 + G(d) (A) computational level is invoked.

Bond	T(Å)	$T \cdot Au_3$ (O_4) (Å)	$T \cdot Au_3$ ($O_2; N_1$) (Å)	$T \cdot Au_3$ ($O_2; N_3$) (Å)	TH_4^+ (Å)	$TH_4^+ \cdot Au_3$ ($O_2; N_1$) (Å)	TH_3^- (Å)	$TH_3^- \cdot Au_3$ ($O_2; N_1$) (Å)
$r(N_1-H_1)$	1.012	1.013	1.029	1.012	1.018	1.064	1.010	1.016
$r(N_1-C_2)$	1.388	1.392	1.362	1.371	1.401	1.379	1.428	1.391
$r(N_1-C_6)$	1.381	1.374	1.384	1.386	1.353	1.348	1.372	1.377
$r(C_2-N_3)$	1.385	1.395	1.367	1.363	1.412	1.397	1.346	1.318
$r(C_2-O_2)$	1.222	1.215	1.253	1.250	1.201	1.225	1.249	1.297
$r(N_3-C_4)$	1.407	1.380	1.417	1.420	1.347	1.353	1.369	1.383
$r(N_3-H_3)$	1.015	1.028	1.016	1.026	1.020	1.020		
$r(C_4-O_4)$	1.225	1.253	1.221	1.219	1.347	1.353	1.253	1.243
$r(C_4-C_5)$	1.468	1.454	1.466	1.470	1.415	1.410	1.492	1.485
$r(C_5-C_6)$	1.354	1.358	1.355	1.352	1.378	1.383	1.353	1.353

- 11) Some selected vibrational modes of thymine: $\nu(N_1-H_1) = 3633 \text{ cm}^{-1}$ (96 km mol^{-1}) and $\nu(N_3-H_3) = 3592 \text{ cm}^{-1}$ (61 km mol^{-1}); the $\nu(C_2=O_2)$ and $\nu(C_4=O_4)$ stretching vibrational modes are centered at 1805 cm^{-1} (798 km mol^{-1}) and 1760 cm^{-1} (644 km mol^{-1}), respectively.

8.6

Cytosine–Gold Interactions

The four conformers shown in Figure 8.12 are located on the PES of the C–Au interaction.

In the case of cytosine, Au₃ strongly anchors its ring nitrogen atom N₃ and forms the nonconventional N₄–H₄···Au₈ H-bond, as shown in Figure 8.13 (see also Tables 8.7 and 8.8). The binding energy $E_b[\text{C} \cdot \text{Au}_3(\text{N}_3)]$ amounts to $25.4 \text{ kcal mol}^{-1}$ (see Table 8.7, which also summarizes the key properties of the N₄–H₄···Au₈ H-bond). Another complex, C·Au₃(O₂; N₁), is weaker with $E_b[\text{C} \cdot \text{Au}_3(\text{O}_2; \text{N}_1)] = 20.0 \text{ kcal mol}^{-1}$. This difference in binding energies arises from a longer anchoring Au₇–O₂ bond (2.177 Å) than the Au₇–N₃ one (2.164 Å) of C·Au₃(N₃). However, in contrast to the latter complex, C·Au₃(O₂; N₁) has a slightly shorter H-bond (2.627 vs. 2.673 Å). The key feature of the complexes C·Au₃(N₃) and C·Au₃(O₂; N₁) is their perfect planarity.

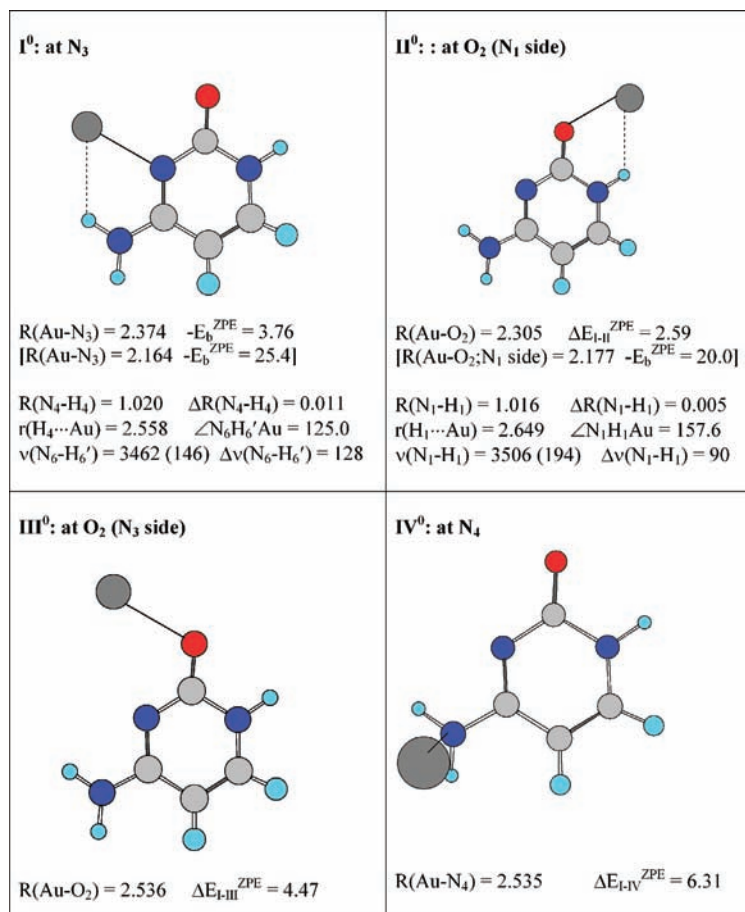


Figure 8.12 The four conformers located on the PES of the C–Au interaction.

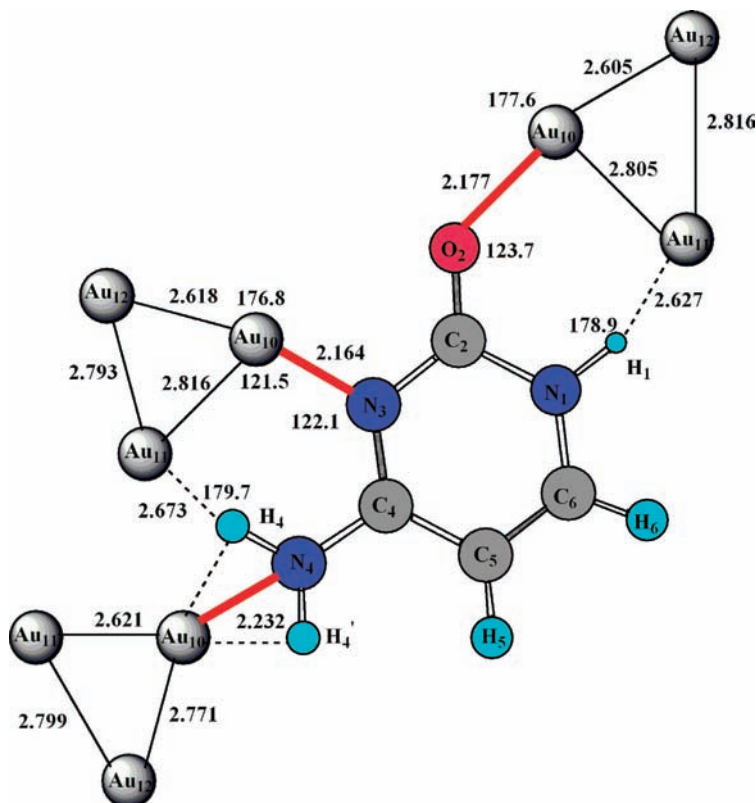


Figure 8.13 Three possible, planar – O₂(N₁), N₃ – and non-planar – N₄ –, binding sites of the gold cluster Au₃ to cytosine. For each complex, the anchor bond is drawn as a thick red line and the nonconventional H-bonds in dotted lines. For the binding site N₄, the anchor bond is to the NH₂ group. The stability ordering of the complexes is (see also Table 8.7): C·Au₃(N₃) > C·Au₃(O₂; N₁) > C·Au₃(N₄). Bond lengths are given in Å and bond angles in degrees. The B3LYP/RECP (gold) ∪ 6-31 + G(d) (DNA base) computational level is invoked. (Reproduced from Figure 4 of Reference [98b] with permission from the American Chemical Society.)

A non-planar coordination of a gold cluster to cytosine arises when Au₃ anchors at the amino group, yielding the complex C·Au₃(N₄) with the bond angle $\angle C_4N_4Au_7 = 114.0^\circ$. Its binding energy amounts only to 11.2 kcal mol^{−1}. Notice that the bond length, $r(Au_7-N_4) = 2.232$ Å, is, however, 0.07 Å smaller than that of C·Au₃(N₃).

8.7

Basic Trends of DNA Base–Gold Interaction

This section discusses the most important features of the interaction between the DNA bases and gold clusters Au₂≤_n≤₆, particularly those that depend on the charge state.

Table 8.7 Basic features of the C-Au₃ complexes calculated at the B3LYP/RECP (Au) ∪ 6-31 + G(d) (A) computational level.^{a)}

Complex	$-E_b$ (kcal mol ⁻¹)	$-\Delta H_f$ (kcal mol ⁻¹)	Anchor bond (Å)	$\Delta R(N-H)$ (Å)	$r(H \cdots Au)$ (Å)	$\angle N-H \cdots Au$ (°)	$-\Delta \nu(N-H)$ (cm ⁻¹)	R_{IR}	$\delta\sigma_{iso}$ (ppm)	$\delta\sigma_{an}$ (ppm)
C-Au ₃ (O ₂ ;N ₁)	20.0	19.5	2.177	0.016	2.627	178.9	306	14.0	-3.2	17.6
C-Au ₃ (N ₃)	25.4	25.1	2.164	0.014	2.673	179.7	232	8.0	-3.2	12.6
C-Au ₃ (N ₄)	11.2	10.9	2.232							
CH ₃ ⁺ ·Au ₃ (O ₂ ;N ₁)	10.0	9.4	2.361	0.042	2.290	178.3	786	31.3		
CH ₄ ⁺ ·Au ₃ (O ₂ ;N ₁)	38.9	38.6	2.107	0.005	3.136	173.7	99	10.4		

a) Few H ··· Au bond lengths exceed the sum of van der Waals radii of 2.86 Å (see condition 4 in Section 8.3.2.1). The binding energy, E_b , and the enthalpy of formation, $-\Delta H_f$, are defined with respect to the infinitely separated monomers; $\Delta \nu(N-H)$ is taken relative to the monomer; R_{IR} is the ratio of the IR activities of the corresponding N-H stretches in the H-bonds in the bases or in the base pairs; $\delta\sigma_{iso}$ and $\delta\sigma_{an}$ are the NMR shifts (in ppm) taken with respect to the corresponding monomers. Extremal values in each column of data are shown in bold. Some selected modes of cytosine: the coupled stretching vibrational modes $\nu(N_4-H_4)$ and $\nu(N_4-H_4')$ are centered at 3590 cm⁻¹ (71 km mol⁻¹) and 3715 cm⁻¹ (43 km mol⁻¹); $\nu(N_1-H_1) = 3611$ cm⁻¹ (66 km mol⁻¹).

Table 8.8 Bond lengths of the complexes $C\cdot Au_3(N_3)$, $C\cdot Au_3(O_2; N_1)$, $C\cdot Au_3(N_4)$, and of some of their selected protonated and deprotonated species; B3LYP/RECP (Au) $\cup$ 6-31 + G(d) (A) computational level is invoked.

Bond	C (Å)	$C\cdot Au_3(N_3)$ (Å)	$C\cdot Au_3(O_2; N_1)$ (Å)	$C\cdot Au_3(N_4)$ (Å)	$CH_3^+ \cdot Au_3$ ($O_2; N_1$) (Å)	CH_4^- (Å)	$CH_4^- \cdot Au_3$ ($O_2; N_1$) (Å)
$r(N_1-H_1)$	1.013	1.013	1.029	1.014	1.017	1.010	1.015
$r(N_1-C_2)$	1.428	1.411	1.395	1.423	1.397	1.442	1.403
$r(N_1-C_6)$	1.356	1.358	1.358	1.355	1.357	1.368	1.374
$r(C_2-N_3)$	1.371	1.388	1.348	1.386	1.414	1.336	1.309
$r(C_2-O_2)$	1.224	1.219	1.261	1.218	1.226	1.250	1.300
$r(N_3-C_4)$	1.321	1.346	1.331	1.304	1.360	1.377	1.390
$r(C_4-N_4)$	1.364	1.339	1.353	1.434	1.332	1.315	1.305
$r(N_4-H_4)$	1.012	1.026	1.011	1.021	1.014		
$r(N_4-H_4')$	1.009	1.010	1.008	1.021	1.013	1.025	1.025
$r(C_4-C_5)$	1.441	1.439	1.436	1.427	1.422	1.480	1.475
$r(C_5-C_6)$	1.361	1.358	1.362	1.366	1.367	1.353	1.351

8.7.1

Anchoring Bond in DNA Base–Gold Complexes

Summarizing the bonding patterns formed between the DNA bases, on the one hand, and the Au atom and Au₃ cluster on the other, we conclude that they are either monofunctional, that is, solely rely on the gold–base anchoring, or bifunctional and involve a nonconventional N–H···Au hydrogen bonding in addition to the anchoring one. In these complexes, the anchoring bonding interaction is unequivocally dominant.

The anchoring bonding arises from the combination of various effects that in particular include a covalent bonding of the Au–N or Au–O type, charge transfer, electrostatic effects and dispersion interactions. The covalent bonding originates from electron sharing between the lone-pair orbitals of the nitrogen or oxygen atoms and the gold 5d and 6s ones. Such sharing and, hence, the strength of covalent bonding both depend obviously on the bond length. For all the most stable nucleobase·Au₃ complexes, the Au–N bonds are shorter than Au–O ones: 2.164 Å [C·Au₃(N₃)], 2.138 Å [A·Au₃(N₃)], 2.153 Å [A·Au₃(N₁)], 2.130 Å [A·Au₃(N₇)], 2.147 Å [G·Au₃(N₃; N₂)], 2.146 Å [C·Au₃(N₃; N₉)] and 2.147 Å [G·Au₃(N₇)] vs. 2.177 Å [C·Au₃(O₂; N₁)], 2.186 Å [G·Au₃(O₆; N₁)], 2.209 Å [T·Au₃(O₄)], 2.218 Å [T·Au₃(O₂; N₁)] and 2.227 Å [T·Au₃(O₂; N₃)].

The shortest Au–N bond – 2.130 Å – is formed in the A·Au₃(N₇) complex, which is not, however, the most stable complex, even in the series of A·Au₃, since its binding energy only amounts to 22.3 kcal mol^{–1}. The shortest Au–O bond (2.177 Å) occurs in the C·Au₃(O₂; N₁) complex characterized by $E_b = 20.0$ kcal mol^{–1}, which is the largest among the DNA base·Au₃ complexes with a Au–O anchoring. Overall, this implies that the covalent bonding definitely contributes to the “anchoring” of the base–gold complexes although it is not a unique factor determining their stabilities.

The charge-transfer effect is larger for gold–nitrogen than gold–oxygen anchorings. To show this, we consider the following two representative complexes, A·Au₃(N₃) and T·Au₃(O₂; N₁), and analyze the changes in the Mulliken atomic charges under the Au₃-anchoring with respect to those of the bare A and T (Table 8.9). It directly follows from Table 8.9 that the stronger character of the Au₁₀–N₃ anchoring in A·Au₃(N₃) is accounted for by a larger change of the Mulliken charges of the N₃ and Au₁₀ atoms. They are $\Delta q^M(N_3) = -0.051 |e|$ and $\Delta q^M(Au_{10}) = 0.184 |e|$, compared, respectively, to $\Delta q^M(O_2) = -0.016 |e|$ and $\Delta q^M(Au_7) = 0.132 |e|$ in T·Au₃(O₂; N₁). Conversely, the nonconventional N₁–H₁···Au₈ hydrogen bonding of T·Au₃(O₂; N₁) is stronger than the N₉–H₉···Au₁₁ one of A·Au₃(N₃). This is explained by the larger $\Delta q^M(N_1) = 0.107 |e|$ and $\Delta q^M(H_1) = 0.017 |e|$ that accompany the formation of the nonconventional H-bond of the former system, in comparison with $\Delta q^M(N_9) = 0.091 |e|$ and $\Delta q^M(H_9) = 0.009 |e|$ for the latter.

Electrostatic effects, such as charge polarization in particular, are also quite significant for the DNA base–gold interaction due to the large electric fields at the

Table 8.9 Mulliken charges q^M of atoms of the complexes $A\cdot Au_3(N_3)$ and $T\cdot Au_3(O_2; N_1)$ near the anchoring and nonconventional hydrogen bonds.

Atom	A/ Au_3 ($ e $)	A- $Au_3(N_3)$ ($ e $)	Atom	T/ Au_3 ($ e $)	T- $Au_3(O_2; N_1)$ ($ e $)
N ₁	−0.381	−0.325	N ₁	−0.595	−0.488
C ₂	0.074	0.118	H ₁	0.429	0.446
N ₃	−0.326	−0.377	C ₂	0.712	0.716
C ₄	0.229	−0.015	O ₂	−0.536	−0.520
N ₉	−0.600	−0.509	N ₃	−0.725	−0.702
H ₉	0.422	0.431	Au ₇	0.122	0.254
Au ₁₀	0.122	0.306	Au ₈	−0.061	−0.224
Au ₁₁	−0.061	−0.245	Au ₉	−0.061	−0.138
Au ₁₂	−0.061	−0.171			

bonding sites of the nucleobases¹²⁾ and the large average polarizabilities of both the bases and Au_3 cluster, being correspondingly equal to 92.5 (A), 79.1 (T), 98.6 (G), 73.9 (C) and 121.0 au. [The average polarizability α_{av} is defined as $\alpha_{av} \equiv (\alpha_{xx} + \alpha_{yy} + \alpha_{zz})/3$.] For comparison, the polarizability of a gold atom evaluated at the PW91/LANL2DZ computational level is 37 au [115a], in fair agreement with the early higher-level calculation yielding 39 au [115b]. An interesting example illustrating the large contribution of the electrostatic interactions to the stabilization of the base–gold complexes is provided by juxtaposing the complexes $T\cdot Au_3(O_2; N_1)$ (upper entry in Figure 8.14) and $C\cdot Au_3(O_2; N_1)$ (lower entry therein). These complexes are structurally similar in the sense of having the same structural unit. Nevertheless, $C\cdot Au_3(O_2; N_1)$ is energetically more favorable by 5.6 kcal mol^{−1} over $T\cdot Au_3(O_2; N_1)$, despite the fact that the hydrogen bond $N_1-H_1 \cdots Au_8$ of $C\cdot Au_3(O_2; N_1)$ is weaker; note that the H-bond lengths are 2.627 Å in $C\cdot Au_3(O_2; N_1)$ and 2.608 Å in $T\cdot Au_3(O_2; N_1)$. The stronger H-bonding of $T\cdot Au_3(O_2; N_1)$ originates from a positive difference of the deprotonation enthalpies of the N_1-H_1 groups of cytosine and thymine [113]: $DPE(N_1-H_1; C) - DPE(N_1-H_1; T) = 11.1$ kcal mol^{−1}.

Another feature of these complexes is that, in contrast, a gold cluster anchors more strongly at O₂ of $C\cdot Au_3(O_2; N_1)$ than of $T\cdot Au_3(O_2; N_1)$. This is a direct consequence of their bond lengths: 2.177 Å in $C\cdot Au_3(O_2; N_1)$ vs. 2.218 Å in $T\cdot Au_3(O_2; N_1)$. The stronger anchoring of gold at $C\cdot Au_3(O_2; N_1)$ mostly results from the following two factors. First: the polarity of C is higher than that of T, as is indicated by their dipole moments of 6.85 and 4.63 D, respectively (note, however, that a higher polarity of C is

12) For the adenine molecule, the magnitude of the electric field at N₁, N₃ and N₇ is 0.0781, 0.0797 and 0.0860 au, respectively. The electric field of thymine is 0.0030 au at N₁, 0.1150 au at O₂, 0.0054 au at N₃ and 0.1121 au at O₄. For guanine, the electric field reaches 0.0793 au at N₃, 0.1135 au at O₆ and finally 0.0838 au at N₇. In cytosine, the electric field distribution is

0.1121 au at O₂, 0.0783 au at N₃ and 0.0032 au at N₄. The electric fields at the N₁ and N₃ atoms of T and N₄ atom of C are very weak. The electric field strength at the atoms of the nucleobase definitely point out those gold-anchoring sites where the strong electrostatic energy contribution to the total binding energy is to be expected.

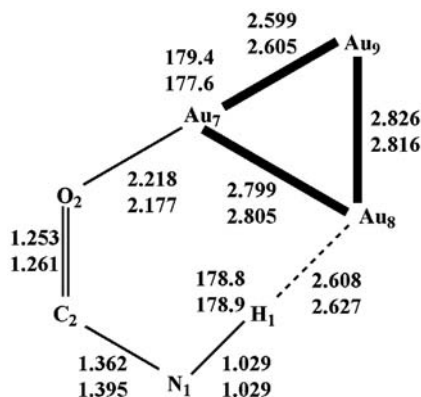


Figure 8.14 Comparison of the anchoring and nonconventional hydrogen bonding characteristics in $T \cdot Au_3(O_2; N_1)$ and $C \cdot Au_3(O_2; N_1)$. Upper entry: $T \cdot Au_3(O_2; N_1)$, lower entry: $C \cdot Au_3(O_2; N_1)$.

partially cancelled by a larger polarizability of T). The second factor is the most decisive. The dipole moment of C aligns almost along the C_2-O_2 bond anchoring the gold cluster with the bond angle $\angle C_2O_2Au_7 = 123.7^\circ$. The latter determines the strength of the bonding dipole-dipole interaction (a negative sign). The total dipole moment of T is approximately equal to a vector sum of the dipole moments of its two carbonyl bonds and approximately directed along the N_3-C_6 bond. With the dipole moment of the Au_7-Au_8 bond it forms an angle of about 40° , resulting in a positive sign for their mutual dipole-dipole interaction, which therefore exhibits a nonbonding (precisely, antibonding) character.

8.7.2

Energetics in $Z = 0$ Charge State

The most remarkable feature of the DNA base-gold interaction is evidently the energetics, which we first analyze for the particular cases of the gold atom and the triangular gold cluster, both in the $Z = 0$ charge state. The strongest complex of those studied is $C \cdot Au_3(N_3)$, which has a binding energy of $25.4 \text{ kcal mol}^{-1}$. A slightly weaker binding, $20.0 \text{ kcal mol}^{-1} \leq E_b \leq 24.4 \text{ kcal mol}^{-1}$, occurs in $A \cdot Au_3(N_3)$, $A \cdot Au_3(N_1)$, $A \cdot Au_3(N_7)$, $G \cdot Au_3(N_3; N_9)$, $G \cdot Au_3(N_3; N_2)$ and in $C \cdot Au_3(O_2; N_1)$. The latter series of complexes shows that the adenine base possesses the highest average affinity to gold, which when averaged over its four anchoring sites amounts to $19.8 \text{ kcal mol}^{-1}$. The guanine base has six anchoring sites and its average affinity to gold is $16.5 \text{ kcal mol}^{-1}$. The binding affinities to gold of thymine and cytosine, both having three anchoring sites, are correspondingly 12.5 and $18.9 \text{ kcal mol}^{-1}$. Therefore, with respect to a Au_3 cluster, the average binding affinities of the nucleobases are ordered as $A > C > G > T$. Note that thymine exhibits the lowest affinity to gold, in agreement with experimental data [94a,70k]. In addition, the purine bases A and G possess a larger number of anchoring sites, in contrast to the pyrimidine ones, C and

T, and therefore the purines are more strongly bonded to gold. In summary, the binding energies of the nucleobases with Au_3 over all anchoring sites lead to the inequality $G > A > C > T$, which correlates with the experimental data on the heats of desorption of the DNA bases from Au thin films [94a]. However, since, as noted in Section 8.1, the DNA bases interact with gold surfaces in a specific, sequence-dependent and rather complex manner [94–96] that likely involves multiple anchorings and different orientations of the nucleobases, not adequately described within the present model invoking the triangular cluster of gold, there is a certain disagreement between the calculated binding energies and the corresponding experimental data. Notice also that since the first ionization potential of a molecule measures its ability to donate the outermost electron the above inequality $G > A > C > T$ of the nucleobase affinities to gold correlates well with their electron donor ability expressed in terms of their first ionization potentials: $G(8.28) > A(8.48) > C(8.65) > T(9.18)$ (in eV; see, for example, Table 2 in Reference [116] and references therein).

The picture of the DNA base–gold interaction we offer in the present chapter would be incomplete without discussing it in terms of two factors that are typically invoked to explain the exceptional reactivity of small gold nanoparticles: a quantum size effect of the gold cluster and an effect of the low coordination of the gold atom [98c].

For this purpose, we examine two series of complexes, $\text{A} \cdot \text{Au}_{2 \leq n \leq 6}(\text{N}_3)$ and $\text{G} \cdot \text{Au}_{3 \leq n \leq 6}(\text{O}_6; \text{N}_1)$, with a Au–N and a Au–O anchoring, respectively. Their properties are summarized in Table 8.10 and Figures 8.15 and 8.16 [98b,c]. The binding energies of the series $\text{A} \cdot \text{Au}_{2 \leq n \leq 6}(\text{N}_3)$ vary from $19.1 \text{ kcal mol}^{-1}$ ($n = 2$) to $24.0 \text{ kcal mol}^{-1}$ ($n = 3$), reach a maximum of $28.8 \text{ kcal mol}^{-1}$ for $n = 4$ (T-shape gold cluster) and go down to $12.7 \text{ kcal mol}^{-1}$ ($n = 5$) and further to $10.9 \text{ kcal mol}^{-1}$ at $n = 6$ (notice that $E_b[\text{A} \cdot \text{Au}_1(\text{N}_3)] = 2.5 \text{ kcal mol}^{-1}$). A similar trend holds for the $\text{G} \cdot \text{Au}_{3 \leq n \leq 6}(\text{O}_6; \text{N}_1)$ series.

However, due to the weaker Au–O anchoring, $E_b[\text{G} \cdot \text{Au}_4^{\text{I}}(\text{O}_6; \text{N}_1)]$ is smaller than $E_b[\text{A} \cdot \text{Au}_4^{\text{I}}(\text{N}_3)]$ by $4.6 \text{ kcal mol}^{-1}$, and there is a sign of a plateau-like behavior of $E_b[\text{G} \cdot \text{Au}_{3 \leq n \leq 6}(\text{O}_6; \text{N}_1)]$ at $n = 5$ and 6 (at least within the studied series of gold clusters).

Since for both series, $\text{A} \cdot \text{Au}_{2 \leq n \leq 6}(\text{N}_3)$ and $\text{G} \cdot \text{Au}_{3 \leq n \leq 6}(\text{O}_6; \text{N}_1)$, the anchored gold atom is two-coordinated – the exception is $n = 5$ for $\text{A} \cdot \text{Au}_{2 \leq n \leq 6}(\text{N}_3)$ where it is three-coordinated – the trend in their binding energies can be attributed to a quantum size effect. Here we confine the treatment of a quantum size effect to the twofold gold coordination and to the gold clusters $\text{Au}_{1 \leq n \leq 6}$, and also exclude the aforementioned effect of multiple anchorings that may likely occur under the interaction of the nucleobases with larger gold clusters. The latter effect appears to be directly related with how effectively the LUMO of the Au_n cluster protrudes into the base [31, 111a] and how the eigenenergies of the HOMO of the base match the LUMO of Au_n . Obviously, the LUMO of the T-shape Au_4^{I} most effectively protrudes into the region of the adenine N_3 atom. It therefore forms the shortest anchor bond ($2.126 \text{ \AA}$) in the series shown in Figure 8.15, although the reinforcement of the anchor bond by the nonconventional H-bond that appears to be quite strong in $\text{A} \cdot \text{Au}_4^{\text{I}}(\text{N}_3)$ must also be taken into account.

Table 8.10 Key features of the planar A-Au₂₋₆(N₃) and G-Au₃₋₆(O₆:N₁) complexes with the N-H...Au nonconventional H-bond at the computational level B3LYP/RECP (Au) U6-31 + + G(d,p) (AUG). The notations are defined in the legend of Table 8.1.^{a)}

Complex	-E _b (kcal mol ⁻¹)	Anchor bond (Å)	ΔR(N-H) (Å)	r(H...Au) (Å)	∠N-H...Au (°)	-Δν(N-H) (cm ⁻¹)	R _{IR}
A-Au ₂ (N ₃)	19.1	2.154	0.003	3.054	102.0	44	1.1
A-Au ₃ (N ₃)	24.0	2.137	0.014	2.691	161.0	270	8.3
A-Au ₄ ^I (N ₃)	28.8	2.126	0.016	2.761	152.4	275	8.3
A-Au ₄ ^{II} (N ₃)	22.1	2.141	0.012	2.698	162.4	218	7.4
A-Au ₅ (N ₃)	12.7	2.184	0.013	2.644	160.0	254	10.3
A-Au ₆ (N ₃)	10.9	2.227	0.005	3.192	155.3	82	3.5
G-Au ₃ (O ₆ :N ₁)	18.4	2.185	0.016	2.568	173.6	324	13.5
G-Au ₄ ^I (O ₆ :N ₁)	24.2	2.157	0.009	2.826	177.2	172	13.2
G-Au ₅ (O ₆ :N ₁)	7.1	2.271	0.011	2.523	174.4	191	6.9
G-Au ₆ (O ₆ :N ₁)	7.2	2.289	0.009	2.877	173.9	183	12.1
				2.801	173.6	191	11.1

a) Relevant gold clusters have the following properties: (i) Au₂: r(Au₁-Au₂) = 2.566 Å, the electronic energy = -271.940755 hartree, ZPE = 0.239 kcal mol⁻¹; (ii) Au₄(C_{2h}): r(Au₁-Au₂) = r(Au₂-Au₃) = 2.759 Å, r(Au₁-Au₃) = 2.626 Å, r(Au₂-Au₄) = 2.573 Å, ∠Au₁Au₂Au₄ = 151.5°, electronic energy = -543.921072 hartree, ZPE = 0.788 kcal mol⁻¹; (iii) Au₄^{II}(D_{2h}): r(Au₁-Au₂) = r(Au₁-Au₃) = r(Au₂-Au₄) = r(Au₃-Au₄) = 2.741 Å, r(Au₂-Au₃) = 2.663 Å, electronic energy = -543.920660 hartree, ZPE = 0.819 kcal mol⁻¹. The energy difference between Au₄^I and Au₄^{II} amounts to only 0.3 kcal mol⁻¹. Properties of the most stable clusters Au₃ and Au₆ are summarized in References [18p, 14b].

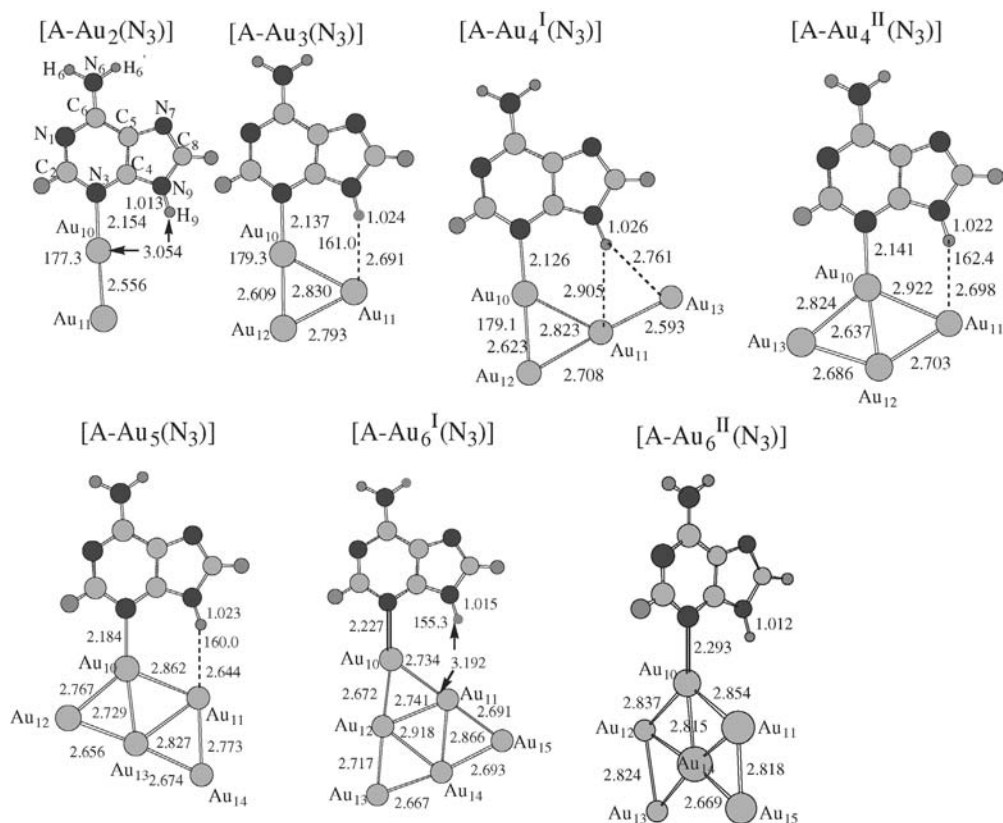


Figure 8.15 Complexes $A\cdot Au_{2\leq n\leq 6}(N_3)$. Bond lengths (Å) and bond angles ($^\circ$) are referred to the computational level B3LYP/RECP (Au) $\cup$ 6-31 + + G(d,p) (A). The structure of the gold cluster formed in the complex $A\cdot Au_6(N_3)$ is unstable in the neutral state [14b,18p]. The

energy difference between the $A\cdot Au_6^{II}(N_3)$ and $A\cdot Au_6^I(N_3)$ structures amounts to $21.1 \text{ kcal mol}^{-1}$. (Reproduced from Figures 1 and 2 of Reference [98c] with the permission from American Chemical Society.)

The strength of the nonconventional H-bond of $A\cdot Au_{3\leq n\leq 6}(N_3)$ is also strongly dependent on the coordination of the proton acceptor gold atom, that is, the strongest H-bond is formed with the singly-coordinated gold atom of Au_4^I while the ones formed with the two-coordinated atom of Au_3 and Au_4^{II} are weaker. The weakest nonconventional hydrogen bond exists with the three-coordinated gold of Au_5 , and none with the four-coordinated Au in Au_6 , as indicated by the fact that H-bond distance in $A\cdot Au_6(N_3)$ (3.19 Å) lies far beyond the van der Waals cutoff (see condition 4 in Section 8.3.2.1). Note that the effect of the anchor-H-bond reinforcement is stronger in the complex $G\cdot Au_4^I(O_6; N_1)$, which is stabilized by two nonconventional hydrogen bonds, instead of a single one that occurs in the DNA base- Au_3 complexes. However, these two nonconventional H-bonds are weaker than the H-bond formed in the complex $G\cdot Au_3(O_6; N_1)$.

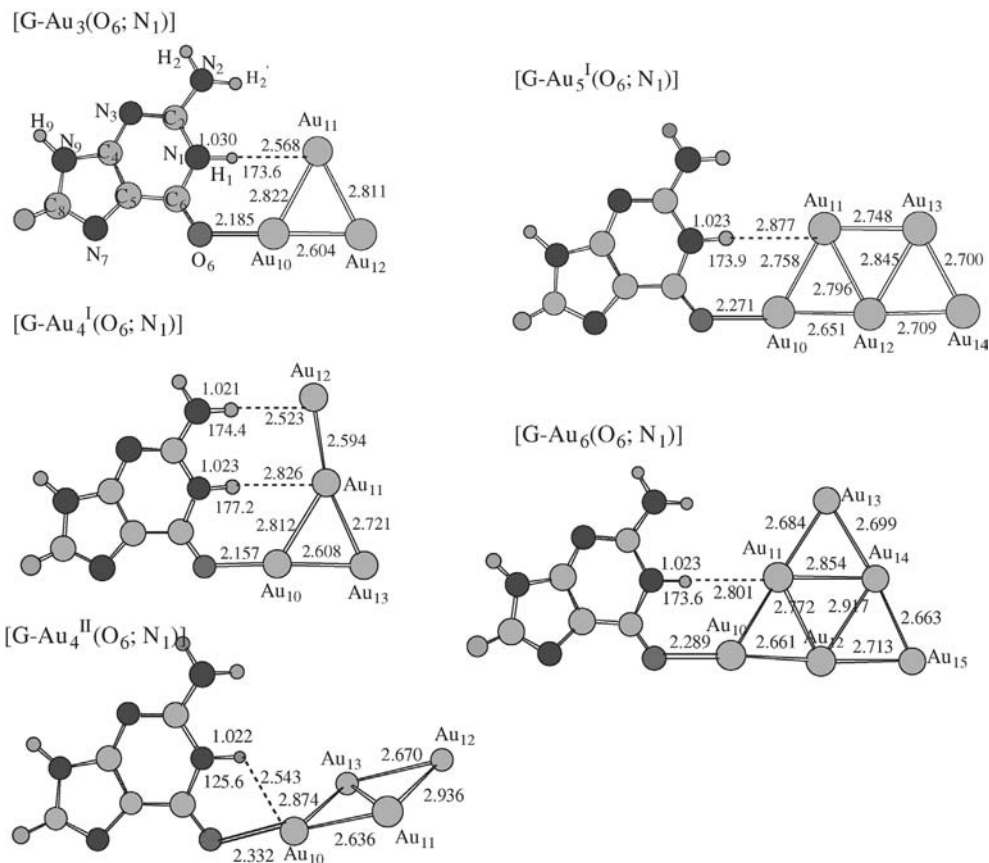


Figure 8.16 Complexes $G\text{-Au}_{3 \leq n \leq 6}(\text{O}_6; \text{N}_1)$. Bond lengths (Å) and bond angles (°) are referred to the computational level B3LYP/RECP (Au) $\cup$ 6-31 + + G(d,p) (G). (Reproduced from Figure 3 of Reference [98c] with the permission from the American Chemical Society.)

8.7.3

Z = −1 Charge State

As emphasized in Section 8.1.1, the charge state of gold NPs can also be the decisive factor in their exceptional reactivity. For this reason, below, we only consider the charge state $Z = -1$ of the complexes of the DNA bases and the gold atom based on previous work [98a,117] since their cationic state $Z = +1$ has been studied rather incompletely and limited to the treatment of the hybridization of Au^+ with the C·A DNA base pair [118] and with the RNA base uracil [117b].

The bonding scenarios that yield the anionic complexes $[\text{DNA base}\cdot\text{Au}]^-$ are collected in Figure 8.17 [98a]. Since the computational electron affinity of the gold atom is high (see References [12, 17]), it is the gold atom of $[\text{DNA}\cdot\text{base}\cdot\text{Au}]^-$ that hosts the most excess electron charge. This is witnessed by the Mulliken charges of gold and therefore, as anticipated, the gold atom mainly exists in $[\text{DNA}\cdot\text{base}\cdot\text{Au}]^-$ as the auride anion Au^- . The latter hence acts as a strong proton acceptor: this can readily be

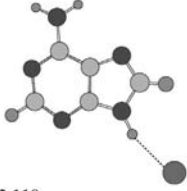
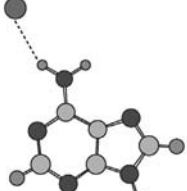
Adenine $\Gamma^{-1} \Rightarrow \Gamma^0$  VDE = 3.110 ADE^{ZPE} = 2.643 $R(N_9-H_9) = 1.053$ $\Delta R(N_9-H_9) = 0.042$ $r(H_9 \cdots Au) = 2.313$ $\angle N_9 H_9 Au = 157.6$ $\nu(N_9-H_9) = 2919$ (2957) $\Delta \nu(N_9-H_9) = 725$ $-E_b^{ZPE} = 14.38$	$\Pi^{-1} \Rightarrow \Pi^0$ and IV^0  VDE = 2.705 ADE^{ZPE} = 2.263 $R(N_6-H_6) = 1.034$ $\Delta R(N_6-H_6) = 0.023$ $r(H_6 \cdots Au) = 2.502$ $\angle N_6 H_6 Au = 146.3$ $\nu(N_6-H_6) = 3232$ (1943) $\Delta \nu(N_6-H_6) = 364$ $\Delta E_{I-II}^{ZPE} (Z=-1) = 8.80$
$III^{-1} \Rightarrow III^0$  VDE = 2.693 ADE^{ZPE} = 2.254 $R(N_9-H_9) = 1.029$ $\Delta R(N_9-H_9) = 0.018$ $r(H_9 \cdots Au) = 2.581$ $\angle N_9 H_9 Au = 137.6$ $\nu(N_9-H_9) = 3321$ (1373) $\Delta \nu(N_9-H_9) = 375$ $\Delta E_{I-III}^{ZPE} (Z=-1) = 9.29$	
Guanine $\Gamma^{-1} \Rightarrow III^0$ and IV^0	$\Pi^{-1} \Rightarrow \Pi^0$

Figure 8.17 Computational bonding scenarios between the auride anion and the DNA bases. The vertical detachment energies, VDE, and adiabatic detachment energy, ADE, are given in eV.¹³⁾ The ZPE-corrected binding energies E_b^{ZPE} and the energy differences are given in kcal mol⁻¹, R

(N–H) and $r(H \cdots Au)$ are in Å, $\angle N-H-Au$ in degrees, and $\nu(N-H)$ in cm⁻¹. The latter is accompanied, in parentheses, by the IR intensity in km·mol⁻¹. The reference asymptote for the complex [DNA base·Au]⁻¹ consists of the infinitely separated Au⁻¹ and the corresponding DNA base.

- 13) Consider a given anionic molecular complex M^{-1} in the anionic charge state $Z = -1$. M^{-1} accesses, directly or indirectly, the ground electronic state of the neutral M^0 , when an excess electron is photodetached from M^{-1} . The electron vertical detachment energy, VDE (or VEDE), is defined as $VEDE[M^{-1}] := E(M^0|G_M^{-1}) - E(M^{-1}|G_M^{-1})$, the energy difference – without the ZPE – between the anionic M^{-1}

and neutral complex M^0 , both taken in the anionic equilibrium geometry G_M^{-1} . The electron adiabatic detachment energy $ADE[M^{-1}] := E(M^0|G_M^0) - E(M^{-1}|G_M^{-1})$. The charge alternation $Z = -1$ is mapped onto $Z = 0$: $Z = -1 \Rightarrow Z = 0$. This charge alternation induces the mapping between the conformational manifolds C_M^{-1} and C_M^0 of M^{-1} and M^0 , respectively, $C_M^{-1} \Rightarrow C_M^0$ where $M^0 = [DNA\ base \cdot Au]$.

seen by juxtaposing the binding energies, for example, with the water dimer. The conformational manifolds $C_{\text{DNA base-Au}}^{-1}$ consist of three conformers of $[\text{A}\cdot\text{Au}]^{-}$ and of $[\text{G}\cdot\text{Au}]^{-}$, and two conformers of $[\text{T}\cdot\text{Au}]^{-}$ and of $[\text{C}\cdot\text{Au}]^{-}$. We thus conclude:

- Such a bonding mechanism predetermines a rather large absolute value of binding energies E_b^{ZPE} , which is typical for the medium and modest ionic conventional hydrogen bonds. As demonstrated in Figure 8.17, E_b^{ZPE} ranges from 6 to almost 20 kcal mol⁻¹. The order of stability of the DNA bases with respect to $[\text{Au}\cdot\text{DNA base}]^{-}$ is $\text{G} > \text{T} > \text{C} > \text{A}$.
- The auride anion is a strong proton acceptor that while interacting with the DNA base significantly perturbs it. This perturbation is manifest in several ways. One is spectroscopic – as a significant redshift that reaches 6–8-hundred wavenumbers.

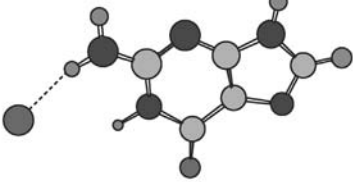
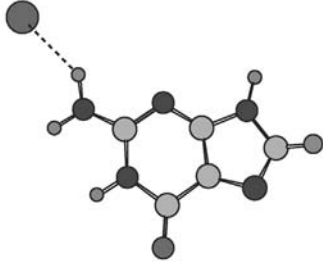
 VDE = 3.302 ADE^{ZPE} = 3.153 $R(\text{N}_2\text{-H}_2') = 1.043$ $\Delta R(\text{N}_2\text{-H}_2') = 0.031$ $r(\text{H}_2'\cdots\text{Au}) = 2.314$ $\angle \text{N}_2\text{H}_2'\text{Au} = 165.5$ $\nu(\text{N}_2\text{-H}_2') = 3317$ (890) $\Delta \nu(\text{N}_2\text{-H}_2') = 245$ $-E_b^{\text{ZPE}} = 19.84$	 VDE = 3.131 ADE^{ZPE} = 2.789 $R(\text{N}_9\text{-H}_9) = 1.056$ $\Delta R(\text{N}_9\text{-H}_9) = 0.045$ $r(\text{H}_9\cdots\text{Au}) = 2.285$ $\angle \text{N}_9\text{H}_9\text{Au} = 165.3$ $\nu(\text{N}_9\text{-H}_9) = 2858$ (3233) $\Delta \nu(\text{N}_9\text{-H}_9) = 782$ $\Delta E_{\text{I-II}}^{\text{ZPE}} (\text{Z}=-1) = 5.07$
III⁻¹ ⇒ I⁰  VDE = 3.049 ADE^{ZPE} = 2.662 $R(\text{N}_2\text{-H}_2) = 1.039$ $\Delta R(\text{N}_2\text{-H}_2) = 0.027$ $r(\text{H}_2\cdots\text{Au}) = 2.440$ $\angle \text{N}_2\text{H}_2\text{Au} = 143.5$ $\nu(\text{N}_2\text{-H}_2) = 3160$ (1923) $\Delta \nu(\text{N}_2\text{-H}_2) = 402$ $\Delta E_{\text{I-III}}^{\text{ZPE}} (\text{Z}=-1) = 6.75$	
Thymine I⁻¹ ⇒ I⁰	II⁻¹ ⇒ II⁰ and III⁰

Figure 8.17 (Continued)

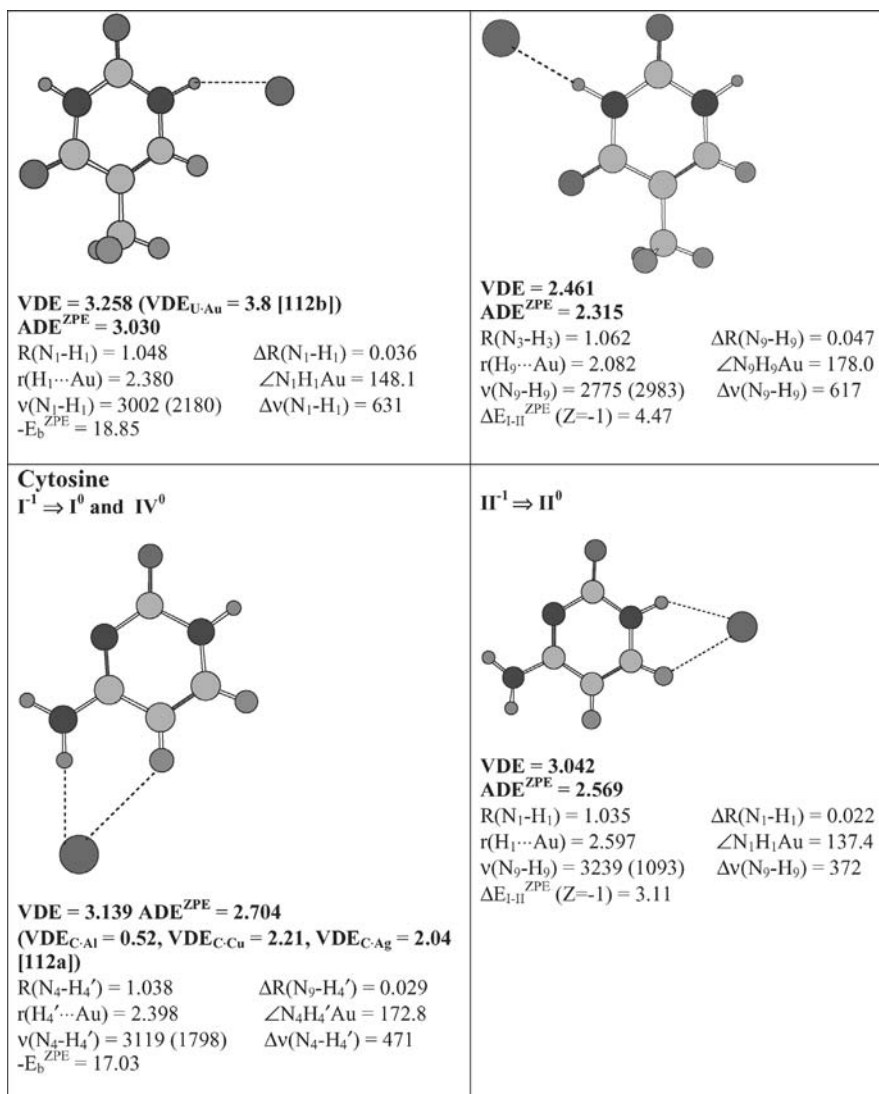


Figure 8.17 (Continued)

- Since the gold atom is the key carrier of the excess electron charge of the complexes [DNA-base·Au]⁻, removal of this charge, formally implying the alternation $Z = -1 \rightarrow Z = 0$ of the charge states, converts the auride anion into the neutral gold atom and often causes the essential structural changes, mainly provided by migration of Au from the nonconventional proton acceptor location in the -1 -charge state to the anchoring location in the 0 -charge state where the gold atom forms, as demonstrated in subsection 8.7.1, the gold–base anchoring bond of the Au–O or Au–N type. For example, Au migration from II⁻¹ to II⁰ occurs over 4 Å.

Despite the high electron affinity of the latter, Au may only induce a small charge transfer from the adjacent oxygen or nitrogen atom as a result of the formation either of Au-O or Au-N anchoring bond. This anchoring bond is very weak, as reflected in the corresponding binding energies. The shortest Au-O and Au-N anchoring bonds of 2.453 and 2.305 Å are correspondingly formed in conformer III⁰ of [G·Au] and in conformer III⁰ of [A·Au]. Since the anchoring interaction is weak, relaxation of the DNA base within [DNA base·Au] is not significant, in contrast to the anionic charge state.

- The conformational manifold C_M^0 consists of four conformers of [A·Au], [G·Au] and [C·Au], and three conformers of [T·Au]. This is in the contrast with C_M^{-1} , which are characterized by $|C_{A-Au}^0| = |C_{G-Au}^0| = 3$ and $|C_{C-Au}^0| = |C_{T-Au}^0| = 2$. Hence, in general, the mapping $|C_M^{-1} \Rightarrow C_M^0$ is not a one-to-one: for example, the first excited state II⁻¹ of [Au·A]⁻¹ is mapped onto the states II⁰ and IV⁰ of [A·Au]⁰. In general, the mapping $|C_M^{-1} \Rightarrow C_M^0$ is not energy preserving: for example, the ground state I⁻¹ of [G·Au]⁻¹ is mapped onto the states III⁰ and IV⁰ of [G·Au]⁰ and the ground state I⁻¹ of [C·Au]⁻¹ is mapped onto the states I⁰ and IV⁰ of [G·Au]⁰. However, for the DNA bases A and T, the ground state I⁻¹ is solely mapped onto the state I⁰. This implies that, after the excess electron is photodetached, the ground-state anions [A·Au]⁻ and [T·Au]⁻ directly access the ground-state neutrals [A·Au] and [T·Au], as anticipated in the experiments on anion photoelectron spectroscopy, with the VDE = 3.110 and 3.258 eV. The ground-state anion [G·Au]⁻ directly accesses only the second and third excited-state of the neutral complex [G·Au] with the VDE 3.302 eV, whereas the ground-state anion [C·Au]⁻ directly accesses the mixture state that is composed of the ground- and third excited-state neutrals [C·Au] with the VDE = 3.139 eV. To access the bottom of the PES⁰ of [G·Au], one has to form III⁻¹. The mapping $|C_M^{-1} \Rightarrow C_M^0$ is undefined on the conformer III⁰ of [C·Au]⁰, which was discussed in Section 8.6. This [C·Au]⁰-conformer III⁰ fragments under the electron attachment into Au⁻ and C.

8.8

Interaction of Watson–Crick DNA Base Pairs with Gold Clusters

8.8.1

General Background

In Sections 8.3–8.7 we have shown that the anchoring and nonconventional N–H···Au hydrogen bondings are actually two fundamental interactions governing the hybridization between the nucleobases and gold clusters (see also the recent work on this theme [89h, 95g j, 96e,f, 119]). The formation of these bonds may quite drastically modify the electron density of the nucleobases, particularly on those nitrogen and oxygen atoms involved in the intermolecular H-bonds with the Watson–Crick (WC) complementary ones [104]. Since the strength of the WC interbase pairing is strongly determined by the proton affinities (PA) of the proton acceptor and the DPEs of the proton donor groups of both complementary bases, we

investigate herein the effect of the base–gold interaction on these PAs and DPEs and attempt to rationalize it [98c].

Let us first consider the WC A·T pair. As known, it is hybridized via the two conventional intermolecular hydrogen bonds $N_6-H_6(A) \cdots O_4(T)$ and $N_3-H_3(T) \cdots N_1(A)$. According to Table 8.11, the Au_3 anchorings at the ring atoms N_3 and N_7 of

Table 8.11 Mulliken charges, PAs and DPEs of the DNA bases and base-gold complexes.

Adenine				
Property	A	A·Au ₃ (N ₃)	A·Au ₃ (N ₆)	A·Au ₃ (N ₇)
$q^M(N_1)$ ($ e $)	−0.381	−0.325	−0.246	−0.339
$q^M(C_2)$ ($ e $)	0.074	0.118	0.020	0.016
$q^M(N_6)$ ($ e $)	−0.826	−0.822	−1.059	−0.804
PA(N ₁) (kcal mol ^{−1})	222.1	208.3		
DPE(N ₆ ·H ₆) (kcal mol ^{−1})	353.0	331.9		
Thymine				
Property	T	T·Au ₃ (O ₂ ;N ₁)	T·Au ₃ (O ₂ ;N ₃)	T·Au ₃ (O ₄)
$q^M(O_2)$ ($ e $)	−0.536	−0.520	−0.539	−0.457
$q^M(N_3)$ ($ e $)	−0.725	−0.701	−0.634	−0.653
$q^M(O_4)$ ($ e $)	−0.511	−0.458	−0.447	−0.479
PA(O ₄) (kcal mol ^{−1})	202.2	198.2		
DPE(N ₃) (kcal mol ^{−1})	343.3	320.2		
Guanine				
Property	G	G·Au ₃ (N ₂)	G·Au ₃ (N ₃ ;N ₂)	G·Au ₃ (N ₃ ;N ₉)
$q^M(N_1)$ ($ e $)	−0.709	−0.606	−0.679	−0.658
$q^M(N_2)$ ($ e $)	−0.769	−1.118	−0.765	−0.769
$q^M(O_6)$ ($ e $)	−0.547	−0.463	−0.468	−0.471
PA(O ₆) (kcal mol ^{−1})	219.2			209.1
DPE(N ₂ ·H ₂ ') (kcal mol ^{−1})	334.7			312.8
DPE(N ₁ ·H ₁) (kcal mol ^{−1})	335.3			313.4
Property	G	G·Au ₃ (O ₆ ;N ₁)	G·Au ₃ (O ₆ ;N ₇)	G·Au ₃ (N ₇)
$q^M(N_1)$ ($ e $)	−0.709	−0.614	−0.650	−0.684
$q^M(N_2)$ ($ e $)	−0.769	−0.773	−0.745	−0.745
$q^M(O_6)$ ($ e $)	−0.547	−0.475	−0.562	−0.473
Cytosine				
Property	C	C·Au ₃ (O ₂ ;N ₁)	C·Au ₃ (N ₄)	
$q^M(O_2)$ ($ e $)	−0.531	−0.480	−0.467	
$q^M(N_3)$ ($ e $)	−0.487	−0.450	−0.372	
$q^M(N_4)$ ($ e $)	−0.803	−0.820	−1.088	
PA(N ₃) (kcal mol ^{−1})	225.0	215.0		
DPE(N ₄) (kcal mol ^{−1})	351.2	332.3		

adenine reduce the Mulliken electron charge $q^M(N_6)$ on N_6 by 0.004 and 0.022 $|e|$, respectively, and as a result the N_6-H_6 bond weakens; that is, its DPE decreases. Additionally, $q^M(N_1)$ decreases by 0.056 and 0.042 $|e|$, respectively, which simply implies that $PA(N_1)$ is lowered too. Similarly, activation of the N_3-H_3 group of T by the Au_3 -anchoring either at O_2 or O_4 reduces $q^M(N_3)$ by 0.024 and 0.072 $|e|$ respectively, which in turn yields a lower $DPE(N_3-H_3; T)$. These two anchorings also weaken the $PA(O_4; T)$ since they decrease $q^M(O_4)$ by 0.053 and 0.032 $|e|$, respectively. On the other hand, a weaker, non-planar coordination of Au_3 to adenine at the amino group likely strengthens the N_6-H_6 bond, although it also reduces the $PA(N_1; A)$.

To verify our above observations, which simply rely on the Mulliken analysis, we examine four representative complexes, the protonated $AH_1^+ \cdot Au_3(N_3)$ and $TH_4^+ \cdot Au_3(O_2; N_1)$ and the deprotonated $AH_6^- \cdot Au_3(N_3)$ and $TH_3^- \cdot Au_3(O_2; N_1)$. Table 8.11 summarizes their relevant properties, from which we arrive at the following key features:

- 1) There exists an overall reduction of the DPEs and PAs caused by the bonding to Au_3 – (i) $DPE[N_6; A \cdot Au_3(N_3)]$ and $DPE[N_3; T \cdot Au_3(O_2; N_1)]$ are lowered by 21.1 and 23.1 kcal mol^{-1} , respectively, compared to the corresponding DPEs of A and T; (ii) $PA[N_1; A \cdot Au_3(N_3)]$ and $PA[O_4; T \cdot Au_3(O_2; N_1)]$ are smaller by 13.8 and 4.0 kcal mol^{-1} with respect to $PA(N_1; A)$ and $PA(O_4; T)$. Since the strength of hydrogen bonding depends more on the proton affinity than the deprotonation energy, we might expect that two simultaneous anchorings of Au_3 clusters at N_3 of A and at $O_2(N_1 \text{ side})$ of T strengthen one interbase hydrogen bond, $N_6-H_6(A) \cdots O_4(T)$, and weaken the other, $N_3-H_3(T) \cdots N_1(A)$.
- 2) While the deprotonation of A and T strengthens the gold interaction with these nucleobases by factor of 2–3, their protonation, conversely, weakens it.

The proposed picture of how the base deprotonation and protonation affect its interaction with a gold cluster is, however, rather crude. It can be summarized as follows: (i) the deprotonation strengthens the anchoring bond and significantly weakens the nonconventional H-bond; (ii) the effect of protonation is opposite, that is, it considerably strengthens the nonconventional hydrogen bond so that the latter even exhibits all features of the moderate-ionic one [with the redshifts reaching 542 cm^{-1} in $AH_1^+ \cdot Au_3(N_3)$ and 861 cm^{-1} in $TH_4^+ \cdot Au_3(O_2; N_1)$] and weakens the anchoring Au–N and Au–O bonds.

The WC G·C base pair is formed via the three conventional intermolecular hydrogen bonds $N_4-H_4(C) \cdots O_6(G)$, $N_1-H_1(G) \cdots N_3(C)$ and $N_2-H_2(G) \cdots O_2(C)$ [104] (see also Reference [112a]). All the information needed to estimate the effect of the gold interaction on the PAs and DPEs of the involved proton donors and acceptors is collected in Table 8.11. As found for A and T, the gold anchoring decreases the Mulliken charges on the N_1 and N_2 atoms of G [see Table 8.11, except for the weak and non-planar complex $G \cdot Au_3(N_2)$], which in turn lowers their DPEs. $DPE[N_1-H_1; G \cdot Au_3(N_3; N_9)]$ and $DPE[N_2-H_2'; G \cdot Au_3(N_3; N_9)]$ are smaller than the corresponding DPEs of G by 21.9 kcal mol^{-1} . Notice that the deprotonated complexes $GH_1^- \cdot Au_3(N_3; N_9)$ and $GH_2'^- \cdot Au_3(N_3; N_9)$ exhibit a very strong binding,

of about 43 kcal mol^{-1} , due to a substantial shortening of their anchoring $\text{Au}_{10}\text{--N}_3$ bonds, as compared to that of $\text{G}\cdot\text{Au}_3(\text{N}_3; \text{N}_9)$. The gold anchoring also weakens the PA ($\text{O}_6; \text{G}$), for example, by $10.1 \text{ kcal mol}^{-1}$ for the complex $\text{G}\cdot\text{Au}_3(\text{N}_3; \text{N}_9)$ whose $\text{H}_6\text{--}$ protonation converts the weak nonconventional $\text{N}_9\text{--H}_9\cdots\text{Au}_{11}$ H-bond into the moderate one (Table 8.3). The PA($\text{N}_3; \text{C}$) of the complex $\text{C}\cdot\text{Au}_3(\text{O}_2; \text{N}_1)$ is almost equally reduced. Its protonated analog, $\text{CH}_3^+\cdot\text{Au}_3(\text{O}_2; \text{N}_1)$, exhibits a rather strong moderate-type nonconventional $\text{N}_1\text{--H}_1\cdots\text{Au}_8$ hydrogen bond showing a significant contraction of the $\text{N}_1\text{--H}_1$ bond by $0.042 \text{ \AA}$ and a redshift of $\nu(\text{N}_1\text{--H}_1)$ equal to 786 cm^{-1} . The H_4' -deprotonation of $\text{C}\cdot\text{Au}_3(\text{O}_2; \text{N}_1)$ lowers the $\text{DPE}[\text{N}_4; \text{C}\cdot\text{Au}_3(\text{O}_2; \text{N}_1)]$ by $18.9 \text{ kcal mol}^{-1}$ with respect to $\text{DPE}(\text{N}_4; \text{C})$. These are the general rules that govern the changes of the WC interbase hydrogen bonds in the A·T and G·C base pairs under their anchoring to gold.

8.8.2

[A·T]·Au₃ Complexes

Some of the bonding patterns formed between the triangular gold cluster Au_3 and the WC A·T pair are shown in Figures 8.18 and 8.19. When interacting with the WC A·T pair, Au_3 changes the WC intermolecular H-bonding pattern in a rather complex manner, the general trend being a weakening of the WC A·T pairing. This effect is easily understood by considering the most stable complex $[\text{A}\cdot\text{Au}_3(\text{N}_3)]\cdot\text{T}$, whose binding energy, taken relative to the infinitely separated A·T and Au_3 , amounts to $19.6 \text{ kcal mol}^{-1}$. According to Table 8.1, this is $4.8 \text{ kcal mol}^{-1}$ lower than the binding energy of the isolated adenine molecule anchoring Au_3 at N_3 . This loss is the result of either a weaker bonding of Au_3 to A within the WC A·T pair or a weakening of the WC pairing, or both. Regarding the former assumption, Table 1 clearly shows that the anchoring and nonconventional H-bonds of $[\text{A}\cdot\text{Au}_3(\text{N}_3)]\cdot\text{T}$ and $\text{A}\cdot\text{Au}_3(\text{N}_3)$ are almost identical, the difference being that the complex $[\text{A}\cdot\text{Au}_3(\text{N}_3)]\cdot\text{T}$ possesses a slightly more elongated (by $0.009 \text{ \AA}$) H-bond $\text{H}_9\cdots\text{Au}_{11}$, resulting in a smaller redshift of its $\nu(\text{N}_9\text{--H}_9)$ stretch (by 6 cm^{-1}). Therefore, the difference in the binding energies is likely to originate from a net weakening of the WC A·T intermolecular H-bonding resulting from the binding of Au_3 at $\text{N}_3(\text{A})$ within the A·T pair.

In geometrical terms, the weakening of the central intermolecular H-bond $\text{N}_3\text{--H}_3(\text{T})\cdots\text{N}_1(\text{A})$ of $[\text{A}\cdot\text{Au}_3(\text{N}_3)]\cdot\text{T}$ with respect to that of A·T is manifested by a shortening of the $\text{N}_3\text{--H}_3$ bond by $0.007 \text{ \AA}$ (which, however, elongates by $0.022 \text{ \AA}$ compared with T) and by a lengthening of the H-bond $\text{H}_6\cdots\text{N}_1$ by $0.034 \text{ \AA}$. The blue-shift of the $\text{N}_3\text{--H}_3$ stretch by 119 cm^{-1} and the weakening of its IR intensity from 1821 to 1631 km mol^{-1} (Table 8.12) is a spectroscopic indicator of such an effect. The above changes in $\text{N}_3\text{--H}_3(\text{T})\cdots\text{N}_1(\text{A})$ are consistent with the physical picture offered in the previous subsection and largely originate from a lowering of the PA(N_1) of adenine under the anchoring of a gold cluster (Table 8.11).

Another intermolecular H-bond, $\text{N}_6\text{--H}_6(\text{A})\cdots\text{O}_4(\text{T})$ of $[\text{A}\cdot\text{Au}_3(\text{N}_3)]\cdot\text{T}$, is, however, strengthened. This is indicated by its stronger directionality ($\Delta\angle\text{N}_6\text{H}_6\text{O}_4 = 2.7^\circ$), an increase of $R(\text{N}_6\text{--H}_6)$ by $0.003 \text{ \AA}$ and a contraction of $r(\text{H}_6\cdots\text{O}_4)$ by $0.027 \text{ \AA}$.

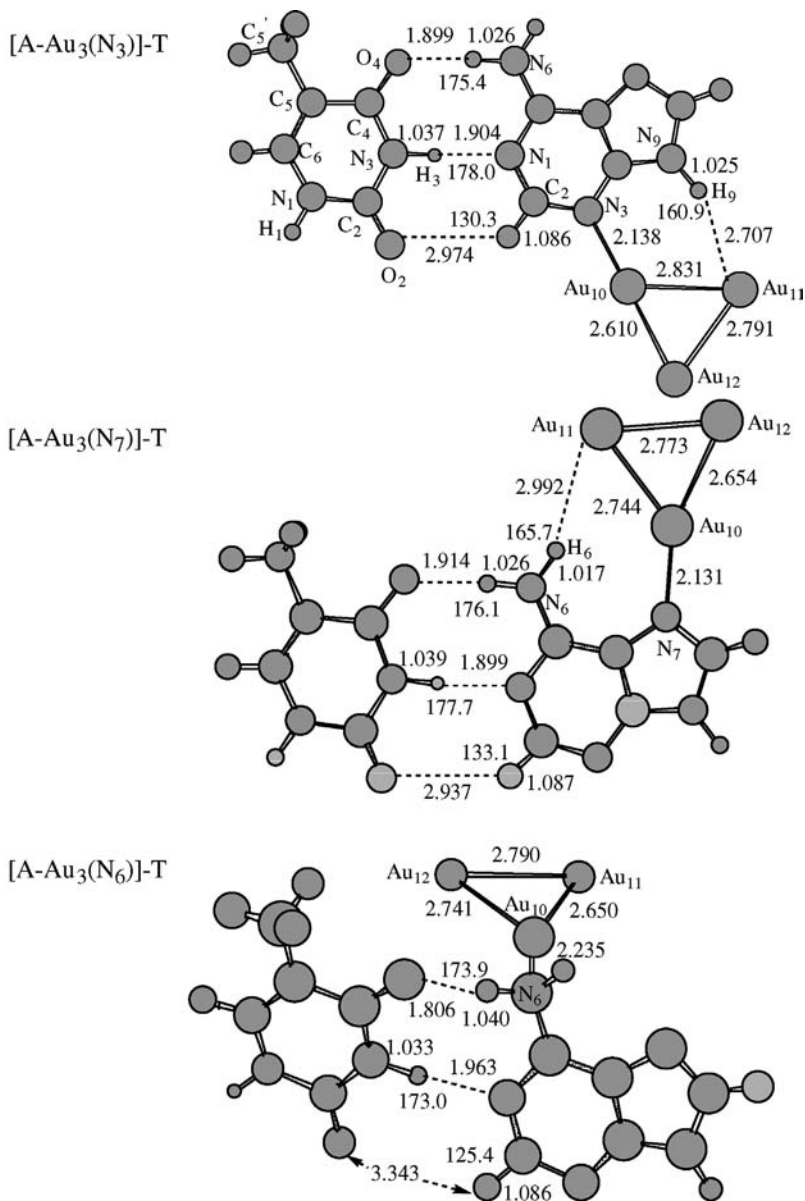


Figure 8.18 Stable [A-Au₃]-T pairs. The WC intermolecular H-bonds of the A-T pair are characterized by the following geometrical parameters: $R[\text{N}_6\text{-H}_6(\text{A})] = 1.023 \text{ \AA}$, $r[\text{H}_6(\text{A}) \cdots \text{O}_4(\text{T})] = 1.926 \text{ \AA}$, $\angle \text{N}_6\text{H}_6(\text{A})\text{O}_4(\text{T}) = 174.1^\circ$; $R[\text{N}_3\text{-H}_3(\text{T})] = 1.044 \text{ \AA}$, $r[\text{H}_3(\text{T}) \cdots \text{N}_1(\text{A})] = 1.822 \text{ \AA}$, $\angle \text{N}_3\text{H}_3(\text{T})\text{N}_1(\text{A}) = 178.5^\circ$; $R[\text{C}_2\text{-H}_2(\text{A})] = 1.087 \text{ \AA}$, $r[\text{H}_2(\text{A}) \cdots \text{O}_2(\text{T})] = 2.937 \text{ \AA}$, $\angle \text{C}_2\text{H}_2(\text{A})\text{O}_2(\text{T}) = 131.9^\circ$. Bond lengths are given in $\text{\AA}$ and bond angles in degrees.

(Reproduced from Figure 4 of Reference [98c] with the permission from the American Chemical Society.)

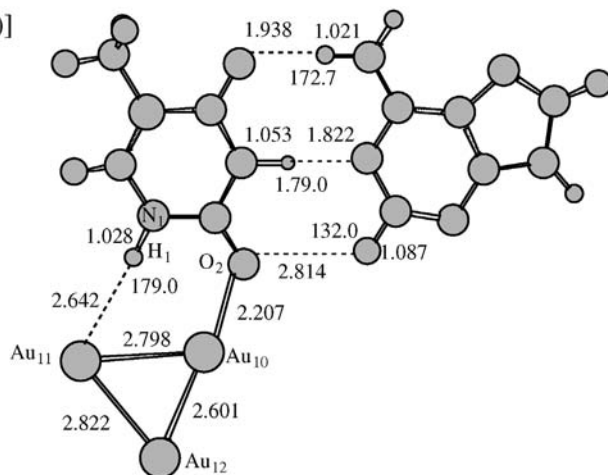
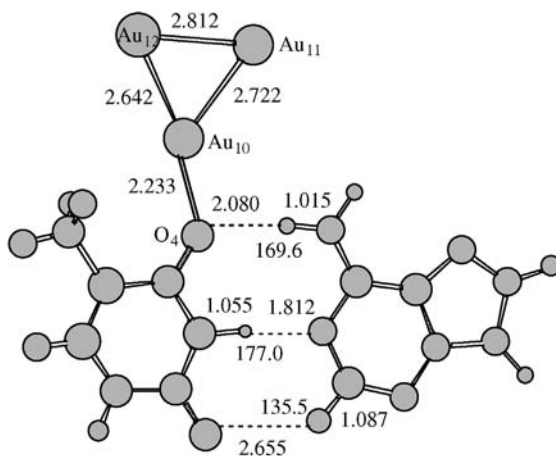
A-[T-Au₃(O₂)]A-[T-Au₃(O₄)]

Figure 8.19 Stable A-[T-Au₃] pairs. Bond lengths are given in Å and bond angles in degrees. (Reproduced from Figure 5 of Reference [98b] with the permission from the American Chemical Society.)

Mirroring these geometrical changes, the $\nu(\text{N}_6-\text{H}_6)$ stretch undergoes a redshift by 45 cm^{-1} (Table 8.12). The way the H-bond $\text{N}_6-\text{H}_6(\text{A}) \cdots \text{O}_4(\text{T})$ is perturbed is due to the lowering of the $\text{DPE}(\text{N}_6; \text{A})$ while A anchors Au₃ to form A·Au₃(N₃), provided that this Au₃-binding does not influence the $\text{PA}(\text{O}_4)$ and $\text{DPE}(\text{N}_3)$ of T. Finally, the very weak H-bond $\text{C}_2-\text{H}_2(\text{A}) \cdots \text{O}_2(\text{T})$ that lies in close to the anchoring $\text{Au}_{10}-\text{N}_3(\text{A})$ bond is weakened too, as indicated by the elongation of its $r(\text{H}_2 \cdots \text{O}_2)$ distance by 0.037 Å and the blue-shift by 26 cm^{-1} of its C_2-H_2 stretch.

Table 8.12 Stretching vibrational modes (in cm^{-1} ; IR activity in km mol^{-1} , in parentheses) of the WC intermolecular hydrogen bonds. The asterisk indicates the mode coupling within the NH_2 group.

Base pair	$\text{N}_3\text{--H}_3(\text{T}) \cdots \text{N}_1(\text{A})$	$\text{N}_6\text{--H}_6(\text{A}) \cdots \text{O}_4(\text{T})$	$\text{C}_2\text{--H}_2(\text{A}) \cdots \text{O}_2(\text{T})$
A·T	3062 (1821)	3420 (1042)	3206 (4)
[A·Au ₃ (N ₃)]·T	3181 (1631)	3375 (1701)	3232 (≈ 0)
[A·Au ₃ (N ₆)]·T	3264 (1433)	3173 (658)	3222 (7)
[A·Au ₃ (N ₇)]·T	3157 (1524)	3374 (1029)*	3211 (6)
A·[T·Au ₃ (O ₂ ;N ₁)]	2915 (2618)	3448 (879)	3211 (5)
A·[T·Au ₃ (O ₄)]	2875 (2374)	3543 (502)*	3215 (2)
Base pair	$\text{N}_1\text{--H}_1(\text{G}) \cdots \text{N}_3(\text{C})$	$\text{N}_4\text{--H}_4(\text{C}) \cdots \text{O}_6(\text{G})$	$\text{N}_2\text{--H}_2(\text{G}) \cdots \text{O}_2(\text{C})$
G·C	3253 (1759)	3195 (558)	3405 (1252)
[G·Au ₃ (N ₃ ;N ₉)]·C	3173 (816)	3276 (792)	3323 (2864)
[G·Au ₃ (N ₇)]·C	3172 (794)	3293 (1206)	3363 (1799)
G·[C·Au ₃ (O ₂ ;N ₁)]	3314 (1474)	3154 (1570)	3505 (898)
[G·Au ₃ (O ₆)]·C	3146 (883)	3429 (1163)	3336 (981)
[G·Au ₃ (N ₂)]·C	3069 (670)	3313 (1177)	3143 (1185)
G·[C·Au ₃ (N ₄)]	3334 (993)	3001 (1871)	3464 (744)
G·[C·Au ₃ (N ₃)]	3495 (13)	3261 (826)	3512 (652)
[G·C]·Au ₆	3305 (855)	3235 (166)*	3409 (713)*
		3237 (438)*	3518 (362)*

The general trend of a net weakening of the WC A·T pairing by at least 4 kcal mol^{-1} as a consequence of the Au₃-binding holds for the rest of the studied complexes, [A·Au₃(N₇)]·T, A·[T·Au₃(O₂; N₁)], [A·Au₃(N₆)]·T and A·[T·Au₃(O₄)], displayed in Figures 8.18 and 8.19. They are characterized by smaller binding energies, 16.7, 9.9, 5.9 and $3.5 \text{ kcal mol}^{-1}$, respectively, than the [A·Au₃(N₃)]·T complex discussed above. In contrast to [A·Au₃(N₃; N₉)]·T, the net weakening of the WC A·T pairing in the above complexes directly relates with noticeable changes in the regions of anchoring and nonconventional H-bonding, compared to the corresponding nucleobase-gold complexes (Table 8.1). For example, in the complex [A·Au₃(N₇)]·T, participation of the $\text{N}_6\text{--H}_6'$ group in the nonconventional hydrogen bonding with Au₃, which is albeit weaker than in A·Au₃(N₇) (e.g., the H-bond $\text{H}_6'(\text{A}) \cdots \text{Au}_{11}$ elongates by $0.176 \text{ \AA}$), lowers the $\text{DPE}(\text{N}_6\text{--H}_6; \text{A})$ and thus enhances $\text{N}_6\text{--H}_6(\text{A}) \cdots \text{O}_4(\text{T})$, in agreement with the reasoning of the previous subsection. As a result, the $\text{N}_6\text{--H}_6$ bond is lengthened by $0.003 \text{ \AA}$ and the H-bond $\text{H}_6 \cdots \text{O}_4$ shrinks by $0.012 \text{ \AA}$. The central intermolecular H-bond $\text{N}_3\text{--H}_3(\text{T}) \cdots \text{N}_1(\text{A})$ of [A·Au₃(N₇)]·T is, however, weakened: its $\text{N}_3\text{--H}_3$ bond undergoes a contraction by $0.005 \text{ \AA}$ while the $\text{H}_3 \cdots \text{N}_1$ one elongates by $0.029 \text{ \AA}$ since $q^{\text{M}}(\text{N}_1)$ reduces by $0.042 |e|$.

A larger weakening of the WC A·T pairing takes place in A·[T·Au₃(O₂; N₁)] where Au₃ anchors at the O₂ atom of T on the N₁ side (which is, however, blocked by the sugar-phosphate backbone in the DNA). Therein, the anchoring $\text{Au}_{10}\text{--O}_2$ bond is slightly stronger (contracted by $0.011 \text{ \AA}$) than in T·Au₃(O₂; N₁), but the nonconven-

tional $N_1-H_1(T) \cdots Au_{11}$ H-bond whose separation $r(H_1 \cdots Au_{11})$ widens by 0.034 Å shows an opposite trend. The intermolecular H-bonds, $N_3-H_3(T) \cdots N_1(A)$ and $C_2-H_2(A) \cdots O_2(T)$, of $A \cdot [T \cdot Au_3(O_2; N_1)]$ become stronger than for the A·T pair, partly as a result of the increase of the DPE($N_3; T$) since $q^M(N_3)$ drops by 0.024 |e|. The other H-bond $N_6-H_6(A) \cdots O_4(T)$, which is placed on the major groove side, weakens, as is accounted for by the lower PA of the O_4 atom of T, whose Mulliken electron charge decreases by 0.053 |e|.

The interbase region of the WC A·T pair undergoes significant damage by the Au_3 -anchoring either at the N_6 atom of the amino group of A or at the O_4 atom of T. The former anchoring leads to the weakening of the proton donor group $N_6-H_6(A)$ [$\Delta R(N_6-H_6) = 0.019$ Å] and a significant strengthening of the H-bond $N_6-H_6(A) \cdots O_4(T)$, as is manifested by a downshift of the $\nu(N_6-H_6)$ stretch by 247 cm^{-1} (Figure 8.17). The intermolecular H-bond $N_3-H_3(T) \cdots N_1(A)$ of $[A \cdot Au_3(N_6)] \cdot T$ becomes weaker. In addition, interestingly, there occurs a cleavage of $C_2-H_2(A) \cdots O_2(T)$ where the distance between $H_2(A)$ and $O_2(T)$ reaches 3.343 Å, thereby pre-opening the $[A \cdot Au_3(N_6)] \cdot T$ pair on the minor groove side. A substantial weakening of the complex $A \cdot [T \cdot Au_3(O_4)]$ by about 9 $kcal\ mol^{-1}$ relative to $T \cdot Au_3(O_4)$ is partly explained by the breaking of the nonconventional $O_4-H_4 \cdots Au_8$ H-bond (in this regard see condition 4 of Section 8.3.2.1).

8.8.3

[G·C]· Au_3 Complexes

The WC pairing between the guanine and cytosine bases prevents them effectively binding a three-gold cluster at the most favorable N_3 -cytosine site and less favorable O_6 -guanine site on the N_1 side. The rest of the sites of the G and C bases are available in the WC G·C duplex to anchor a gold cluster; Figures 8.20 and 8.21 show the resulting complexes. The most stable are $[G \cdot Au_3(N_3; N_9)] \cdot C$ and $[G \cdot Au_3(N_7)] \cdot C$, characterized by binding energies of 19.3 and 18.0 $kcal\ mol^{-1}$, respectively.¹⁴⁾ Interestingly, the complexes $[A \cdot Au_3(N_3)] \cdot T$ and $[G \cdot Au_3(N_3; N_9)] \cdot C$ are quasi isoenergetic since $E_b([A \cdot Au_3(N_3)] \cdot T) \approx E_b([G \cdot Au_3(N_3; N_9)] \cdot C)$. This implies that the favorable Au_3 -anchoring eliminates the well-known stronger bonding character of the WC G·C pair compared to the A·T one [120].

Let us consider the complex $[G \cdot Au_3(N_3; N_9)] \cdot C$ in detail. Its anchor and nonconventional H-bondings are somewhat stronger than the unpaired to C, viz., the $G \cdot Au_3(N_3; N_9)$ complex (e.g., the anchoring bond and the H-bond distance are shorter by 0.008 and 0.009 Å, respectively; see Table 8.3), but its binding energy is 1.6 $kcal\ mol^{-1}$ smaller. By analogy with the Au_3 -anchored A·T pairs, this small decrease in the binding energy is partly a direct result of the weakening of the intermolecular $N_4-H_4(C) \cdots O_6(G)$ H-bond due to lowering of the PA($O_6; G$) under the Au_3 -anchoring (as follows from Table 8.11 the Mulliken electron charge reduces

14) Notice that the N_9-H_9 group of G is blocked in the DNA molecule [104]. The complex $[G \cdot Au_3(N_3; N_2)] \cdot C$ does not exist – under optimization it converts into $[G \cdot Au_3(N_3; N_9)] \cdot C$.

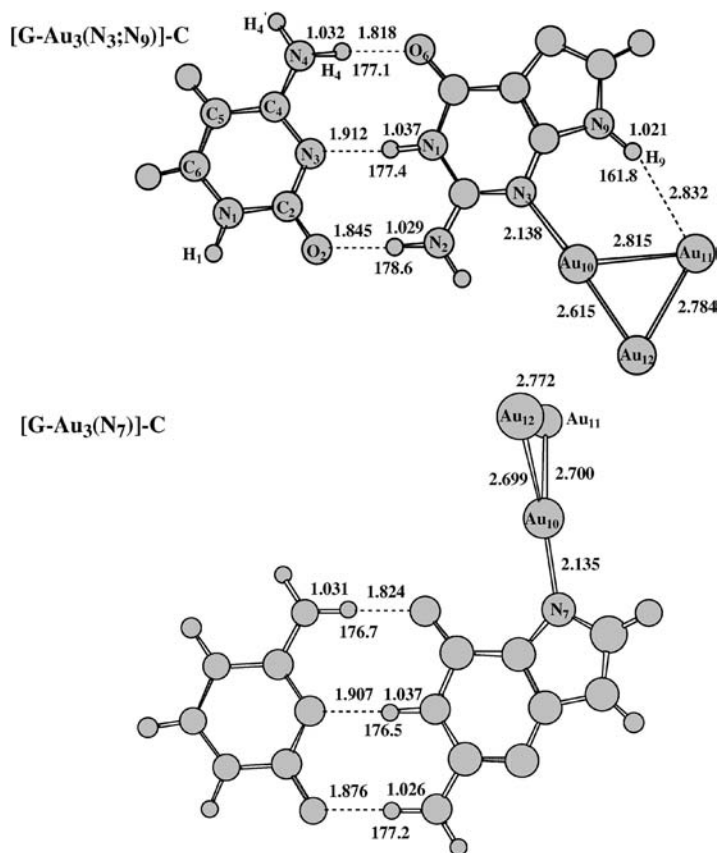


Figure 8.20 Stable $[G\cdot Au_3]\cdot C$ pairs. The WC intermolecular H-bonds of the G-C pair are characterized by the following geometrical parameters: $R[N_4-H_4(C)] = 1.036 \text{ \AA}$, $r[H_4(C) \cdots O_6(G)] = 1.789 \text{ \AA}$, $\angle N_4H_4(C)O_6(G) = 178.9^\circ$; $R[N_1-H_1(G)] = 1.033 \text{ \AA}$, $r[H_1(G) \cdots N_3(C)] = 1.936 \text{ \AA}$, $\angle N_1H_1(G)N_3(C) = 177.3^\circ$; R

$[N_2-H_2(G)] = 1.024 \text{ \AA}$, $r[H_2(G) \cdots O_2(C)] = 1.920 \text{ \AA}$, $\angle N_2H_2(G)O_2(C) = 178.2^\circ$. Bond lengths are given in $\text{\AA}$ and bond angles in degrees. (Reproduced from Figure 6 of Reference [98c] with permission from the American Chemical Society.)

by $0.076 |e|$). The $\nu(N_4-H_4)$ stretch is blue-shifted by 81 cm^{-1} (Table 8.12). The other two H-bonds of $[G\cdot Au_3(N_3; N_9)]\cdot C$ are, however, strengthened. Specifically, the $N_1-H_1(G) \cdots N_3(C)$ one has a shorter (by $0.024 \text{ \AA}$) H-bond separation that results from a decrease of the DPE of the N_1 atom of the $G\cdot Au_3(N_3; N_9)$ complex (the corresponding Mulliken electron charges drops accordingly by $0.051 |e|$). The strengthening of $N_2-H_2(G) \cdots O_2(C)$ is indicated by the shortening of its H-bond by $0.075 \text{ \AA}$ and $\Delta\nu(N_2-H_2) = -92 \text{ cm}^{-1}$ (Figure 8.17).

A net weakening of the WC pairing in the G-C duplex under its interaction with a gold cluster is also predicted when Au_3 anchors either at the N_2 , N_7 or O_6 of G or at the O_2 of C (Tables 8.3 and 8.7 Figure 8.17). By analogy with the WC A-T pair and

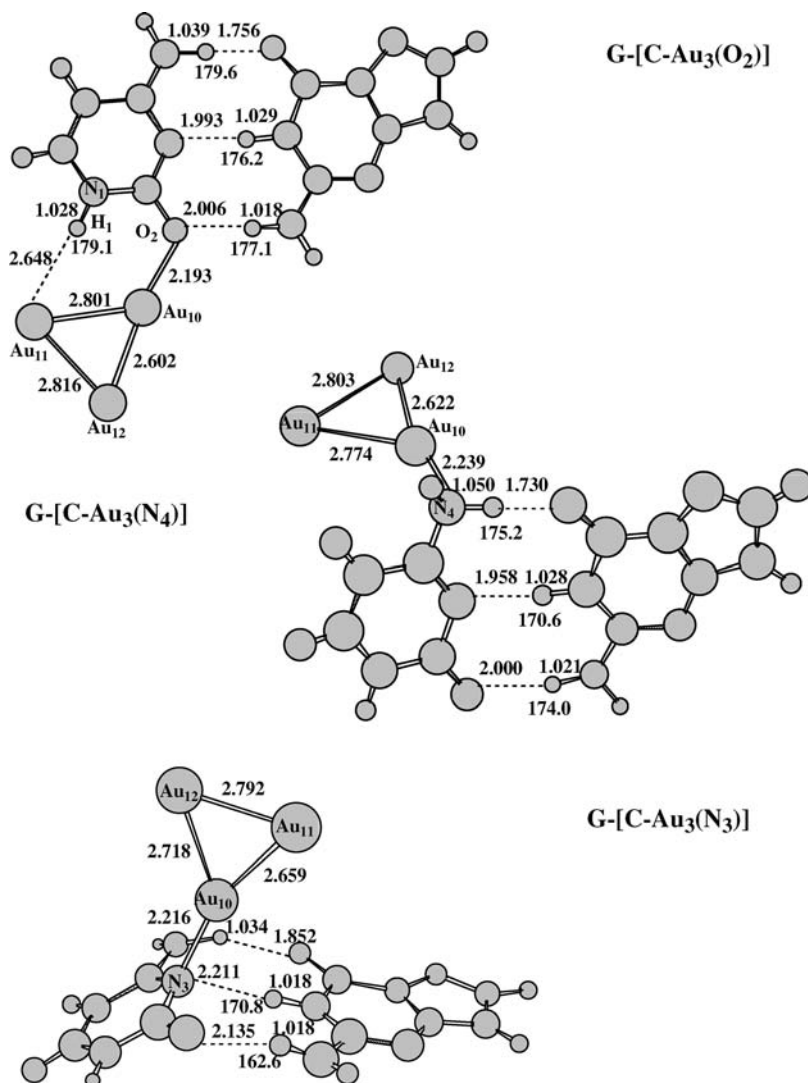


Figure 8.21 Stable G·[C-Au₃] pairs. Bond lengths are given in Å and bond angles in degrees. (Reproduced from Figure 7 of Reference [98c] with permission from the American Chemical Society.)

the [G·Au₃(N₃)]·C one, the origin of this trend probably arises from that fact that, in general, the bonding of Au₃ to the DNA base lowers the base PAs (Table 8.11). The WC pairing in G·C markedly weakens under anchoring of a gold cluster at N₃ or N₄ of cytosine, resulting in the very low binding energies of about 2–3 kcal mol^{−1}.

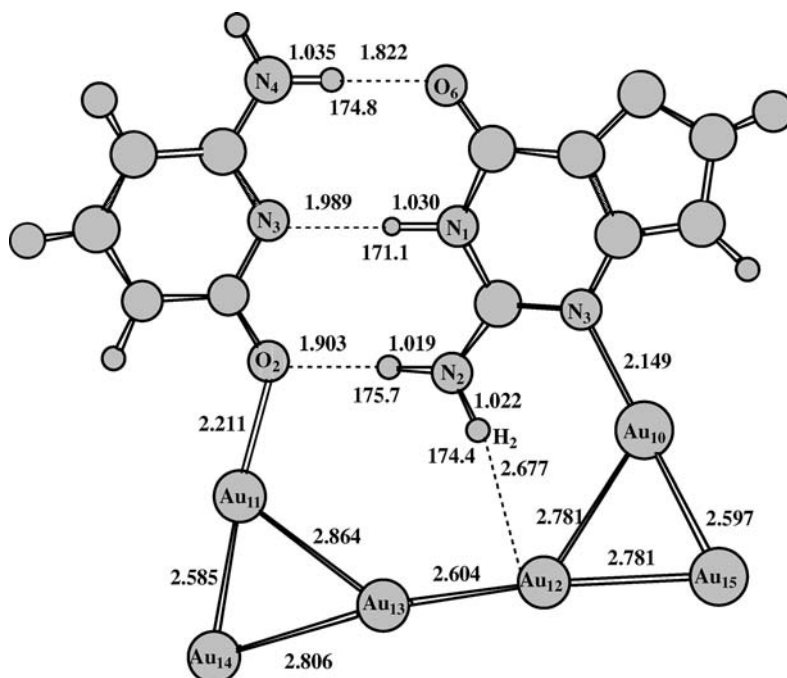


Figure 8.22 Complex $[G \cdot C] \cdot Au_6$. Bond lengths are given in Å and bond angles in degrees. (Reproduced from Figure 8 of Reference [98c] with permission from the American Chemical Society.)

8.8.4

Au₆ Cluster Bridges the WC G·C Pair

In all complexes between the WC pairs A·T and G·C and a three-gold cluster that were examined in Sections 8.8.2 and 8.8.3, the latter – Au₃ – is too small to be accommodated within the interbase region and to link both WC-paired bases together via an additional gold–gold bond (i.e., multiple anchorings to the base pairs), as likely occurs in experiments on adsorption of the DNA bases on Au nanoparticles and surfaces.¹⁵⁾ To illustrate the formation of such an interbase gold–gold bond and to investigate its effect (if it exists) on the WC pairing patterns, we consider the WC hybridization of G·Au₃(N₃; N₂) with C·Au₃(O₂; N₁). The resultant complex is displayed in Figure 8.22. Its rather large binding energy of 62.4 kcal mol^{−1}, taken relative to the isolated species, can obviously mostly be attributed to the formation of the strong interbase gold–gold bond, whose length amounts only to 2.604 Å, and to the formation of the

¹⁵⁾ The different bonding scenarios when each of two Au₃ clusters binds to each monomer of the WC pairs, that is, when the WC pair is trapped between two gold clusters that mimic Au electrodes, has been treated computationally [95g].

Au₆ cluster.¹⁶⁾ On the one hand, this bond reinforces the nonconventional N₂–H₂'(G)···Au₁₂ hydrogen bond and, on the other, it breaks the other, N₁–H₁(C)···Au₁₄. It additionally changes the WC pairing patterns. The two remote bonds, N₄–H₄(C)···O₆(G) and N₁–H₁(G)···N₃(C), are weakened, mostly due to lengthening of their H-bond distances: $r(\text{H}_4 \cdots \text{O}_6)$ by 0.033 Å and $r(\text{H}_1 \cdots \text{N}_3)$ by 0.063 Å, compared to those in the WC G–C pair. The related stretches, $\nu(\text{N}_4\text{--H}_4; \text{C})$ and $\nu(\text{N}_1\text{--H}_1; \text{G})$, are blue-shifted by ≈ 40 and 52 cm^{-1} , respectively. The effect of the interbase gold–gold bond on the nearby H-bond N₂–H₂(G)···O₂(C) is more complex: both the N₂–H₂(G) and H₂···O₂ bonds are compressed, by 0.005 and 0.017 Å, respectively. Overall, the net effect of this interbase gold–gold bond consists in a weakening of the WC G–C pairing.

8.9

Summary and Perspectives

The computational picture of the interaction of DNA bases and Watson–Crick base pairs with small neutral gold clusters Au_{2≤n≤6} has been thoroughly described, via analyzing various features – in particular, the geometrical, spectroscopic and energetic. The key conclusion we have drawn from this picture is that it is true – the interaction of the DNA with gold is rather specific, as the experiments claimed, primarily due to the existence of the two major bonding mechanisms of interactions and their interplay under the charge alternation. These are: the *anchoring*, either of the Au–N or Au–O type, and the nonconventional N–H···Au *hydrogen bonding*.

Anchoring bonding is the leading interaction in the neutral and cationic charge states and results in stronger binding and coplanar coordination when the ring nitrogen atoms of the nucleobases are involved. The anchoring bond predetermines the formation of the nonconventional H-bonding via prearranging the charge distribution within the entire interacting system and “galvanizing” an unanchored atom of the gold cluster to act as a nonconventional proton acceptor, through its lone-pair-like 5d_{±2} and 6s orbitals.

The presented picture of interaction demonstrates another, non-specific type of interaction – the nonconventional hydrogen bonding as a new type of bonding that, on the one hand, originates from the recently revealed propensity of gold to act as a nonconventional proton acceptor with conventional proton donors and, on the other hand, sustains and even reinforces the anchoring one. These bonding interactions are, generally speaking, entangled and separable only in few particular cases of the whole bonding scenario and in some charge states.

The presented picture opens perspectives to manipulate the DNA–gold bonding patterns and to propose concrete experiments, particularly the experiments on anion photoelectron spectroscopy of the DNA base–gold and DNA base pair–gold complexes, partly described in our computational “thought” *mise en scènes*, which are

16) Hybridization of the A·T and G·C WC DNA base pairs with Au₄ and Au₈ has been studied computationally [95j].

actually referred to as “Negative ion – to Neutral” experiments within the well-known general NeNePo (A “Negative ion – to Neutral – to Positive ion”) experimental technique (see Reference [121] and references therein).

Acknowledgment

I gratefully thank Francoise Remacle, Kit Bowen, Alfred Karpfen, Javier F. Luque, Pekka Pyykkö, Camille Sandorfy[†], Lucjan Sobczyk, George V. Yukhnevich and Georg Zundel[†] for encouraging discussions, useful suggestions and valuable comments, and Chérif F. Matta for his kind invitation to contribute to the present book.

References

- 1 A rephrasing of “Small is different” from Reference Nat. Nanotechnol. 2006,1,1.
- 2 A rephrasing of “Tiny Is Beautiful” K. Chang, *New York Times* 2005, February 22.
- 3 The term “nanotechnology” as the science of manipulating atoms and single molecules was first coined by Norio Taniguchi from Tokyo Science University, in 1974: Taniguchi, N. ‘On the Basic Concept of “Nano-Technology”’ Proc. Intl. Conf. Prod. London, Part II, British Society of Precision Engineering, 1974.
- 4 Royal Society & Royal Academy of Engineering (2004) *Nanoscience and Nanotechnologies: Opportunities and Uncertainties*, The Royal Society, London, www.nanotec.org.uk/finalReport.htm.
- 5 Kearnes, M., Macnaghten, P., and Wilsdon, J. (2006) *Governing at the Nanoscale*, Demos, London.
- 6 Yang, P. (ed.) (2003) *The Chemistry of Nanostructured Materials*, World Scientific, Singapore.
- 7 Cao, G. (2004) *Nanostructures and Nanomaterials. Synthesis, Properties and Applications*, World Scientific, Singapore.
- 8 Ozin, G.A. and Arsenault, A.C. (2005) *Nanochemistry: A Chemical Approach to Nanomaterials*, RSC Publishing, Cambridge, UK.
- 9 Heiz, U. and Landman, U. (2006) *Nanocatalysis*, Springer, New York.
- 10 (a) Schmidbaur, H. (ed.) (1999) *Gold: Progress in Chemistry, Biochemistry, and Technology*, John Wiley & Sons, Inc., New York, (b) Bond, G.C., Luo, C., and Thompson, D.T. (2006) *Catalysis by Gold*, World Scientific, Singapore.
- 11 This is the energy-consistent 19-5s²5p⁶5d¹⁰6s¹-valence-electron relativistic effective core potential (RECP) of Ermler, Christiansen and co-workers Ross, R. B.; Powers, J. M.; Atashroo, T.; Ermler, W. C.; LaJohn, L. A.; Christiansen, P. A.J. *Chem. Phys.* 1990, **93**, 6654.
- 12 Torchilin, V.P. (ed.) (2007) *Nanoparticulates as Drug Carriers*, World Scientific, Singapore.
- 13 Joachim, C. and Plévert, L. (2008) *Nanosciences: La Révolution Invisible*, Seuil.
- 14 (a) Schmidbaur, H., Cronje, S., Djordjevic, B., and Schuster, O. (2005) *Chem. Phys.*, **311**, 151; (b) Remacle, F. and Kryachko, E.S. (2004) *Adv. Quantum Chem.*, **47**, 421.
- 15 (a) Hammer, B. and Nørskov, J.K. (1995) *Nature*, **376**, 238; (b) Valden, M., Lai, X., and Goodman, D.W. (1998) *Science*, **281**, 1647; (c) Sanchez, A., Abbet, S., Heiz, U., Schneider, W.-D., Häkkinen, H., Barnett, R.N., and Landman, U. (1999) *J. Phys. Chem. A*, **103**, 9573; (d) Schmid, G. and Corain, B. (2003) *Eur. J. Inorg. Chem.*, 3081.
- 16 (a) Pyykkö, P. (2004) *Angew. Chem. Int. Ed.*, **43**, 4412; (b) Pyykkö, P. (2005) *Inorg. Chim. Acta*, **358**, 4113; (c) Pyykkö, P. (2008) *Chem. Soc. Rev.*, **37**, 1967.
- 17 (a) Pyykkö, P. (2002) *Angew. Chem. Int. Ed.*, **41**, 3573; (b) Pyykkö, P. (2000) *Relativistic Theory of Atoms and Molecules*,

- vol. III, Springer, Berlin; (c) Pyykkö, P. (1988) *Chem. Rev.*, **88**, 563; (d) Pyykkö, P. (1997) *Chem. Rev.*, **97**, 597; (e) Schmidbaur, H. (1995) *Chem. Soc. Rev.*, **24**, 391; the DFT estimates $IE_1(\text{Au})$ rather accurately, as $IE_1^{\text{DFT}}(\text{Au}) = 9.323 \text{ eV}$: (f) (fa) Lide, D.R. (ed.) (1992) Ionization potentials of atoms and atomic ions, in *Handbook of Chemistry and Physics*, CRC Press, Boca Raton, FL; (fb) Korgaonkar, A.V., Gopalaraman, C.P., and Rohatgi, V.K. (1981) *Int. J. Mass. Spectrom. Ion Phys.*, **40**, 127; (fc) Barakat, K.A., Cundari, T.R., Rabaâ, H., and Omary, M.A. (2006) *J. Phys. Chem. B*, **110**, 14645; (fd) Jackschath, C., Rabin, I., and Schulze, W. (1992) *Ber. Bunsenges. Phys. Chem.*, **96**, 1200 and references therein; (g) Neogrády, P., Kellö, V., Urban, M., and Sadlej, A. (1997) *J. Int. J. Quantum Chem.*, **63**, 557; the experimental value of $EA^{\text{exp}}(\text{Au}) = 2.30 \pm 0.10 \text{ eV}$ according to: (h) (ha) Ganteför, G., Krauss, S., and Eberhardt, W. (1998) *J. Electron Spectrosc. Relat. Phenom.*, **88**, 35, 2.308664 $\pm$ 0.000044 eV according to: (hb) Jotop, H. and Lineberger, W.C. (1985) *J. Phys. Chem. Ref. Data*, **14**, 731 2.927 and $\pm$ 0.050 eV according to: (hc) Taylor, K.J., Pettiettehall, C.L., Cheshnovsky, O., and Smalley, R.E. (1992) *J. Chem. Phys.*, **96**, 3319; $EA^{\text{theor}}(\text{Au}) = 2.33 \text{ eV}$: (hd) Buckart, S., Ganteför, G., Kim, Y.D., and Jena, P. (2003) *J. Am. Chem. Soc.*, **125**, 14205; $EA^{\text{theor}}(\text{Au}) = 2.166 \text{ eV}$: (he) Joshi, A.M., Delgass, W.N., and Thomson, K.T. (2005) *J. Phys. Chem. B*, **109**, 22392; (hf) with the used basis set, MP2 yields 1.536 eV; (hg) the $EA^{\text{theor}}(\text{Au}) = 1.86 \text{ eV}$ was calculated at the MCPDF computational level in: Bauschlicher, C.W. Jr., Langhoff, S.R., and Partridge, H.J. (1990) *J. Chem. Phys.*, **93**, 8133; (hh) the PW91PW91 DFT level in conjunction with the basis set used in the present work yields 2.25 eV and 2.31 eV with the LANL2DZ basis set, as reported in: Walker, A.V. (2005) *J. Chem. Phys.*, **122**, 094310; (i) Antušek, A., Urban, M., and Sadlej, A.J. (2003) *J. Chem. Phys.*, **119**, 7247; (j) Bilić, A., Reimers, J.R., Hush, N.S., and Hafner, J. (2002) *J. Chem. Phys.*, **116**, 8981; (k) Gollisch, H. (1984) *J. Phys. B*, **17**, 1463; (l) Schwerdtfeger, P., Dolg, M., Schwarz, W.H.E., Bowmaker, G.A., and Boyd, P.W.D. (1989) *J. Chem. Phys.*, **91**, 1762; (m) Marian, C.M. (1990) *Chem. Phys. Lett.*, **173**, 175.
- 18** (a) Häkkinen, H. and Landman, U. (2000) *Phys. Rev. B*, **62**, R2287; (b) Häkkinen, H., Moseler, M., and Landman, U. (2002) *Phys. Rev. Lett.*, **89**, 033401; (c) Häkkinen, H., Yoon, B., Landman, U., Li, X., Zhai, H.J., and Wang, L.C. (2003) *J. Phys. Chem. A*, **107**, 6168; (d) Bonačić-Koutecký, V., Burda, J., Mitric, R., Ge, M.F., Zampella, G., and Fantucci, P. (2002) *J. Chem. Phys.*, **117**, 3120; (e) Furche, F., Ahlrichs, R., Weis, P., Jacob, C., Gilb, S., Bierweiler, T., and Kappes, M.M. (2002) *J. Chem. Phys.*, **117**, 6982; (f) Gilb, S., Weis, P., Furche, F., Ahlrichs, R., and Kappes, M.M. (2002) *J. Chem. Phys.*, **116**, 4094; (g) Lee, H.M., Ge, M., Sahu, B.R., Tarakeshwar, P., and Kim, K.S. (2003) *J. Phys. Chem. B*, **107**, 9994; (h) Wang, J.L., Wang, G.H., and Zhao, J.J. (2002) *Phys. Rev. B*, **66**, 035418; (i) Xiao, L. and Wang, L. (2004) *Chem. Phys. Lett.*, **392**, 452; (j) Olson, R.M., Varganov, S., Gordon, M.S., Metiu, H., Chretien, S., Piecuch, P., Kowalski, K., Kucharski, S., and Musial, M. (2005) *J. Am. Chem. Soc.*, **127**, 1049; (k) Koskinen, P., Häkkinen, H., Huber, B., Issendorff, B.V., and Moseler, M. (2007) *Phys. Rev. Lett.*, **98**, 015701; (l) Han, V.K. (2006) *J. Chem. Phys.*, **124**, 024316; (m) Fernández, E.M., Soler, J.M., Garzón, I.L., and Balbás, L.C. (2004) *Phys. Rev. B*, **70**, 165403; (n) Fernández, E.M., Soler, J.M., and Balbás, L.C. (2006) *Phys. Rev. B*, **73**, 235433; (o) Remacle, F. and Kryachko, E.S. (2004) *Adv. Quantum Chem.*, **47**, 421; (p) Remacle, F. and Kryachko, E.S. (2005) *J. Chem. Phys.*, **122**, 044304; this reference demonstrates that the size threshold for the 2D–3D coexistence already develops for the cationic clusters Au_5^+ and Au_7^+ , and for the neutral at Au_9 ; the latter conclusion was also drawn by Häkkinen *et al.* [14c]; for Au_N^- , 3D appears at $N \geq 11$ [14k]. (r) Johansson, M.P., Lechten, A., Schooss, D., Kappes, M.M., and Furche, F. (2008) *Phys. Rev. A*, **77**, 053202; (s) Häkkinen, H. (2008) *Chem. Soc. Rev.*, **37**, 1847.
- 19** The T-Au₃ complexes where the gold cluster Au₃ “hooks” T at its N₁ and N₃ atoms are unstable within the present computational approach. This

- computational result agrees with the experimental one reported by Lindsay and co-workers Tao, N. J.; de Rose, J. A.; Lindsay, S.M. *J. Phys. Chem.* 1993, **97**, 910; see also Gourishankar, A.; Shukla, S.; Ganesh, K. N.; Sastry, M. *J. Am. Chem. Soc.* 2004, **126**, 13186f.
- 20 (a) Feynman, R. (1960) *Eng. Sci. (Caltech)*, **23**, 22; (b) Crommie, M.F., Lutz, C.P., and Eigler, D.M. (1993) *Science*, **262**, 218; (c) Avouris, P. and Lyo, I.-W. (1994) *Science*, **264**, 942; (d) Khanna, S.N. and Castleman, A.W. (eds) (2003) *Quantum Phenomena in Clusters and Nanostructures*, Springer-Verlag, Heidelberg.
 - 21 de Heer, W.A. (1993) *Rev. Mod. Phys.*, **65**, 611.
 - 22 (a) Haruta, M., Kobayashi, T., Sano, H., and Yamada, N. (1987) *Chem. Lett.*, 405; (b) Haruta, M., Yamada, N., Kobayashi, T., and Iijima, S. (1989) *J. Catal.*, **115**, 301; (c) Haruta, M., Tsubota, S., Kobayashi, T., Kageyama, H., Genet, M.J., and Delmon, B. (1993) *J. Catal.*, **144**, 175; (d) Haruta, M. (1997) *Catal. J. Today*, **36**, 153; (e) Iizuka, Y., Tode, T., Takao, T., Yatsu, K.I., Takeuchi, T., Tsubota, S., and Haruta, M. (1999) *Catal. J. Today*, **187**, 50; (f) Shiga, A. and Haruta, M. (2005) *Appl. Catal. A: General*, **291**, 6; (g) Date, M., Okumura, M., Tsubota, S., and Haruta, M. (2004) *Angew. Chem. Int. Ed.*, **43**, 2129.
 - 23 (a) Chen, M.-S. and Goodman, D.W. (2004) *Science*, **306**, 252; (b) See also Jacoby, M. (2004) *C&EN*, **30**, 9; (c) Chen, M., Cai, Y., Yan, Z., and Goodman, D.W. (2006) *J. Am. Chem. Soc.*, **128**, 6341.
 - 24 (a) Alivisatos, A.P. (1996) *Science*, **271**, 933; (b) Coulthard, I., Degen, I.S., Zhu, Y., and Sham, T.K. (1998) *Can. J. Phys.*, **76**, 1707.
 - 25 (a) Bäumer, M. and Freund, H.-J. (1999) *Progr. Surf. Sci.*, **61**, 127; (b) Clair, T.P. St. and Goodman, D.W. (2000) *Top. Catal.*, **13**, 5.
 - 26 Gardea-Torresday, J.L., Parson, J.G., Gomez, E., Peralta-Videa, J., Troiani, H.E., Santiago, P., and Yacaman, M.J. (2002) *Nano Lett.*, **2**, 397.
 - 27 Choudhary, T.V. and Goodman, D.W. (2002) *Top. Catal.*, **20**, 35.
 - 28 Griesel, R.J.H., Kooyman, P.J., and Nieuwenhuys, B.E. (2000) *J. Catal.*, **191**, 430.
 - 29 Salama, T.M., Ohnishi, R., and Ichikawa, M. (1996) *J. Chem. Soc., Faraday Trans.*, **92**, 301.
 - 30 Hayashi, T., Tanaka, K., and Haruta, M. (1998) *J. Catal.*, **178**, 566.
 - 31 Chrétien, S., Gordon, M.S., and Metiu, H. (2004) *J. Chem. Phys.*, **121**, 3756.
 - 32 Heiz, U., Sanchez, A., Abbet, S., and Schneider, W.-D. (2000) *Chem. Phys.*, **262**, 189.
 - 33 Heiz, U. and Schneider, W.-D. (2000) *J. Phys. D: Appl. Phys.*, **33**, R85.
 - 34 Valden, V., Pak, S., Lai, X., and Goodman, D.W. (1998) *Catal. Lett.*, **56**, 7.
 - 35 Lai, X. and Goodman, D.W. (2000) *J. Mol. Catal. A*, **162**, 33.
 - 36 Kim, Y.D. (2004) *Int. J. Mass Spectrom.*, **238**, 17.
 - 37 Lemire, C., Meyer, R., Shaikhutdinov, S., and Freund, H.-J. (2004) *Angew. Chem. Int. Ed.*, **43**, 118.
 - 38 (a) Lopez, N., Janssens, T.V.W., Clausen, B.S., Xu, Y., Mavrikakis, M., Bligaard, T., and Nørskov, J.K. (2004) *J. Catal.*, **223**, 232; (b) Lopez, N., Nørskov, J.K., Janssens, T.V.W., Carlsson, A., Puig-Molina, A., Clausen, B.S., and Grunwaldt, J.-D. (2004) *J. Catal.*, **225**, 86.
 - 39 Bocuzzi, F., Chiorino, A., and Manzoli, M. (2001) *Mater. Sci. Eng. C*, **15**, 215.
 - 40 Mavrikakis, M., Stoltze, P., and Nørskov, J.K. (2000) *Catal. Lett.*, **64**, 10.
 - 41 Haruta, M. (2002) *CATTECH*, **6**, 102.
 - 42 Pietron, J.J., Stroud, R.M., and Rolison, D.R. (2002) *Nano Lett.*, **2**, 545.
 - 43 Lopez, N. and Nørskov, J.K. (2002) *J. Am. Chem. Soc.*, **124**, 11262.
 - 44 Molina, L.M. and Hammer, B. (2003) *Phys. Rev. Lett.*, **90**, 206102.
 - 45 Molina, L.M. and Hammer, B. (2004) *Phys. Rev. B*, **69**, 155424.
 - 46 Molina, L.M., Rasmussen, M.D., and Hammer, B. (2004) *J. Chem. Phys.*, **120**, 7673.
 - 47 Häkkinen, H., Abbet, S., Sanchez, A., Heiz, U., and Landman, U. (2003) *Angew. Chem. Int. Ed.*, **42**, 1297.
 - 48 Cho, A. (2003) *Science*, **299**, 1684.
 - 49 Guzman, J. and Gates, B.C. (2004) *J. Am. Chem. Soc.*, **126**, 2672.
 - 50 Cox, D.M., Brickman, R., Creegan, K., and Kaldor, A. (1991) *Z. Phys. D*, **19**, 353.
 - 51 Wallace, W.T. and Wetten, R.L. (2002) *J. Am. Chem. Soc.*, **124**, 7499.

- 52 Mills, G., Gordon, M.S., and Metiu, H. (2002) *Chem. Phys. Lett.*, **359**, 493.
- 53 Yoon, B., Häkkinen, H., and Landman, U. (2003) *J. Phys. Chem. A*, **107**, 4066.
- 54 Socaciu, L.D., Hagen, J., Bernhardt, T.M., Wöste, L., Heiz, U., Häkkinen, H., and Landman, U. (2003) *J. Am. Chem. Soc.*, **125**, 10437.
- 55 Stolicic, D., Fischer, M., Ganteför, G., Kim, Y.D., Sun, Q., and Jena, P. (2003) *J. Am. Chem. Soc.*, **125**, 2848.
- 56 Kim, Y.D., Fischer, M., and Ganteför, G. (2003) *Chem. Phys. Lett.*, **377**, 170.
- 57 Yoon, B., Häkkinen, H., Landman, U., Wörz, A.S., Antonietti, J.-M., Abbet, S., Judai, K., and Heiz, U. (2005) *Science*, **307**, 403.
- 58 Schwerdtfeger, P. (2003) *Angew. Chem. Int. Ed.*, **42**, 1892.
- 59 Schwarz, H. (2003) *Angew. Chem. Int. Ed.*, **42**, 4442.
- 60 (a) Kroto, H.W., Heath, J.R., O'Brien, S.C., Curl, R.F., and Smalley, R.E. (1985) *Nature*, **318**, 162; (b) Heath, J.R., Zhang, Q., O'Brien, S.C., Curl, R.F., Kroto, H.W., and Smalley, R.E. (1987) *J. Am. Chem. Soc.*, **109**, 359; (c) Kroto, H.W., Heath, J.R., O'Brien, S.C., Curl, R.F., and Smalley, R.E. (1987) *Astrophys. J.*, **314**, 352.
- 61 (a) Pyykkö, P. and Runeberg, N. (2002) *Angew. Chem. Int. Ed.*, **41**, 2174; (b) Johansson, M.P., Sundholm, D., and Vaara, J. (2004) *Angew. Chem. Int. Ed.*, **43**, 2678; (c) Li, X., Kiran, B., Li, J., Zhai, H.J., and Wang, L.S. (2002) *Angew. Chem. Int. Ed.*, **41**, 4786; (d) Zhai, H.J., Li, J., and Wang, L.S. (2004) *J. Chem. Phys.*, **121**, 8369; (e) Autschbach, J., Hess, B.A., Johansson, M.P., Neugebauer, J., Patzschke, M., Pyykkö, P., Reiher, P., and Sundholm, D. (2004) *Phys. Chem. Chem. Phys.*, **6**, 11; (f) Sun, Q., Wang, Q., Jena, P., and Kawazoe, Y. (2008) *ACS Nano*, **2**, 341; (g) Stener, M., Nardelli, A., and Fronzoni, G. (2008) *J. Chem. Phys.*, **128**, 134307; (h) Yoon, B., Koskinen, P., Huber, B., Kostko, O., Issendorff, B.v., Häkkinen, H., Moseler, M., and Landman, U. (2007) *Chem. Phys. Chem.*, **8**, 157; (i) Qiu, Y.-X., Wang, S.-G., and Schwarz, W.H.E. (2004) *Chem. Phys. Lett.*, **397**, 374; (j) Gao, Y., Bulusu, S., and Zeng, X.C. (2005) *J. Am. Chem. Soc.*, **127**, 156801; (k) Bulusu, S., Li, X., Wang, L.-S., and Zeng, X.C. (2006) *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 8326, 40; (l) Wang, D.-L., Sun, X.-P., Shen, H.-T., Hou, D.-Y., and Zhai, Y.-C. (2008) *Chem. Phys. Lett.*, **457**, 366; (m) Wang, J., Jellinek, J., Zhao, J., Chen, Z., King, R.B., and Schleyer, P.v.R. (2005) *J. Phys. Chem. A*, **109**, 9265; (n) Tian, D.X., Zhao, J.J., Wang, B.L., and King, R.B. (2007) *J. Phys. Chem. A*, **111**, 411; (o) Gao, Y. and Zeng, X.C. (2005) *J. Am. Chem. Soc.*, **127**, 3698; (p) Häkkinen, H. and Moseler, M. (2006) *Comp. Mat. Sci.*, **35**, 332; (q) Karttunen, A.J., Linnolahti, M., Pakkanen, T.A., and Pyykkö, P. (2008) *Chem. Commun.*, 465; (r) Kryachko, E.S. and Remacle, F. (2007) *Int. J. Quantum Chem.*, **107**, 2922; (s) Kryachko, E.S. and Remacle, F. (2009) *J. Phys. Chem. C*, **113**, 0000.
- 62 (a) Li, J., Li, X., Zhai, H.-J., and Wang, L.-S. (2003) *Science*, **299**, 864; (b) King, R.B., Chen, Z., and Schleyer, P.v.R. (2004) *Inorg. Chem.*, **43**, 4564.
- 63 (a) Everts, M., Saini, V., Leddon, J.L., Kok, R.J., Stoff-Khalili, M., Preuss, M.A., Millican, C.L., Perkins, G., Brown, J.M., Bagaria, H., Nikles, D.E., Johnson, D.T., Zharov, V.P., and Curiel, D.T. (2006) *Nano Lett.*, **6**, 587; (b) Willner, B., Katz, E., and Willner, I. (2006) *Curr. Opin. Biotechnol.*, **17**, 589; (c) Lévy, R. (2006) *ChemBioChem*, **7**, 1141; (d) Templeton, A.C., Welfing, W.P., and Murray, R.W. (2000) *Acc. Chem. Res.*, **33**, 27; (e) Kamat, P.V. (2002) *J. Phys. Chem. B*, **106**, 7729; (f) Thomas, K.G. and Kamat, P.V. (2003) *Acc. Chem. Res.*, **36**, 888; (g) Shenhar, R. and Rotello, V.M. (2003) *Acc. Chem. Res.*, **36**, 549; (h) Drechsler, U., Erdogan, B., and Rotello, V.M. (2004) *Chem. Eur. J.*, **10**, 5570; (i) Eustis, S. and El-Sayed, M.A. (2006) *Chem. Soc. Rev.*, **35**, 209; (j) Lee, D., Donkers, R.L., Wang, G., Harper, A.S., and Murray, R.W. (2004) *J. Am. Chem. Soc.*, **126**, 6193; (k) Guo, R. and Murray, R.W. (2005) *J. Am. Chem. Soc.*, **127**, 12140; (l) Wang, G., Huang, T., Murray, R.W., Menard, L., and Nuzzo, R.G. (2005) *J. Am. Chem. Soc.*, **127**, 812; (m) Dulkeith, E., Niedereichholz, T., Klar, T.A., and Feldmann, J. (2004) *Phys. Rev. B*, **70**, 205424; (n) Wang, G., Guo, R., Kalyuzhny, G., Choi, J.-P., and Murray, R.W. (2006) *J. Phys. Chem. B*, **110**, 20282; (o) Cheng, P.P.H., Silvester, D., Wang, G.,

- Kalyuzhny, G., Douglas, A., and Murray, R.W. (2006) *J. Phys. Chem. B*, **110**, 4637; (p) Montalti, M., Zaccheroni, N., Prodi, L., O'Reilly, N., and James, S.L. (2007) *J. Am. Chem. Soc.*, **129**, 2418; (q) Battistini, G., Cozzi, P.G., Jalkanen, J.-P., Montalti, M., Prodi, L., Zaccheroni, N., and Zerbetto, F. (2008) *ACS Nano*, **2**, 77.
- 64 Daniel, M.C. and Astruc, D. (2004) *Chem. Rev.*, **104**, 293–346.
- 65 (a) Seeman, N.C. (2003) *Nature*, **421**, 427; (b) Smith, L.M. (2006) *Nature*, **440**, 283; (c) Rothmund, P.W.K. (2006) *Nature*, **440**, 297; see also (d) Dusastre, V. (2008) *Nature*, **451**, 770.
- 66 (a) Niemeyer, C.M. and Mirkin, C.A. (2004) *NanoBiotechnology: Concepts, Methods and Applications*, Wiley-VCH Verlag GmbH, Weinheim, (b) Nalwa, H.S. (ed.) (2005) *Handbook of Nanostructured Biomaterials and their Applications in Nanobiotechnology*, American Scientific Publishers, Stevenson Ranch, CA
- 67 (a) Mirkin, C.A., Letsinger, R.L., Mucic, R.C., and Storhoff, J.J. (1996) *Nature*, **382**, 607; (b) Storhoff, J.J., Elghanian, R., Mucic, R.C., Mirkin, C.A., and Letsinger, R. (1998) *J. Am. Chem. Soc.*, **120**, 1959; (c) Demers, L.M., Mirkin, C.A., Mucic, R.C., Reynolds, R.A., Letsinger, R.L., Elghanian, R., and Viswanadham, G. (2000) *Anal. Chem.*, **72**, 5535; (d) Storhoff, J.J., Lazarides, A.A., Mucic, R.C., Mirkin, C.A., Letsinger, R.L., and Schatz, G.C. (2000) *J. Am. Chem. Soc.*, **122**, 4640; (e) Storhoff, J.J., Mucic, R.C., and Mirkin, C.A. (1997) *J. Clust. Sci.*, **8**, 179; (f) Elghanian, R., Storhoff, J.J., Mucic, R.C., Letsinger, R.L., and Mirkin, C.A. (1997) *Science*, **277**, 1078; (g) Mucic, R.C., Storhoff, J.J., Mirkin, C.A., and Letsinger, R.L. (1998) *J. Am. Chem. Soc.*, **120**, 12674; (h) Mitchell, G.P., Mirkin, C.A., and Letsinger, R.L. (1999) *J. Am. Chem. Soc.*, **121**, 8122.
- 68 (a) Reynolds, R.A., Mirkin, C.A., and Letsinger, R.L. (2000) *J. Am. Chem. Soc.*, **122**, 3795; (b) Taton, T.A., Mucic, R.C., Mirkin, C.A., and Letsinger, R.L. (2000) *J. Am. Chem. Soc.*, **122**, 6305; (c) Storhoff, J.J. and Mirkin, C.A. (1999) *Chem. Rev.*, **99**, 1849; (d) Lazarides, A.A. and Schatz, G.C. (2000) *J. Phys. Chem. B*, **104**, 460; (e) Lazarides, A.A. and Schatz, G.C. (2000) *J. Chem. Phys.*, **112**, 2987; (f) Park, S.-J., Lazarides, A.A., Mirkin, C.A., and Letsinger, R.L. (2001) *Angew. Chem. Int. Ed.*, **40**, 2909; (g) Reynolds, R.A. III, Mirkin, C.A., and Letsinger, R.L. (2000) *Pure Appl. Chem.*, **72**, 229; (h) Li, Z., Jin, R., Mirkin, C.A., and Letsinger, R.L. (2002) *Nucleic Acid. Res.*, **30**, 1558.
- 69 (a) Cao, Y.W.C., Jin, R., and Mirkin, C.A. (2002) *Science*, **297**, 1536; (b) Park, S.-J., Taton, T.A., and Mirkin, C.A. (2002) *Science*, **295**, 1503; (c) Jun, R.C., Wu, G.S., Li, Z., Mirkin, C.A., and Schatz, G.C. (2003) *J. Am. Chem. Soc.*, **125**, 1643; (d) Nam, J.-M., Thaxton, C.S., and Mirkin, C.A. (2003) *Science*, **301**, 1884; (e) Niemeyer, C.M., Ceyhan, B., Gao, S., Chi, L., Peschel, S., and Simon, U. (2001) *Colloid Polym. Sci.*, **279**, 68; (f) Peschel, S., Ceyhan, B., Niemeyer, C.M., Gao, S., Chi, L., and Simon, U. (2002) *Mater. Sci. Eng. C*, **19**, 47; (g) Niemeyer, C.M. (2001) *Angew. Chem. Int. Ed.*, **40**, 4129; (h) Niemeyer, C.M., Burger, W., and Peplies, J. (1998) *Angew. Chem. Int. Ed.*, **37**, 2265; (i) Yang, J., Yang, L., Too, H.-P., Chow, G.-M., and Gan, L.M. (2006) *Chem. Phys.*, **323**, 304.
- 70 (a) Parak, W.J., Pellegrino, T., Micheel, C.M., Gerion, D., Williams, S.C., and Alivisatos, A.P. (2003) *Nano Lett.*, **3**, 33; (b) Alivisatos, A.P., Johnsson, K.P., Peng, X., Wislon, T.E., Loweth, C.J., Bruchez, M.P. Jr., and Schultz, G.C. (1996) *Nature*, **382**, 609; (c) Pirrung, M.C. (2002) *Angew. Chem. Int. Ed.*, **41**, 1277; (d) Basir, R. (2001) *Superlattices Microstruct.*, **29**, 1; (e) Hölzel, R., Gajovic-Eichelmann, N., and Bier, F.F. (2003) *Biosens. Bioelectron.*, **18**, 555; (f) Xiao, S., Liu, F., Rosen, A.E., Hainfeld, J.F., Seeman, N.C., Musier-Forsyth, K., and Kiehl, R.A. (2002) *J. Nanopart. Res.*, **4**, 313; (g) Harnack, O., Ford, W.E., Yasuda, A., and Wessels, J.M. (2002) *Nano Lett.*, **2**, 919; (h) Daniel, M.-C. and Astruc, D. (2004) *Chem. Rev.*, **104**, 293 and references therein; (i) Seeman, N.C. (2003) *Nature*, **421**, 427; (j) Alivisatos, A.P. (2004) *Nat. Biotechnol.*, **22**, 47; (k) Gourishankar, A., Shukla, S., Ganesh, K.N., and Sastry, M. (2004) *J. Am. Chem. Soc.*, **126**, 13186.

- 71 (a) Sato, K., Hosokawa, K., and Maeda, M. (2003) *J. Am. Chem. Soc.*, **125**, 8102; (b) Maeda, Y., Tabata, H., and Kawai, T. (2001) *Appl. Phys. Lett.*, **79**, 1181; (c) Yonezawa, T., Onoue, S.-Y., and Kimizuka, N. (2002) *Chem. Lett.*, 1172; (d) Gearheart, L.A., Ploehn, H.J., and Murphy, C.J. (2001) *J. Phys. Chem. B*, **105**, 12609; (e) Petty, J.T., Zheng, J., Hud, N.V., and Dickson, R.M. (2004) *J. Am. Chem. Soc.*, **126**, 5207; (f) Liu, D., Park, S.H., Reif, J.H., and LaBean, T.H. (2004) *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 717; (g) Yan, H., Park, S.H., Finkelstein, G., Reif, J.H., and LaBean, T.H. (2003) *Science*, **301**, 1882.
- 72 Richter, J. (2003) *Physica E*, **16**, 157 and references therein.
- 73 (a) Park, S.Y. and Stroud, D. (2003) *Phys. Rev. B*, **67**, 212202; (b) Park, S.Y. and Stroud, D. (2003) *Physica B*, **338**, 353; (c) Park, S.Y. and Stroud, D. (2003) *Phys. Rev. B*, **68**, 224201.
- 74 Slocik, J.M., Moore, J.T., and Wright, D.W. (2002) *Nano Lett.*, **2**, 169.
- 75 Tarlov, M.J. and Steel, A.B. (2003) *Biomolecular Films: Design, Function, and Applications*, vol. 111 (ed. J.F. Rusling), Marcel Dekker, New York, pp. 545–608.
- 76 Liu, Y., Meyer-Zaika, W., Franzka, S., Schmid, G., Tsoli, M., and Kuhn, H. (2003) *Angew. Chem. Int. Ed.*, **42**, 2853.
- 77 Wolf, L.K., Gao, Y., and Georgiadis, R.M. (2004) *Langmuir*, **20**, 3357.
- 78 Niemeyer, C.M. (2001) *Angew. Chem. Int. Ed.*, **40**, 4128.
- 79 (a) Braun, E., Eichen, Y., Sivan, U., and Ben-Yoseph, G. (1998) *Nature*, **391**, 775; (b) Joachim, C., Gimzewski, J.K., and Aviram, A. (2000) *Nature*, **408**, 541.
- 80 Niemeyer, C.M. and Adler, M. (2002) *Angew. Chem. Int. Ed.*, **41**, 3779.
- 81 Yang, X., Vologodskii, A.V., Liu, B., Kemper, B., and Seeman, N.C. (1998) *Biopolymers*, **45**, 69.
- 82 Mao, C., Sun, W., Shen, Z., and Seeman, N.C. (1999) *Nature*, **397**, 144.
- 83 Niemeyer, C.M., Adler, M., Lenhert, S., Gao, S., Fuchs, H., and Chi, L.F. (2001) *ChemBioChem*, **2**, 260.
- 84 Liu, D. and Balasubramanian, S. (2003) *Angew. Chem. Int. Ed.*, **42**, 5734.
- 85 (a) Yurke, B., Turberfield, A.J., Mills, A.P. Jr., Simmel, F.C., and Neumann, J.L. (2000) *Nature*, **406**, 605; (b) Li, J.J. and Tan, W. (2002) *Nano Lett.*, **2**, 315; (c) Yan, H., Zhang, X., Shen, Z., and Seeman, N.C. (2002) *Nature*, **415**, 62; (d) Dittmer, W.U., Reuter, E., and Simmel, F.C. (2004) *Angew. Chem. Int. Ed.*, **43**, 3554; (e) Chen, Y., Wang, M., and Mao, C. (2004) *Angew. Chem. Int. Ed.*, **43**, 3550; (f) Hazarika, P., Ceyhan, B., and Niemeyer, C.M. (2004) *Angew. Chem. Int. Ed.*, **43**, 6469.
- 86 Sato, K., Hosokawa, K., and Maeda, M. (2003) *J. Am. Chem. Soc.*, **125**, 8102.
- 87 (a) Lee, J.-S., Stoeva, S.I., and Mirkin, C.A. (2006) *J. Am. Chem. Soc.*, **128**, 8899; (b) Hurst, S.J., Lytton-Jean, A.K.R., and Mirkin, C.A. (2006) *Anal. Chem.*, **78**, 8313; (c) Dilenback, L.M., Goodrich, G.P., and Keating, C.D. (2006) *Nano Lett.*, **6**, 16; (d) Niemeyer, C.M. and Simon, U. (2005) *Eur. J. Inorg. Chem.*, 3641; (e) Lee, J.-S., Han, M.S., and Mirkin, C.A. (2007) *Angew. Chem. Int. Ed.*, **46**, 4093; (f) Cerruti, M.G., Sauthier, M., Leonard, D., Liu, D., Duscher, G., Feldheim, D.L., and Franzen, S. (2006) *Anal. Chem.*, **78**, 3282; (g) Rosi, N.L. and Mirkin, C.A. (2005) *Chem. Rev.*, **105**, 1547; (h) He, L., Musick, M.D., Nicewarner, S.R., Salinas, F.G., Benkovic, S.J., Natan, M.J., and Keating, C.D. (2000) *J. Am. Chem. Soc.*, **122**, 9071; (i) Liu, J. and Lu, Y. (2003) *J. Am. Chem. Soc.*, **125**, 6642; (j) Pavlov, V., Xiao, Y., Shlyahovsky, B., and Willner, I. (2004) *J. Am. Chem. Soc.*, **126**, 11768; (k) Rosi, N.L., Giljohann, D.A., Thaxton, C.S., Lytton-Jean, A.K.R., Han, M.S., and Mirkin, C.A. (2006) *Science*, **312**, 1027; (l) Seferos, D.S., Giljohann, D.A., Rosi, N.L., and Mirkin, C.A. (2007) *ChemBioChem*, **8**, 1230; (m) Lee, J.-S., Seferos, D.S., Giljohann, D.A., and Mirkin, C.A. (2008) *J. Am. Chem. Soc.*, **130**, 5430.
- 88 Liu, Y., Meyer-Zaika, W., Franzka, S., Schmid, G., Tsoli, M., and Kuhn, H. (2003) *Angew. Chem. Int. Ed.*, **42**, 2853.
- 89 (a) Andersson, D., Hammarstrom, P., and Carlsson, U. (2001) *Biochemistry*, **40**, 2653; (b) Simizu, T. and Takada, A. (1997) *Polym. Networks*, **5**, 267; (c) Pindur, U. and Fischer, G. (1996) *Curr. Med. Chem.*, **3**, 379; (d) Hudson, B.P. and Barton, J.K. (1998) *J. Am. Chem. Soc.*, **120**, 6877; (e) Yang, X.L. and Wang, A.H. (1999)

- Pharmacol. Therap.*, **83**, 181; (f) Hambley, T.W. and Jones, A.R. (2001) *Coord. Chem. Rev.*, **212**, 35; (g) Bashir, R. (2001) *Superlattices Microstruct.*, **29**, 1; (h) Garzón, I.L., Artacho, E., Beltrán, M.R., García, A., Junquera, J., Michaelian, K., Ordejón, P., Rovira, C., Sánchez-Portal, D., and Soler, J.M. (2001) *Nanotechnology*, **12**, 126; (i) Patolsky, F., Weizmann, Y., Lioubashevski, O., and Willner, I. (2002) *Angew. Chem. Int. Ed.*, **41**, 2323; (j) Mirkin, C.A. (2000) *Inorg. Chem.*, **39**, 2258; (k) Braun, E., Eichen, Y., Sivan, U., and Ben-Yoseph, G. (1998) *Nature*, **391**, 775.
- 90** (a) Coffey, J.L., Bigham, S.R., Pinizzotto, R.F., and Yang, H. (1992) *Nanotechnology*, **3**, 69; (b) Coffey, J.L., Bigham, S.R., Li, X., Pinizzotto, R.F., Rho, Y.G., Pirtle, R.M., and Pirtle, I.L. (1996) *Appl. Phys. Lett.*, **69**, 3851; (c) Braun, E., Eichen, Y., Sivan, U., and Ben-Yoseph, G. (1998) *Nature*, **391**, 775; (d) Cassell, A.M., Scrivens, W.A., and Tour, J.M. (1998) *Angew. Chem. Int. Ed.*, **37**, 1528; (e) Cao, Y.W., Jin, R.C., and Mirkin, C.A. (2001) *J. Am. Chem. Soc.*, **123**, 7961; (f) Park, S.J., Lazarides, A.A., Mirkin, C.A., Brazis, P.W., Kannewurf, C.R., and Letsinger, R.L. (2000) *Angew. Chem. Int. Ed.*, **39**, 3845; (g) Taton, T.A., Mirkin, C.A., and Letsinger, R.L. (2000) *Science*, **289**, 1757; (h) Sauthier, M.L., Carroll, R.L., Gorman, C.B., and Franzen, S. (2002) *Langmuir*, **18**, 1825; (i) Maeda, Y., Tabata, H., and Kawai, T. (2001) *Appl. Phys. Lett.*, **79**, 1181; (j) Nolan, C., Harris, N.C., and Kiang, C.-H. (2005) *Phys. Rev. Lett.*, **95**, 046101; (k) Sun, Y., Harris, N.C., and Kiang, C.-H. (2005) *Physica (Amsterdam)*, **350A**, 89; (l) Sun, Y., Harris, N.C., and Kiang, C.-H. (2005) *Physica (Amsterdam)*, **354A**, 1.
- 91** Pannopard, P., Khongpracha, P., Probst, M., and Limtrakul, J. (2008) *J. Mol. Graph. Modell.*, **26**, 1066.
- 92** (a) Porath, D., Bezryadin, A., de Vries, S., and Dekker, C. (2000) *Nature*, **403**, 635; (b) Fink, H.-W. and Schonenberger, C. (1999) *Nature*, **398**, 407; (c) Kasumov, A.Yu., Kociak, M., Guéron, S., Reulet, B., Volkov, V.T., Klinov, D.V., and Bouchiat, H. (2001) *Science*, **291**, 280; (d) Reichert, J., Ochs, R., Beckmann, D., Weber, H.B., Mayor, M.v., and Löhneysen, H. (2001) *Phys. Rev. Lett.*, **88**, 176804; (e) Xu, B. and Tao, N.J. (2003) *Science*, **301**, 122; (f) Hais, W., Nichols, R.J., van Zalingen, H., Higgins, S.J., Bethell, D., and Schiffrin, D.J. (2004) *Phys. Chem. Chem. Phys.*, **6**, 4330; (g) Piva, P.G., DiLabio, G.A., Pitters, J.L., Zikovsky, J., Rezeq, M., Dogel, S., Hofer, W.A., and Wolkow, R.A. (2005) *Nature*, **435**, 658; (h) Dadosh, T., Gordin, Y., Krahne, R., Khivrich, I., Mahalu, D., Frydman, V., Sperling, J., Yacobi, A., and Bar-Joseph, I. (2005) *Nature*, **436**, 677.
- 93** Zhang, Y., Austin, R.H., Kraeft, J., Cox, E.C., and Ong, N.P. (2002) *Phys. Rev. Lett.*, **89**, 198102.
- 94** (a) Demers, L.M., Östblom, M., Zhang, H., Jam, N.-H., Liedberg, B., and Mirkin, C.A. (2002) *J. Am. Chem. Soc.*, **124**, 11248; (b) Storhoff, J.J., Elghanian, R., Mirkin, C.A., and Letsinger, R.L. (2002) *Langmuir*, **18**, 6666; (c) Östblom, M., Liedberg, B., Demers, L.M., and Mirkin, C.A. (2005) *J. Phys. Chem. B*, **109**, 15150; (d) Hurst, S.J., Lytton-Jean, A.K.R., and Mirkin, C.A. (2006) *Anal. Chem.*, **78**, 8313.
- 95** (a) Kimura-Suda, H., Petrovykh, D.Y., Tarlov, M.J., and Whitman, L.J. (2003) *J. Am. Chem. Soc.*, **125**, 9014; (b) Petrovykh, D.Y., Kimura-Suda, H., Whitman, L.J., and Tarlov, M.J. (2003) *J. Am. Chem. Soc.*, **125**, 5219; (c) Yang, J., Pong, B.-K., Lee, J.Y., and Too, H.-P. (2007) *J. Inorg. Biochem.*, **101**, 824; (d) Brown, K.A., Sunho Park, S., and Hamad-Schifferli, K. (2008) *J. Phys. Chem. C*, **112**, 7517; (e) Yonezawa, T., Onoue, S.-Y., and Kimizuka, N. (2002) *Chem. Lett.*, 1172; (f) Weightman, P., Dolan, G.J., Smith, C.I., Cuquerella, M.C., Almond, N.J., Farrell, T., Fernig, D.G., Edwards, C., and Martin, D.S. (2006) *Phys. Rev. Lett.*, **96**, 086102; (g) Mohan, P.J., Datta, A., Mallajosyula, S.S., and Pati, S.K. (2006) *J. Phys. Chem. B*, **110**, 18661; (h) Pergolese, B., Bonifacio, A., and Bigotto, A. (2005) *Phys. Chem. Chem. Phys.*, **7**, 3610; (i) Hadjiladis, N., Pneumatikakis, G., and Basosi, R. (1981) *J. Inorg. Biochem.*, **14**, 115; (j) Kumar, A., Mishra, P.C., and Suhai, S. (2006) *J. Phys. Chem. A*, **110**, 7719.
- 96** (a) Chen, Q., Frankel, D.J., and Richardson, N.V. (2002) *Langmuir*, **18**, 3219; (b) Giese, B. and McNaughton, D. (2002) *J. Phys. Chem. B*, **125**, 1112;

- (c) Rapino, S. and Zerbetto, F. (2005) *Langmuir*, **21**, 2512; (d) Otero, R., Schöck, M., Molina, L.M., Lægsgaard, E., Stensgaard, I., Hammer, B., and Besenbacher, F. (2005) *Angew. Chem. Int. Ed.*, **44**, 2270; (e) Piana, S. and Bilic, A. (2006) *J. Phys. Chem. B*, **110**, 23467; (f) Otero, R., Xu, W., Lukas, M., Kelly, R.E.A., Lægsgaard, E., Stensgaard, I., Kjems, J., Kantorovich, L.N., and Besenbacher, F. (2008) *Angew. Chem. Int. Ed.*, **47**, 9673; (g) Otero, R., Lukas, M., Kelly, R.E.A., Xu, W., Stensgaard, I., Kantorovich, L.N., and Besenbacher, F. (2008) *Science*, **319**, 312.
- 97 Wells, D.H. Jr., Delgass, W.N., and Thomson, K.T. (2004) *J. Catal.*, **225**, 69 and references therein.
- 98 (a) Kryachko, E.S. (2009) *Pol. J. Chem.*, **000**, 000; (b) Kryachko, E.S. and Remacle, F. (2005) *Nano Lett.*, **5**, 735; (c) Kryachko, E.S. and Remacle, F. (2005) *J. Phys. Chem. B*, **109**, 22746.
- 99 Hadži, D. and Thompson, W.H. (eds) (1959) *Hydrogen Bonding*, Pergamon Press, London.
- 100 Pimentel, C.G. and McClellan, A.L. (1960) *The Hydrogen Bond*, Freeman, San Francisco.
- 101 Hamilton, W.C. and Ibers, J.A. (1968) *Hydrogen Bonding in Solids*, Benjamin, New York.
- 102 Schuster, P., Zundel, G., and Sandorfy, C. (eds) (1976) *The hydrogen bond, in Recent Developments in Theory and Experiments*, North-Holland, Amsterdam.
- 103 Schuster, P. (1978) *Intermolecular Interactions: From Diatomics to Biopolymers* (ed. B. Pullman), John Wiley & Sons, Ltd., Chichester, p. 363; (b) Schuster, P. (guest ed.) (1984) *Top. Curr. Chem.*, **120**.
- 104 Jeffrey, G.A. and Saenger, W. (1991) *Hydrogen Bonding in Biological Structures*, Springer, Berlin.
- 105 Jeffrey, G.A. (1997) *An Introduction to Hydrogen Bonding*, Oxford University Press, Oxford.
- 106 Scheiner, S. (1997) *Hydrogen Bonding. A Theoretical Perspective*, Oxford University Press, Oxford.
- 107 Hadži, D. (ed.) (1997) *Theoretical Treatment of Hydrogen Bonding*, John Wiley & Sons, Inc., New York.
- 108 (a) Steiner, T. and Desiraju, G.R. (1998) *Chem. Commun.*, 891; (b) Steiner, T. (2002) *Angew. Chem. Int. Ed.*, **41**, 48.
- 109 Desiraju, G.R. and Steiner, T. (1999) *The Weak Hydrogen Bond in Structural Chemistry and Biology*, Oxford University Press, Oxford.
- 110 (a) Bondi, A. (1964) *J. Phys. Chem.*, **68**, 441; (b) Rowland, R.S. and Taylor, T. (1996) *J. Phys. Chem.*, **100**, 7384.
- 111 (a) Kryachko, E.S. and Remacle, F. (2005) in *Theoretical Aspects of Chemical Reactivity* (ed. A. Toro Labbe), Theoretical and Computational Chemistry, vol. 16 (series ed. P. Politzer), Elsevier, Amsterdam, p. 219; (b) Kryachko, E.S. and Remacle, F. (2005) *Chem. Phys. Lett.*, **404**, 142; (c) Kryachko, E.S., Karpfen, A., and Remacle, F. (2005) *J. Phys. Chem. A*, **109**, 7309; (d) Kryachko, E.S. and Remacle, F. (2006) *Recent Advances in the Theory of Chemical and Physical Systems* (eds J.-P. Julien, J. Maruani, D. Mayou, S. Wilson, and G. Delgado-Barrio), Theoretical and Computational Chemistry, vol. 15, Springer, Dordrecht, p. 433; (e) Kryachko, E.S. and Remacle, F. (2007) *Topics in the Theory of Chemical and Physical Systems*, (eds S. Lahmar, J. Maruani, S. Wilson, and G. Delgado-Barrio), Progress in Theoretical Chemistry and Physics, vol. 16, Springer, Dordrecht, p. 161; (f) Kryachko, E.S. and Remacle, F. (2007) *J. Chem. Phys.*, **127**, 194305; (g) Kryachko, E.S. and Remacle, F. (2008) *Mol. Phys.*, **106**, 521; (h) Kryachko, E.S. (2008) *J. Mol. Struct.*, **880**, 23; (i) Kryachko, E.S. (2008) *Collect. Czech. Chem. Commun. R. Zahradnik Festschr.*, **73**, 000; (j) E.S. Kryachko (2009) in *Gold in Hydrogen Bonding Motif—Fragments of Essay. Demonstration of Nonconventional Hydrogen Bonding Patterns Between Gold and Clusters of Conventional Proton Donors* (eds N. Russo, V. Ya. Antonchenko, and E.S., Kryachko), The Proceedings of the NATO ARW “Molecular Self-Organization in Micro-, Nano-, and Macro-Dimensions: From Molecules to Water, to Nanoparticles, DNA and Proteins” dedicated to Alexander S. Davydov

- 95th birthday (June 8–12, 2008, Kiev, Ukraine) Self-Organization of Molecular Systems. From Molecules and Clusters to Nanotubes and Proteins. NATO Science for Peace and Security Series A: Chemistry and Biology, **14**, Springer, 315–334.
- 112** (a) Kryachko, E.S. and Sabin, J.R. (2003) *Int. J. Quantum Chem.*, **91**, 695; (b) Kryachko, E.S. (2003) *Fundamental World of Quantum Chemistry: A Tribute Volume to the Memory of Per-Olov Löwdin* (eds E.J. Brändas and E.S. Kryachko), Kluwer, Dordrecht, vol. **2**, p. 583.
- 113** (a) Chandra, A.K., Nguyen, M.T., Uchimar, T., and Zeegers-Huyskens, T. (1999) *J. Phys. Chem. A*, **103**, 8853; (b) Kryachko, E.S., Nguyen, M.T., and Zeegers-Huyskens, T. (2001) *J. Phys. Chem. A*, **105**, 1288, 1934.
- 114** Li, W., Haiss, W., Floate, S., and Nichols, R. (1999) *Langmuir*, **15**, 4875.
- 115** (a) Bilic, A., Reimers, J.R., Hush, N.S., and Hafner, J. (2002) *J. Chem. Phys.*, **116**, 8981; (b) Gollisch, H. (1984) *J. Phys. B*, **17**, 1463.
- 116** Close, D.M. (2003) *J. Phys. Chem. B*, **107**, 864 and references therein.
- 117** (a) Vázquez, M.-V. and Martínez, A. (2008) *J. Phys. Chem. A*, **112**, 1033; (b) Valdespino-Saenz, J. and Martínez, A. (2008) *J. Phys. Chem. A*, **112**, 2408; (c) Martínez, A., Dolgounitcheva, O., Zakrzewski, V.G., and Ortiz, J.V. (2008) *J. Phys. Chem. A*, **A112**, 10399.
- 118** Burda, J.V., Šponer, J., and Hobza, P. (1996) *J. Phys. Chem.*, **100**, 7250.
- 119** (a) Šponer, J., Sabat, M., Burda, J.V., Leszczynski, J., Hobza, P., and Lippert, B. (1999) *J. Biol. Inorg. Chem.*, **4**, 537; (b) Yanson, I., Teplitsky, A., and Sukhodub, L. (1979) *Biopolymers*, **18**, 1149.
- 121** (a) Wolf, S., Sommerer, G., Rutz, S., Schreiber, E., Leisner, T., Wöste, L., and Berry, R.S. (1995) *Phys. Rev. Lett.*, **74**, 4177; (b) Socaciu-Siebert, L.D., Hagen, J., Le Roux, J., Popolan, D., Vaida, M., Vajda, S., Bernhardt, T.M., and Wöste, L. (2005) *Phys. Chem. Chem. Phys.*, **7**, 2706; (c) Mitrić, R., Hartmann, M., Stanca, B., Bonačić-Koutecký, V., and Fantucci, P. (2001) *J. Phys. Chem. A*, **105**, 8892; (d) Bernhardt, T.M., Hagen, J., Socaciu-Siebert, L.D., Mitrić, R., Heidenreich, A., Le Roux, J., Popolan, D., Vaida, M., Wöste, L., Bonačić-Koutecký, V., and Jortner, J. (2005) *Chem. Phys. Chem.*, **6**, 243.

9

Quantum Mechanical Studies of Noncovalent DNA–Protein Interactions

Lesley R. Rutledge and Stacey D. Wetmore

9.1

Introduction

DNA–protein interactions play key roles in various processes that are vital to the survival of living organisms [1]. For example, repair of DNA damage, which can arise from exposure to external agents (medical X-rays, UV sunlight, tobacco smoke) or natural processes (replication errors), relies on nucleobase–amino acid interactions to selectively identify and remove damaged bases, while leaving natural bases intact [2, 3]. Furthermore, gene expression is regulated by protein switches that bind to specific DNA sequences [4, 5], which has led to proposals that DNA–protein interactions can be used to target genetic diseases through rational drug design [4, 6].

In attempts to understand the remarkable specificity with which proteins recognize DNA sequences [1], structural analysis of numerous DNA–protein complexes has been performed [7–9]. These studies reveal that it is not possible to establish a simple set of rules for predetermining interactions between DNA and protein building blocks [10, 11]. Indeed, nucleobases often interact with several amino acid side chains upon binding to a protein [12]. In addition, proteins often readily undergo structural changes to accommodate different nucleobases [10, 13, 14], where each substrate exploits unique active site interactions to promote binding [10]. Owing to this complexity, noncovalent interactions have been proposed to govern DNA–protein contacts [4, 15].

Noncovalent interactions are ideal for biological processes [16], which generally have a high dependency on the ease of DNA–protein complex formation. Specifically, to fulfill their function, noncovalent complexes must be stable, while at the same time they must readily degrade upon function completion. For example, damaged nucleobases must easily enter the active sites of DNA repair enzymes, but must also be easily removed to allow the protein to continue its function. This justifies the use of noncovalent interactions to facilitate important biological processes [16].

Examination of experimental structures reveals that one-third of direct DNA–protein interactions are specific hydrogen bonds between DNA base pairs and amino

acid side chains [4, 15]. The nature of these hydrogen-bonding interactions has been widely studied, and is fairly well understood. The remaining interactions, which include stacking, T-shaped, cation– π , electrostatic, hydrophobic and charge transfer, are believed to provide important clues for understanding the complete picture of DNA–protein binding since they likely contribute significantly to the overall complex stability [15]. Unfortunately, much less is known about these weaker noncovalent interactions [17].

A key factor missing in our quest to understand DNA–protein contacts and their implications is knowledge of the relative magnitude of different noncovalent interactions [12]. Computational chemistry (or molecular modeling) is an ideal tool for studying the structure and strength of molecular binding interactions. However, even with yearly improvements in computer power, large model systems are difficult to describe with a high level of accuracy, where computational resources (time, memory and disk) rapidly increase with both accuracy and model size. Furthermore, as suggested above, there are many different interactions and factors that contribute to DNA–protein recognition and binding.

A feasible computational approach for studying DNA–protein interactions is to start with the simplest system (two interacting monomers) and account for additional factors (synergy between contacts or environmental effects) in a step-by-step manner [12]. This approach allows scientists to use the highest-level (*ab initio*) quantum mechanical or density functional (DFT) techniques to obtain the most accurate structures and magnitudes of interaction possible. Through understanding each interaction at the molecular level, each contribution to the total stability of DNA–protein systems can be characterized, and vital clues about the relative importance of contacts will be revealed. This information can subsequently be used to understand fundamental biological processes, as well as how to exploit these interactions in applications ranging from protein design to drug discovery [17].

This chapter will focus on recent studies of DNA–protein noncovalent interactions using high-level computational techniques and small computational models. Specifically, we discuss and compare the magnitude of hydrogen-bonding, stacking, T-shaped and cation– π interactions between DNA (nucleobase and sugar–phosphate) and protein (amino acid backbone and side chain) components (Figure 9.1). We note that discrete interactions with water also play a large role in DNA–protein interactions; however, these interactions are beyond the scope of the present chapter and interested readers are directed to reviews on this topic [18–20]. Before discussing recent literature that analyzes direct contacts between DNA and protein components, the next section highlights the diverse range of computational approaches used to model these interactions.

9.2

Computational Approaches for Studying Noncovalent Interactions

Noncovalent interactions have been widely studied with various computational methodologies. These works span many different interactions and molecular

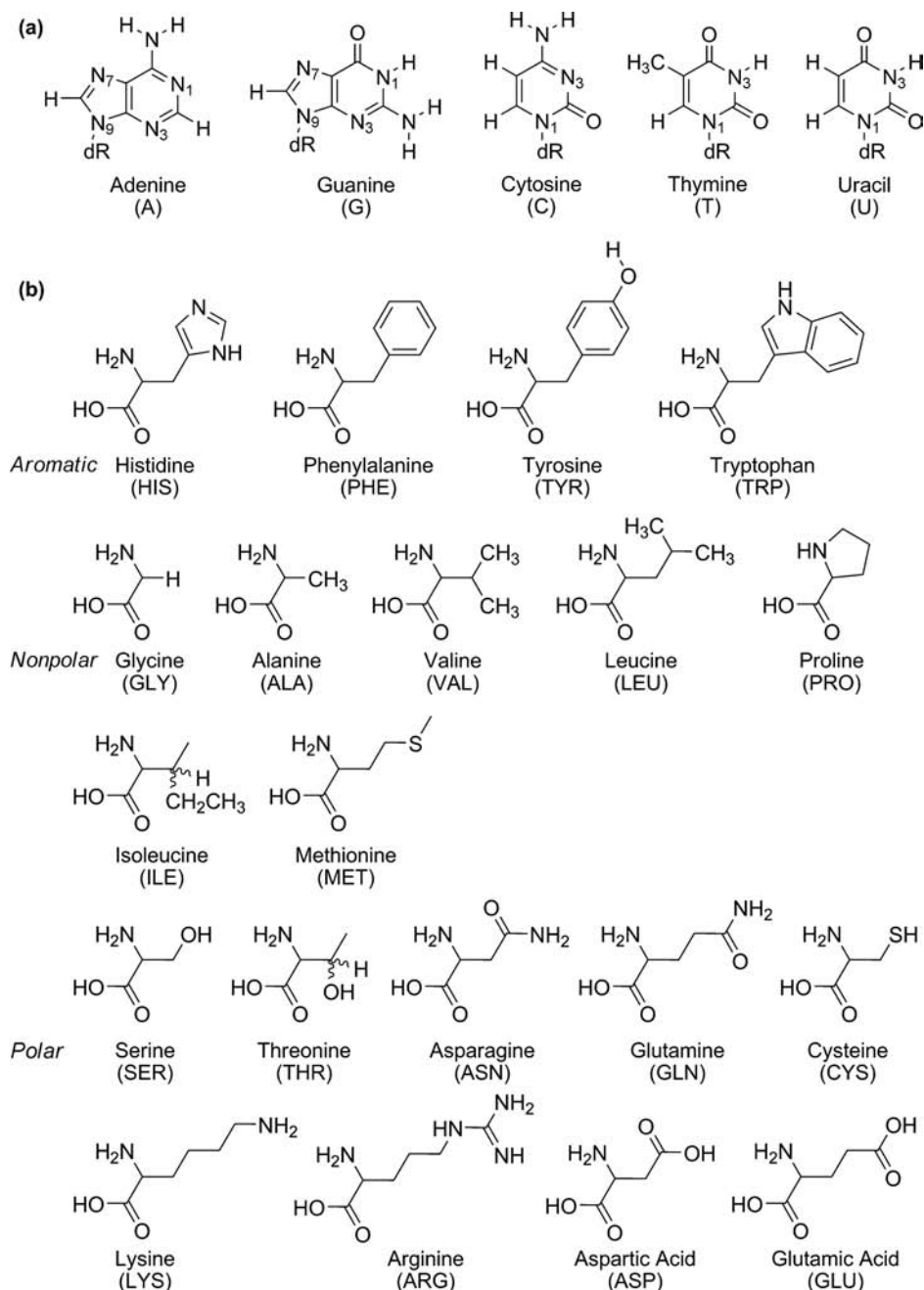


Figure 9.1 Structure and atomic numbering of (a) DNA and RNA nucleobases and (b) amino acids.

systems, including biomolecules. Indeed, a recent perspective provides a detailed discussion of computational approaches for studying noncovalent interactions between biomolecules [16]. This section summarizes key points from this perspective, as well as emphasizing methodologies that have been used to study DNA–protein complexes.

When modeling noncovalent interactions, the computational approach implemented depends on both the intrinsic interaction and the size of the model. For example, density functional theory (DFT) has been used to study hydrogen-bonding interactions, and has been proven to accurately describe the structures and strengths of a diverse range of complexes (when compared to reference data from experiment or wavefunction theories) [21, 22]. Owing to this accuracy, as well as reasonable cost for full geometry optimizations, DFT is typically the method of choice for studying hydrogen-bonding interactions between amino acids and nucleobases. The accuracy of DFT for hydrogen bonding is considered a consequence of error cancellations. Specifically, most currently used functionals do not properly account for (non-local) dispersion interactions, which can play an important role in hydrogen bonding, but rather these effects are accounted for through an overestimated attraction by the exchange functional [23].

Although DFT works well for hydrogen-bonded systems, “weaker” noncovalent interactions, like stacking and T-shaped, present greater challenges for computational modeling. Since these interactions include an even larger non-local dispersion contribution, their calculation requires the use of methods that recover a very large portion of the total electron correlation. Although DFT accounts for electron correlation, these methods completely fail to accurately describe “weak” noncovalent interactions due to an improper description of dispersion. Instead, higher-level quantum mechanical techniques must be used, which require an abundance of computer resources. Therefore, we are presently limited to studying dimers of biomolecules, although a few recent papers have considered trimers (or larger complexes) [24–26].

Among “weaker” noncovalent contacts, stacking interactions are perhaps the best studied to date. Several computational groups have rigorously examined benzene or substituted benzenes to identify the most accurate, yet cost effective, methods for modeling stacking of simple aromatics [27–30]. Stacking of natural or modified nucleobases has also been studied [31–39]. In principle, *ab initio* methods can be used to describe these interactions. However, in practice, due to small stabilization energies, only the most accurate methods should be used. Indeed, coupled cluster theory (CC) has been extensively implemented, where CCSD(T) has been found to be the “golden standard” [16].

Although the size of the system to which CCSD(T) can be applied grows with computer power, these calculations are expensive. When CCSD(T) is no longer practical, Møller–Plesset perturbation theory (MP2) with small basis sets has been shown to work well, while the use of larger basis sets leads to an overestimation of the MP2 interaction energy [16]. Perhaps the most widely used and successful MP2 basis set is the 6-31G*(0.25) variant, which replaces the standard d-exponent (0.8) with a value of 0.25 [32]. Although the success of this combination

is due to a cancellation of errors, this method has proven to rival the accuracy of CCSD(T) [31, 37].

As mentioned above, in addition to the level of theory [DFT, MP2, CCSD(T)], the choice of basis set used to describe the molecular orbitals is very important. Unfortunately, when a finite number of basis functions is used, basis set superposition error (BSSE) arises, where the basis functions from one molecule compensate for the basis set incompleteness on the other molecule and vice versa [40]. This results in a total dimer energy that is artificially too low or, in other words, the binding energy of the complex is overestimated [40]. Since BSSE can be a very large contribution to the energy [16], it must be eliminated using, for example, the counterpoise correction procedure of Boys and Bernardi [41].

Owing to large BSSE effects, it would be advantageous to use infinite basis sets when studying noncovalent interactions. In practice, this is done by extrapolating the total energy to the complete basis set (CBS) limit. Several extrapolation schemes have been suggested, among which that developed by Helgaker *et al.* [42] is probably the most commonly used today. This extrapolation uses systematically improved basis sets such as aug-cc-pVDZ and aug-cc-pVTZ (or even aug-cc-pVQZ for smaller systems). Since CCSD(T) calculations with these basis sets are not practical and CCSD(T) and MP2 energies have a similar dependence on the size of the basis set [36, 37], MP2 energies are generally extrapolated to the CBS limit. Subsequently, the difference between CCSD(T) and MP2 interaction energies is calculated using a medium (or small) sized basis set and added to the MP2/CBS energy to yield the CCSD(T)/CBS result [36, 37].

Although the MP2/6-31G*(0.25) method is suitable for reproducing higher-level stacking energies, this basis set cannot be used in geometry optimizations since it is not properly balanced [43]. Indeed, several factors must be considered to determine the structures of stacked complexes. First, owing to its large effect on the magnitude of “weak” interactions, BSSE will likely also have a large effect on geometry optimizations. Although small systems with high symmetry can be fully optimized on BSSE-free surfaces, the computational expense makes these calculations unfeasible for large systems with low symmetry, which includes most biologically-relevant models. Second, optimizations of stacked structures without BSSE corrections generally lead to hydrogen-bonded arrangements [32]. This phenomenon occurs for several reasons, such as the relative strength of stacking versus hydrogen-bonding interactions or the inability of computational methods to properly weight the importance of these interactions. Alternatively, poor starting guesses are likely used due to our currently limited understanding of optimal stacking orientations and lack of experimental data. In addition, if stacked structures are successfully optimized, individual monomers are often distorted, which may or may not be chemically relevant and cannot be verified in the absence of accurate experimental data. Furthermore, full optimizations of small models overlook natural structural constraints of, for example, protein or DNA backbones [43].

One approach for determining structures of complexes bound by “weak” noncovalent interactions is to develop new computational techniques. Indeed, recent literature is devoted to developing new density functional methods that correctly

describe dispersion [16, 44–47]. Owing to the efficiency of DFT, these new techniques may allow reliable geometry optimizations. Another more recent approach is resolution-of-identity MP2 (RI-MP2), which decreases calculation time while retaining MP2 accuracy [24, 48]. This technique has been mostly used for hydrogen-bonding interactions, as well as stacking interactions, in select systems [24, 49]. Although these new methodologies show promise, full geometry optimizations with these methods still suffer from some of the problems outlined above.

Another approach to gain information about stacking interactions between biomolecules is to use experimental crystal structures. Unfortunately, the resolution of protein structures does not allow identification of hydrogen atoms. Furthermore, addition of hydrogen atoms to the structure (either manually or through modeling programs) often leads to molecular distortion. An approach commonly used to resolve this issue is to overlay optimized monomer geometries onto experimental crystal structures and subsequently calculate the interaction energies. This approach generally leads to more realistic binding strengths than using fully optimized dimers.

A different approach for determining the structure of stacked complexes is to perform a potential energy surface (PES) scan. Hobza and Šponer have developed a technique that systematically varies the relative orientation of one monomer with respect to the other and uses a series of BSSE-free single-point calculations to identify the “optimal” interaction energy [32]. The structure(s) with the strongest interaction can subsequently be used as a starting point in full optimizations. A systematic approach for scanning the PES provides a very detailed understanding of how relative monomer orientations alter the interaction energy and thereby reveals important information about the nature of weak interactions that can only be conjectured from full optimizations [32]. PES scans can also provide estimates of the strengths of weaker contacts that occur in nature (e.g., due to protein folding), but are not necessarily the strongest interaction between two molecules in isolation.

Our group has adapted the technique developed by Hobza and Šponer [32] to study stacking between various biomolecules [38, 39, 50–52]. In our approach, three variables are systematically considered (Figure 9.2a): the vertical separation (R_1), the angle of rotation (α) and the horizontal shift (R_2). From a carefully defined starting dimer, the preferred R_1 is determined while holding α fixed and, subsequently, the preferred α is determined while holding R_1 at the optimal distance. In our scans, R_1 is typically varied between 3 and 4 Å by 0.1 Å increments, while α is altered in 30° increments from 0 to 360°. Although R_2 is held fixed at a “zero” defined by stacking the monomers via their centers of mass in these initial scans, R_2 is subsequently considered by moving one monomer in its molecular plane across a 3×3 Å grid in 0.5 Å increments, starting from the structure with the preferred R_1 and α , where the centers of mass are aligned in the middle of the grid. In general, we find that our method accurately searches the potential energy surfaces of stacked systems and that even more rigorous searches that simultaneously alter all variables are not required.

More recently, computational studies have begun to consider so-called T-shaped ($X-H \cdots \pi$, where $X = N, O, C$) interactions. Initial studies have focused on the

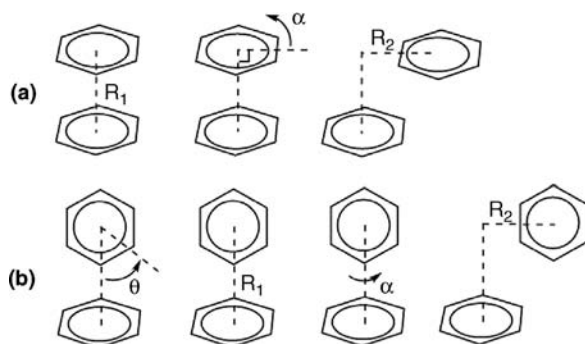


Figure 9.2 Definition of the variables considered in potential energy surface scans for DNA–protein complexes with (a) stacking (face-to-face) interactions [vertical separation (R_1), angle of rotation (α) and horizontal displacement (R_2)]; (b) T-shaped (edge-to-face) interactions [angle of “edge” rotation (θ), vertical separation (R_1), angle of rotation (α) and horizontal “edge” displacement (R_2)].

interactions between various small molecules and different aromatic rings [53–55]. For example, the CCSD(T)/CBS T-shaped interaction between benzene and methane has been estimated to be -6.1 kJ mol^{-1} [56]. These studies reveal that T-shaped interactions can be significant and therefore could be important in biology, a hypothesis that has been confirmed with studies on biomolecules or biomolecular fragments [56–59]. However, T-shaped interactions can also occur between two aromatic systems. Indeed, the stacking and T-shaped orientations of the benzene dimer have been extensively studied, where the T-shaped complex is isoenergetic with the stacked dimer [27–30]. These studies have shown that techniques implemented to study stacking interactions are also viable for T-shaped interactions. Nevertheless, there is little literature on T-shaped (or edge-to-face) interactions between two aromatic rings (other than benzene) [60] or between aromatic amino acids and nucleobases [61–63].

In our group, we have employed T-shaped potential energy surface scans analogous to those discussed above for stacking [62, 63]. Two major differences in the scans are (i) an additional variable (θ) is included (Figure 9.2b), which identifies the edge interacting with the π -system, where a number defines the bond directed at the π -system and a letter defines the edge bridging the π -system (Figure 9.3); and (ii) a more detailed R_2 shift is performed, where the “edge” monomer is shifted by $0.5 \text{ \AA}$ in four directions over the entire π -system of the “face” monomer.

Although most calculations are performed in the gas phase, environmental effects can significantly alter binding strengths. There are several ways that solvent can be accounted for in high-level quantum chemical calculations. For example, solvent molecules can be explicitly included in the computational model. Although this is potentially the most accurate way to treat solvent, it is difficult to determine the number of solvent molecules to include, while balancing computational costs. It is also difficult to determine the relative orientation of solvent molecules with respect to themselves, as well as to the solute. To avoid these problems,

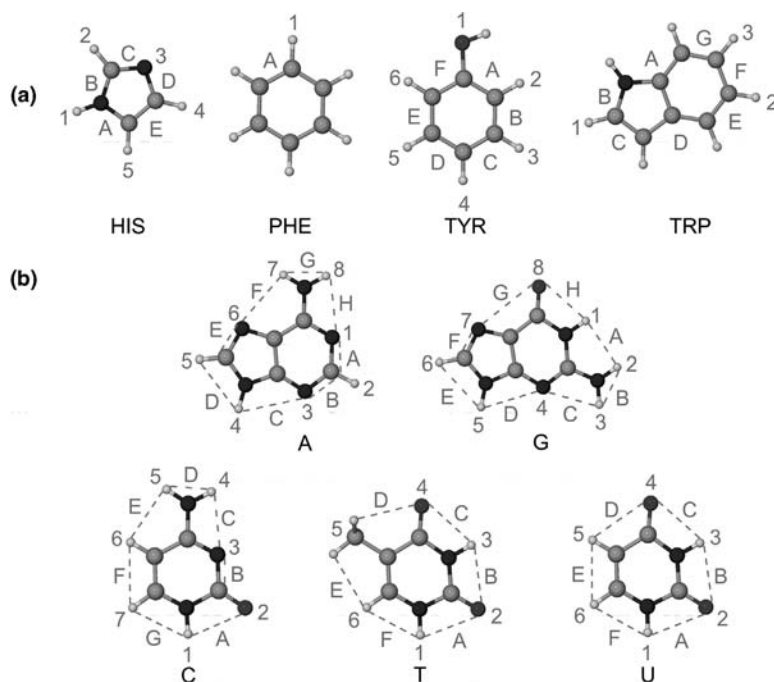


Figure 9.3 Definition of θ (the angle of "edge" rotation) for (a) amino acid edges and (b) nucleobase edges considered in potential energy surface scans of T-shaped contacts, where a number defines the bond directed at the π -system and a letter defines the edge bridging the π -system.

computational techniques that implicitly include solvent effects have also been developed. Perhaps the most widely used of these so-called continuum methods is the polarized continuum model (PCM), where solvent effects are modeled by an average field represented by a dielectric constant and the solute cavity is represented as a series of overlapping spheres centered on solute atoms [64]. A major problem with methods such as PCM is choosing dielectrics to mimic protein-like environments.

Before discussing specific literature, it should be reemphasized that the small model approach gives a detailed understanding of intrinsic interactions between DNA and protein components. At the same time, the high levels of theory used on these small models ensure accuracy of the calculated interaction energies. Indeed, computational studies on DNA nucleobases reveal that the highest levels of electron correlation (CCSD(T)) and largest basis sets (CBS limit) must be employed to reliably compare hydrogen-bonding and stacking interactions [16], while studies on the benzene dimer yield similar conclusions for stacking and T-shaped interactions [27–30]. Therefore, the small model approach is currently the best way to obtain information about the relative magnitude of "weak" noncovalent DNA–protein interactions.

9.3

Hydrogen-Bonding Interactions

Four different classes of hydrogen bonds occur at DNA–protein interfaces, which include those between (i) the protein and DNA backbones, (ii) the protein backbone and DNA bases, (iii) protein side chains and the DNA backbone and (iv) protein side chains and DNA bases [65, 66]. Interactions between the protein and DNA backbones commonly occur between pyrimidine nucleosides and hydrophobic amino acids such as Gly, Ala and Val due to their small size and lack of strong hydrogen-bond donors or acceptors [65]. Since, to the best of our knowledge, these interactions have not been studied using high-level calculations, we focus our discussion on the three remaining types of DNA–protein hydrogen-bonding interactions. We also note that hydrogen-bonding interactions between uracil and protein components have been heavily examined due to interest in RNA structure and function, as well as the structural similarities between uracil and thymine. Therefore, this section will also discuss hydrogen-bonding interactions involving the RNA nucleobase uracil.

9.3.1

Interactions between the Protein Backbone and DNA Nucleobases

The nature of interactions between the protein backbone and DNA nucleobases has been revealed through analysis of structures in the protein data bank (PDB). These hydrogen-bonding interactions typically occur between adenine or thymine and Ala or Gly [65]. Guanine is also frequently observed bound with Gly or Asn, while cytosine can interact with Lys, although this interaction is much less frequent [65].

In computational studies, the protein backbone is commonly modeled as formamide. Rozas' group have investigated the interactions between this model backbone and the four natural RNA nucleobases with B3LYP/6-31+G(d,p) [67]. The backbone was found to interact with each nucleobase through two medium strength hydrogen bonds (Figure 9.4). This leads to very stable structures that increase in strength according to $U (-48.0 \text{ kJ mol}^{-1}) \approx A (-48.4 \text{ kJ mol}^{-1}) < C (-63.7 \text{ kJ mol}^{-1}) < G (-78.6 \text{ kJ mol}^{-1})$. A similar trend was obtained by Alkorta and Elguero using a larger (2-formylaminoacetamide) backbone model [68]. Interestingly, few backbone–cytosine hydrogen-bonded complexes have been found despite the large calculated interaction [67].

Larger complexes involving more than one backbone or nucleobase have also been considered in the literature [68, 69]. For example, a backbone model has been bound to the Watson–Crick base pairs due to the potential importance of this binding in the recognition process [68]. These calculations show that the base pair structure does not significantly change upon backbone binding and that there are no cooperativity effects compared to binding with the individual bases. In another study, the interactions between adenine and up to seven formamide molecules were considered since multiple hydrogen bonds may also play a role in base recognition [69].

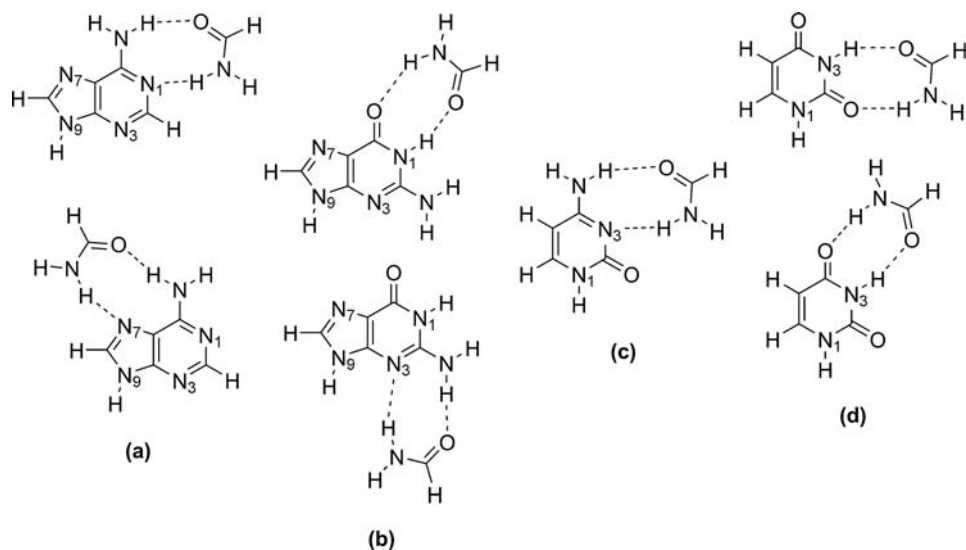


Figure 9.4 Hydrogen bonding between the protein backbone (formamide) and the nucleobases (a) adenine, (b) guanine, (c) cytosine and (d) uracil investigated in Reference [67].

9.3.2

Interactions between Protein Side Chains and DNA Backbone

Interactions between amino acid side chains and the DNA backbone are responsible for approximately half of all DNA–protein hydrogen-bonding interactions [65, 70]. Indeed, proteins initially recognize and bind to DNA through interactions with the charged phosphate groups [65]. These interactions are mainly responsible for stabilizing DNA–protein complexes rather than specificity [65]. These interactions may also play a structural role by inducing arrangements between DNA and proteins that align the macromolecules and allow specific interactions between side chains and nucleobases [65].

Interactions between protein side chains and the DNA backbone are primarily believed to involve the aromatic amino acids, as well as Arg, which interacts with the backbone through a strong electrostatic attraction [65]. Some of these interactions have been characterized with B3LYP/6-31G(d) by modeling the phosphate as $\text{OP}(\text{OH})_3$ and Lys as CH_3NH_2 , Arg as $\text{CH}_3(\text{NH})_2\text{CNH}_2$ and His as $\text{CH}_3\text{C}_3\text{N}_2\text{H}_3$ (Figure 9.5) [71]. The binding strengths of complexes involving two hydrogen bonds were found to form [deprotonated DNA phosphate] $^-$ [protonated side chain] $^+$ complexes for Arg ($-125.0 \text{ kJ mol}^{-1}$) and Lys ($-74.8 \text{ kJ mol}^{-1}$). In the case of His, no proton transfer from the DNA backbone was observed, but the calculated binding strength is significant ($-82.8 \text{ kJ mol}^{-1}$). These binding strengths support the potential importance of these stabilizing contacts. Interestingly, when these dimers are solvated, the binding strengths are independent of whether discrete water molecules surround the dimer or are located between the monomers, which validates

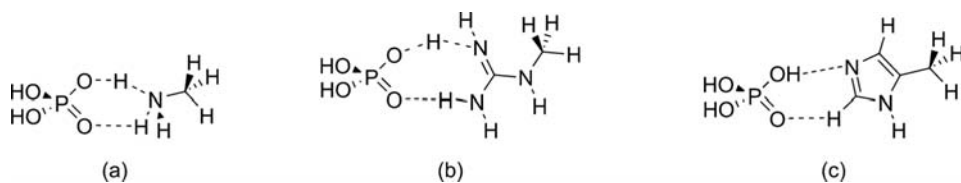


Figure 9.5 Hydrogen-bonding interactions between the phosphate backbone with (a) Lys, (b) Arg and (c) His investigated in reference [71].

suggestions that these interactions are for structure and stabilization of DNA–protein complexes rather than specificity [71].

9.3.3

Interactions between Protein Side Chains and DNA Nucleobases

Hydrogen bonding between protein side chains and the edges of DNA nucleobases plays a vital role in DNA substrate recognition [65]. These interactions typically occur between side chains, such as Asp, Glu, Asn, Glu, Lys and Arg, and the nucleobase edge atoms that appear in the major and minor grooves [65]. Since guanine has the largest number of potential hydrogen bonding atoms, it participates in these contacts most frequently, while contacts involving the smaller pyrimidines are observed less often [65]. Computational studies on these interactions are varied in their focus. For example, some studies consider a range of nucleobases and select amino acids [72–75], while others consider all possible binding modes between a range of amino acids and one nucleobase [76, 77]. We outline a few of these studies below.

The simplest hydrogen-bonded system that can be investigated is the uracil–glycine dimer, which involves two hydrogen bonds (Figure 9.6) [78–80]. Using B3LYP/6-31++G(d,p) calculations (including thermal corrections at 298 K), Rak's group have determined that these interactions range from -42.6 to -65.2 kJ mol $^{-1}$. Although the strongest interaction occurs at N1 of uracil, only slightly weaker binding is observed at other sites [N3(O4) and O2(N3)], which are more relevant to nucleosides and nucleotides. Our group also found very strong binding for a

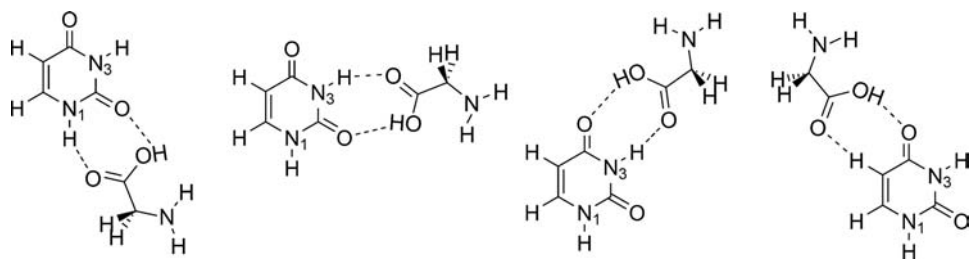


Figure 9.6 Hydrogen bonding in uracil–glycine dimers investigated in Reference [78].

complete range of amino acids at uracil sites that do not involve the glycosidic bond [77].

The second most commonly studied side chain is Asn or Gln. For example, the formamide model used by Rozas' group to consider interactions between the RNA nucleobases and the backbone is also a suitable model for studying nucleobase–Asn (Gln) interactions [67, 75]. Indeed, the large calculated interaction energies for guanine compared with the other nucleobases are in agreement with experiments that show the most frequent interactions occur between guanine and Asn (as well as Glu). Similarly, the calculated interactions with uracil support the small hydrogen-bond distances and frequent U: Asn contacts reported experimentally. In another study, Adamowicz's group used acrylamide to determine the role of the amide group in nucleobase–Asn(Gln) interactions using MP2 and B3LYP optimizations [72]. The calculated interactions decrease as $G (-65.0 \text{ kJ mol}^{-1}) > C (-54.8 \text{ kJ mol}^{-1}) > U (-51.4 \text{ kJ mol}^{-1}) > A (-40.9 \text{ kJ mol}^{-1})$, which shows a strong correlation with the trend in experimental enthalpies of formation ($G > C > A > U$). Based on thermodynamics, these latter results suggest that the amide group can distinguish between bases in single-stranded DNA.

Perhaps the most complete study of hydrogen-bonding interactions between various side chains and nucleobases has been made by Frankel's group [73, 74]. Specifically, interactions between RNA bases (A, C, G, U, A^+ and C^+) and many amino acids (Asp/Glu, His, His $^+$, Lys, Asn/Gln, Arg, Ser/Thr, Tyr, Trp) were considered using LMP2/6-31G(d,p)//HF/6-31G(d,p). Among interactions involving neutral side chains, G: Asn and C: Asn contacts are the most favorable (-85.6 and $-79.2 \text{ kJ mol}^{-1}$); however, these have not been observed in crystal structures. Alternatively, A: Asn interactions commonly found in DNA–protein complexes were found to bind with a similar strength ($-65.2 \text{ kJ mol}^{-1}$) as A: Ser complexes, which are not as common. These examples show that there is no correlation between binding strength and natural abundance of the contact. This suggests that sterics plays a key role in dictating the nature of DNA–protein hydrogen-bonding interactions.

9.4

Interactions between Aromatic DNA–Protein Components

In addition to hydrogen-bonding interactions, aromatic DNA or protein components can participate in stacking or T-shaped interactions, and these interactions may be strong enough to play roles in biological processes [4, 15, 81]. Indeed, stacking between nucleobases has been recognized to provide similar stabilization to DNA helices as hydrogen bonding in Watson–Crick base pairs [32]. Although various stacking or T-shaped interactions between different nucleobases and amino acids appear in nature and have been considered with computational methods, we will illustrate the major findings by first focusing on the interactions between adenine and histidine. These interactions have been thoroughly investigated by several groups since adenine is a fundamental building block commonly used to study

protein structure [17], while histidine can be neutral or protonated ($pK_a = 6.1$) and participates in many different noncovalent networks [61]. Our discussion will subsequently summarize important findings about interactions between all nucleobases and (aromatic) amino acids.

9.4.1

Stacking Interactions

A survey of interactions between adenine and histidine has been performed by Rooman's group, where a total of 14 different A: His contacts were found in a range of X-ray crystal structures (Figure 9.7) [61]. On average, the angle between the planes of the molecules was found to be 42° , which suggests that there is a slight preference for stacking interactions between these two molecules. The interaction energies of these complexes were estimated using MP2/6-31G(2d(0.8,0.2),p) single-point calculations on geometries obtained by overlaying HF/6-31G(d) optimized monomers onto crystal structure coordinates, where His was modeled as imidazole and adenine as the nucleobase. The (gas-phase) average interaction energies between A

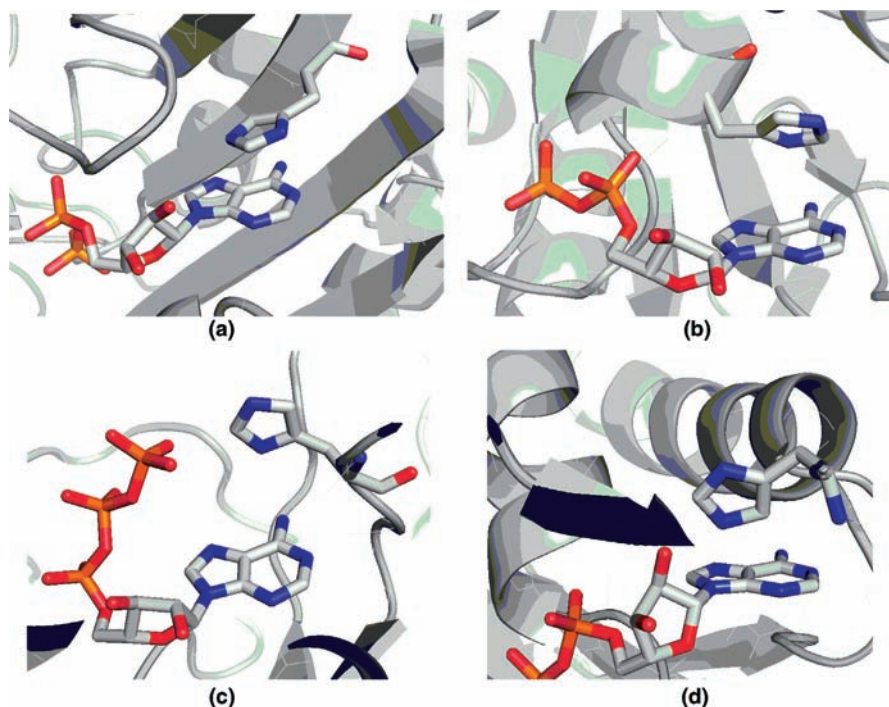


Figure 9.7 Representative examples of adenine–histidine stacking, (a) and (b), or T-shaped, (c) and (d), interactions identified in the PDB by Rooman *et al.* [61] PDB code: (a) 1BG0 [91]; (b) 2KIN [92]; (c) 1B8A [93]; (d) 1ZIN [94].

and His observed in these crystal structures is -15 kJ mol^{-1} . Furthermore, the strongest stacking interaction (-21 kJ mol^{-1}) was determined to be stabilized by significant dispersion.

In the same study, the interactions in stacked arrangements of adenine and histidine were studied by shifting His across the face of A, where two different tautomers of His were considered. Specifically, the geometric center of the His ring was placed in 12 positions over adenine, which correspond to the centers of the two adenine rings and the midpoints of all bonds in the purine ring. In these calculations, the molecular planes of both species were maintained in a parallel arrangement and separated by $3.5 \text{ \AA}$, the most common separation observed in X-ray structures. This study revealed that the interaction energy is greatly dependent on the position of the His ring, where the strongest interactions occur when the center of His is aligned with the center of the C4–C5 adenine bond.

Our group has performed a more thorough potential energy surface scan to investigate the stacking interactions between adenine and histidine using the same models [50, 51]. In our study, the relative orientation of optimized monomers was considered as a function of the vertical separation, angle of rotation and horizontal shift (Section 9.2 and Figure 9.2a), and two different orientations of histidine with respect to adenine were examined (Figure 9.8). The strongest stacking interactions were found to occur when the His ring is generally centered over the C4–C5 bond, which verifies the conclusion reported by Rooman’s group. Most importantly, the strongest interactions are -27 and -29 kJ mol^{-1} , which show that the optimal arrangement between two aromatic systems can lead to very large stacking energies. Furthermore, comparison to the results of Rooman’s group obtained using crystal structure orientations reveals that the interactions occurring in different biological systems are only approximately 9 kJ mol^{-1} weaker than the “optimal” (or largest possible) interaction.

Although Rooman’s study discussed above only identified adenine–histidine contacts in various X-ray structures, the search did not explicitly consider other amino acids [61]. However, the Hu group searched the PDB for structures that contained adenine to understand how binding proteins recognize this nucleobase [82]. Their

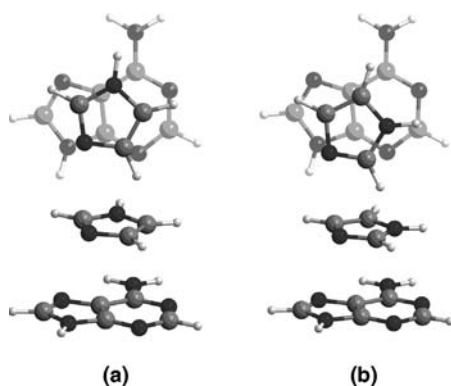


Figure 9.8 Orientations leading to the strongest (“optimal”) interaction for (a) A: His and (b) A: His; in Reference [51].

data mining revealed 68 complexes with adenine, where on average 2.7 hydrogen bonds, 1.0 stacking interactions, and 0.8 cation– π interactions were identified for each adenine. In total, 66 aromatic amino acids (Phe, Tyr, and Trp) residues were located within 5.6 Å of the adenine base in 44 structures. They further investigated the π – π stacking interactions of 9 adenine–aromatic amino acid complexes found in crystal structures (6 with Phe, 2 with Tyr, and 1 with Trp) using BSSE corrected MP2/6-311+G(d) calculations. These calculations reveal that the interactions depend on the intermolecular distance, orientation, and extent of π – π overlap, where the strongest crystal structure interaction was found to be $-26.6 \text{ kJ mol}^{-1}$ (adenine–phenylalanine complex modeled as the adenine–toluene heterodimer).

Very recently, 26 crystal structures from the PDB were identified by the Tschumper group that involve stacking interactions between adenine and phenylalanine [83]. The interaction energies of these 26 complexes (modeled as 9-methyladenine–toluene dimers) were investigated at the MP2 level and extrapolated to the CCSD(T)/CBS level of theory, and determined to range between -13.3 and $-28.3 \text{ kJ mol}^{-1}$. This study also optimized these complexes (initially with MP2/STO-3G++ (with even-tempered diffuse functions) and subsequently with MP2/DZP++), which resulted in only 6 unique structures with CCSD(T)/CBS binding strengths ranging from -24.8 to $-29.5 \text{ kJ mol}^{-1}$.

Due to the strength of these interactions, our group used detailed potential energy surface scans, like those discussed for the adenine–histidine system, to examine the stacking interactions between adenine and PHE, TYR or TRP [50, 51]. The strongest interactions were found for Trp (up to $-35.0 \text{ kJ mol}^{-1}$), followed by Tyr $\approx$ His, and Phe has the weakest interactions ($-24.3 \text{ kJ mol}^{-1}$). This trend was found to be related to the dipole moments of the amino acids, and also due to the relative size of their π -systems. Most importantly, even though Phe was found to have the weakest stacking energy, the interaction is still significant, which suggests that adenine can stack very strongly with all four aromatic amino acids.

Although Rooman's groups have examined many X-ray structures involving adenine His, contacts with nucleobases other than adenine were not identified [61, 82, 83]. However, since adenine is the most commonly used nucleobase for inhibitor or ligand building blocks, it is not surprising that these contacts were the majority of those found and does not necessarily rule out the possibility of other contacts occurring in natural systems. Indeed, based on our results for adenine, associations between the other nucleobases and aromatic amino acids have the potential to play important roles in biological processes, and therefore our group investigated these interactions (Figure 9.9) [51]. Our calculations indicate that regardless of the nucleobase considered the stacking interactions in nucleobase–(aromatic) amino acid dimers decrease as Trp > Tyr $\approx$ His > Phe. Furthermore, the trend with respect to the nucleobase was generally found to be G > A > T $\approx$ U $\approx$ C for all amino acids, which agrees with the relative dipole moments and size of the natural nucleobases, as well as trends in the binding strengths of natural nucleobase dimers.

Perhaps the most important finding regarding nucleobase–(aromatic) amino acid stacking is the magnitude of these interactions. The strongest binding strengths

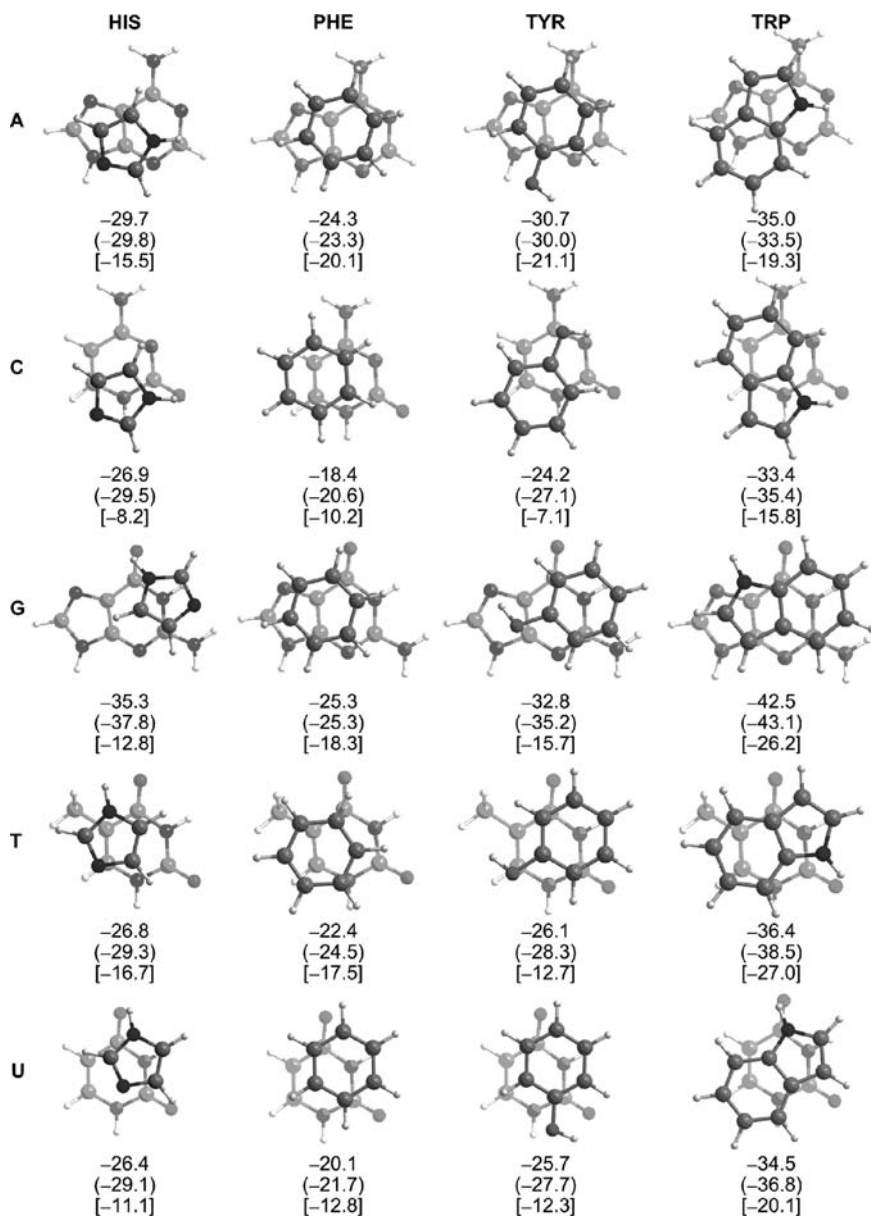


Figure 9.9 Orientations leading to the strongest (“optimal”) stacking interactions identified through MP2/6-31G* (0.25) potential energy surface scans. The MP2/6-31G* (0.25) interaction energies (kJ mol^{-1}) are reported below the appropriate structure, where the values in parentheses are the CCSD(T)/CBS interaction energies and values in square brackets are the interaction energies in (THF) solvent.

range between -18 and -43 kJ mol^{-1} , where two-thirds of these interactions are over -25 kJ mol^{-1} . Therefore, these interactions are very strong and approach the adenine–thymine Watson–Crick hydrogen-bond strength at the same level of theory ($-50.6 \text{ kJ mol}^{-1}$). This suggests that these interactions can play a much bigger role in biological processes than currently accepted. For example, due to differences in the stacking strengths between amino acids and nucleobases, these noncovalent interactions could be involved in recognition or specificity.

To ensure that conclusions based on MP2/6-31G*(0.25) binding strengths are legitimate, the stacking interactions between the aromatic amino acids and the natural nucleobases have been expanded to the CCSD(T)/CBS limit (Figure 9.9) [63]. The stacking interactions calculated at this most accurate level of theory are very close to the MP2/6-31G*(0.25) results. Indeed, the MP2/6-31G*(0.25) stacking energies were found to recover 89–104% of the CCSD(T)/CBS results. In addition to supporting our major conclusions, this justifies the use of MP2/6-31G*(0.25) to study these interactions and verifies that the interaction energies are being calculated on very accurate PESs. We note that, more recently, Cysewski has studied the stacking between the aromatic amino acids, as well as Arg, and cytosine or uracil using MP2/aug-cc-pVDZ scans and full optimizations [84]. Comparison of structures from both studies indicates that the geometries are identical and/or lead to binding strengths within 0.1 kJ mol^{-1} when calculated at the same level of theory.

Although the above stacking energies were calculated in the gas phase, environmental effects have been examined using the PCM solvation model and a range of dielectric constants [$\epsilon = 2$ (CCl_4) to 78 (water), where $\epsilon = 1$ corresponds to gas phase] [52]. Our calculations show that the interaction energy generally decreases with an increase in the dielectric constant of the surrounding medium, as shown for adenine dimers in Figure 9.10. Specifically, there is a large drop in the gas-phase stacking energy when solvents with small dielectrics [CCl_4 ($\epsilon = 2$) and diethyl ether ($\epsilon = 4$)] are considered, but as the dielectric constant further increases to THF ($\epsilon = 7$) and acetone ($\epsilon = 21$), the effects of increasing the dielectric constant decrease. Indeed, the environmental effects plateau for dielectric constants greater than acetone ($\epsilon = 21$), as binding strengths in acetone and DMSO ($\epsilon = 47$) are nearly equal. Despite the decrease in interaction energy upon inclusion of environmental effects, the stacking interactions between the natural nucleobase and (aromatic) amino acids are significant (up to -28 kJ mol^{-1}) in protein-like environments [THF (Figure 9.10) and acetone], as well as polar solvents. Therefore, the important conclusions from our gas-phase results still hold true. Specifically, stacking interactions likely play a large role in DNA–protein binding and could play a role in other processes, such as nucleobase recognition. However, roles in addition to stability remain to be determined.

9.4.2

T-Shaped Interactions

As for stacking, we focus our initial discussion of T-shaped contacts on interactions between adenine and histidine. Among the 14 A: His contacts identified in

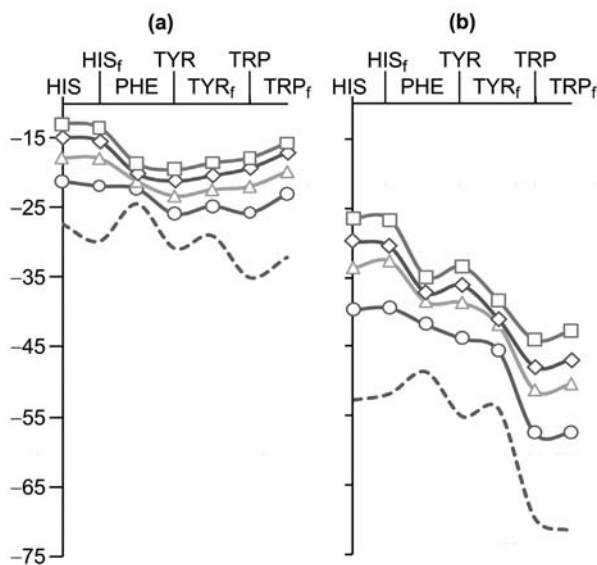


Figure 9.10 Interaction energies (kJ mol^{-1}) for the strongest stacked dimers between the aromatic amino acids and (a) adenine or (b) 3-methyladenine in the gas phase (dashed line), CCl_4 (circle), diethyl ether (triangle), THF (diamond) or acetone (square).

X-ray crystal structures by Rooman's group, six were found to correspond to T-shaped arrangements, which suggests that nucleobase-(aromatic) amino acid T-shaped interactions appear almost as often as stacking interactions [61]. Although systematic calculations were not performed on the A: His dimer, calculations on the His: Phe dimer suggest that His prefers to interact with the π -system of Phe through its acidic N–H bond. Furthermore, these T-shaped interactions were found to be comparable to His: Phe stacking interactions in the gas-phase and non-polar solvents, where the T-shaped interactions were dampened upon inclusion of environmental effects.

To gain more information about nucleobase-(aromatic) amino acid T-shaped contacts, our group recently investigated the interactions between the adenine π -system (face) and all His edges (Figure 9.3) [62]. We employed MP2/6-31G*(0.25) potential energy surface scans (Section 9.2 and Figure 9.2b) and therefore these interactions can be directly compared to those previously discussed for stacking. As suggested by Rooman's group, we found that the most favorable histidine (edge)-adenine (face) interaction occurs for the edge involving the acidic N–H bond. However, our results reveal that a bridged structure involving both N–H and C–H bonds directed towards adenine is slightly more favorable than the structure with only the N–H bond directed at the nucleobase (Figure 9.11). This is a very important finding for a general understanding of T-shaped interactions since most studies, including that by Rooman's group, do not consider these bridged structures. Perhaps even more importantly, the largest A: His T-shaped interaction ($-22.5 \text{ kJ mol}^{-1}$) is

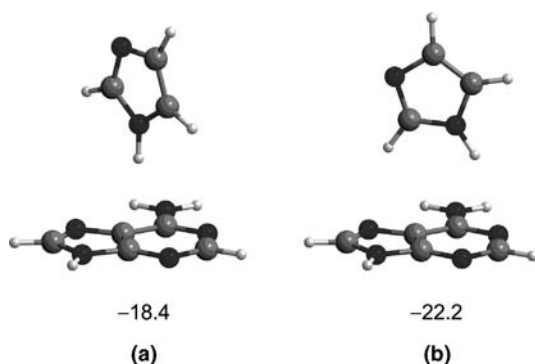


Figure 9.11 Structures and MP2/6-31G* (0.25) interaction energies (kJ mol^{-1}) for A(face): His(edge) dimers where (a) the N–H bond of His is directed towards adenine ($\theta = 1$) and (b) the N–H bond of His is bridging adenine ($\theta = \text{B}$).

almost the same as the largest stacking interactions (-27.2 and $-29.7 \text{ kJ mol}^{-1}$). Additionally, since our study considered all possible bond directed and bridged His edges interacting with adenine, we revealed the potential importance of a range of T-shaped interactions in nature. Indeed, His edge T-shaped interactions range between -9 and -23 kJ mol^{-1} .

We have also considered the interactions between all bond directed and bridged adenine edges and the His face [62]. Similarly to the His edge calculations, the strongest interaction occurs when the most acidic adenine edge is directed towards His. Since the most acidic adenine edge involves the model N–H glycosidic bond, the strongest interaction that is relevant for nucleoside or nucleotide substrates involves the adenine amino group. Both of these interactions are extremely stabilizing ($-33.6 \text{ kJ mol}^{-1}$ for the glycosidic bond and $-22.6 \text{ kJ mol}^{-1}$ for the amino group), and are almost as strong as, or stronger than, the stacking interactions between adenine and histidine, which justifies their potential importance.

As discussed in the previous section, interactions can also occur between adenine and the other aromatic amino acids (Phe, Tyr and Trp), as well as between the amino acids and other nucleobases. Therefore, our group has performed a full study of all of these T-shaped interactions [63]. As found for the A: His case study, the edge of the amino acid or nucleobase that leads to the strongest interactions with the π -system is the most acidic edge of the monomer. This corresponds to the edge including the N–H bond of His and Trp, the O–H bond of Tyr, and the glycosidic (N–H) bond of all nucleobases except guanine, where interactions with the N1 and N2 acidic protons lead to stronger interactions. In the case of nucleobase edge interactions, the strongest binding site that does not involve the (model) glycosidic bond involves the second most acidic nucleobase protons.

Nucleobase–amino acid T-shaped interactions involving an amino acid edge range between -10 (U: Phe) and -30 (C: Trp) kJ mol^{-1} , which suggest that these are extremely stabilizing contacts. In general, the maximum T-shaped interaction involving an amino acid edge decreases with amino acid according to $\text{Trp} > \text{His} >$

Tyr > Phe. This is the same trend discussed for stacking interactions and is dictated by the relative acidity of the edges involved. Interactions involving a nucleobase edge are even stronger (up to -48 kJ mol^{-1}), and are very close in magnitude to the corresponding stacking interaction.

As discussed for stacking, preliminary results for adenine indicate that extrapolation of MP2/6-31G*(0.25) to the CCSD(T)/CBS limit changes T-shaped binding strengths by a very small amount. However, T-shaped interactions are strengthened at CCSD(T) (by up to 5%), while stacking interactions are weakened at CCSD(T) (by up to 10%) [62]. Therefore, at the highest level of analysis possible, we find that adenine T-shaped interactions are at least as strong as, and sometimes slightly stronger than, the corresponding stacking interactions. This finding suggests that a variety of both stacking and T-shaped interactions should be considered when attempting to elucidate the role of DNA–protein contacts in biological processes. We note that this result has thus far only been found in the gas phase and future work must confirm these results in various environments.

9.5

Cation– π Interactions between DNA–Protein Components

Another extremely important class of noncovalent interactions in biology and chemistry is cation– π interactions. Over the past 20 or so years, many biological processes have been shown to rely on cation– π interactions. Ligand–antibody binding and receptor–ligand interactions are two examples [85]. Additionally, cation– π interactions between aromatic (Phe, Tyr, His, Trp) and cationic (Lys or Arg) amino acid side chains are frequently found in protein crystal structures [61]. Interfaces between proteins and DNA can also rely on these interactions, where the nucleobase acts as the aromatic moiety [61]. Alternatively, nucleobases can adopt cationic forms and thereby participate in cation– π contacts. For example, DNA repair enzymes that remove cationic (alkylated) nucleobases have been proposed to rely on cation– π interactions between charged nucleobases and aromatic amino acids to selectively remove damaged bases over the natural counterparts [2, 3]. The following sections briefly summarize selected computational studies that have attempted to elucidate DNA–protein cation– π interactions.

9.5.1

Cation– π Interactions between Charged Nucleobases and Aromatic Amino Acids

Our group has studied cation– π interactions between charged nucleobase and aromatic amino acids due to their potential role in DNA repair [2, 3]. Specifically, although highly polar groups capable of forming strong hydrogen bonds with substrates are found in the active sites of DNA glycosylases that remove (neutral) damaged nucleobases arising through oxidation or deamination [2, 3], the active sites of proteins that remove (cationic) alkylated nucleobases are lined with aromatic amino acids [2, 3]. This has led to suggestions that cation– π interactions are used

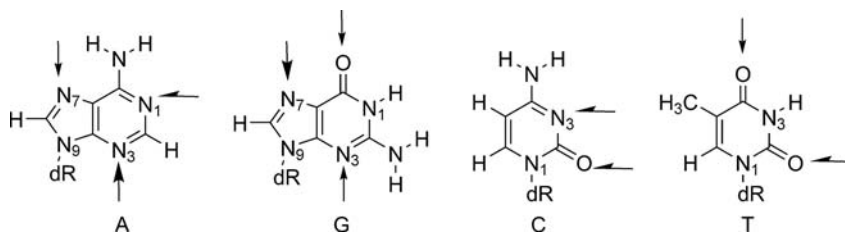


Figure 9.12 Structure of the natural nucleobases, with arrows indicating common methylation sites; large arrowheads identify the most common methylation sites and single-headed arrows indicate methylation sites that occur in single-stranded DNA.

to selectively recognize and remove alkylated nucleobases. Since little is known about the strength of these associations, our group has investigated the stacking interactions between the aromatic amino acids and the ten most common alkylated bases (Figure 9.12) [52].

To allow direct comparison to our studies of stacking interactions between natural nucleobases and amino acids, and thereby reveal the effects of nucleobase alkylation (cationic charge) on these interactions, MP2/6-31G*(0.25) potential energy surface scans were performed for dimers between damaged nucleobases and aromatic amino acids (Section 9.2 and Figure 9.2a). We observed that nucleobase methylation increases the stacking interaction energy by up to 40 kJ mol^{-1} , which corresponds to an increase of up to 135%. More specifically, the maximum stacking interactions with the amino acids for the natural bases range between -18 and -43 kJ mol^{-1} , but increase to -38 to -85 kJ mol^{-1} upon methylation (Table 9.1). These results indicate that the increase in stacking upon alkylation may be large enough for DNA repair enzymes to selectively remove the (cationic) alkylated bases over the natural (neutral) bases.

Our second major finding regarding the stacking interactions between the aromatic amino acids and cationic (alkylated) nucleobases is that the interaction energies vary with the alkylation site by up to 20 kJ mol^{-1} . For example, Figure 9.13 compares the stacking energies of guanine, 3-methylguanine, 7-methylguanine or O6-methylguanine and the aromatic amino acids. Interestingly, the stacking energies are not heavily dependent on whether methylation occurs at a ring nitrogen or exocyclic carbonyl. Instead, the magnitude of binding depends on the relative dipole moments of the adducts and the proton affinity of the alkylation site. Perhaps more importantly, the effects of methylation (up to -43 kJ mol^{-1}) are larger than the effects of the methylation site (up to -20 kJ mol^{-1}). Therefore, the differences in the stacking energies with respect to the alkylation site are small enough to explain why enzymes can remove various alkylation adducts, while leaving the natural bases intact.

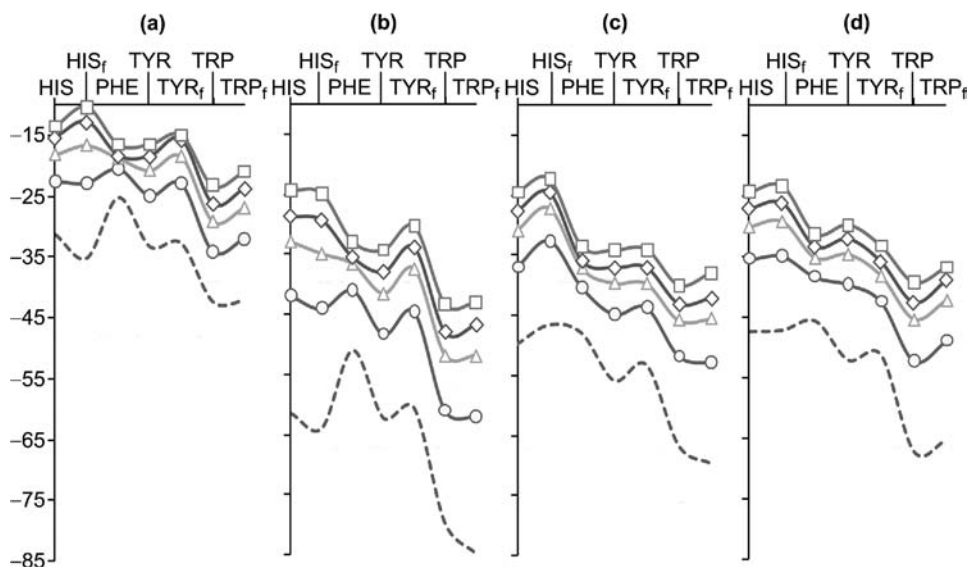
Our study also examined the effects of immersing the dimers in different solvents using PCM. As discussed previously for stacking interactions between two neutral aromatic systems, our calculations show that the stacking interactions decrease as the polarity of the solvent increases (Figure 9.13). Nevertheless,

Table 9.1 Strongest gas-phase MP2/6-31G* (0.25) stacking interactions (ΔE , kJ mol⁻¹) between the amino acids and the natural or methylated nucleobases as determined from potential energy surface scans.^{a),b)}

	His	His _f	Phe	Tyr	Tyr _f	Trp	Trp _f
Adenine	-27.2	-29.7	-24.3	-30.7	-28.9	-35.0	-32.0
1-Methyladenine	-49.8	-51.0	-45.9	-54.0	-53.3	-70.2	-69.6
3-Methyladenine	-52.7	-51.7	-48.6	-55.0	-3.9	-69.7	-71.5
7-Methyladenine	-61.9	-57.8	-51.0	-58.8	-59.5	-74.7	-71.3
Cytosine	-26.0	-26.9	-18.4	-22.7	-24.2	-32.9	-33.4
3-Methylcytosine	-52.1	-48.2	-43.3	-51.3	-50.2	-69.4	-70.5
O2-Methylcytosine	-38.9	-38.2	-38.5	-42.9	-42.7	-57.6	-59.9
Guanine	-31.4	-35.3	-25.3	-33.4	-32.8	-42.5	-42.4
3-Methylguanine	-61.5	-63.9	-51.0	-62.2	-60.6	-79.7	-84.7
7-Methylguanine	-49.8	-46.6	-48.0	-55.9	-53.3	-66.8	-69.8
O6-Methylguanine	-47.5	-47.3	-45.8	-52.2	-51.4	-67.2	-65.3
Thymine	-26.8	-25.0	-22.4	-25.5	-26.1	-36.4	-36.0
O2-Methylthymine	-54.1	-54.7	-48.9	-53.1	-55.6	-77.4	-74.7
O4-Methylthymine	-51.9	-47.8	-49.4	-54.5	-53.9	-77.7	-77.2

a) See Section 9.2 and Figure 9.2a for the definition of variables altered during MP2/6-31G* (0.25) potential energy surface scans.

b) Reference [52].

**Figure 9.13** Interaction energies for the strongest stacked dimers between the aromatic amino acids and (a) guanine, (b) 3-methylguanine, (c) 7-methylguanine or (d) O6-methylguanine in the gas phase (dashed line), CCl₄ (circle), diethyl ether (triangle), THF (diamond) or acetone (square).

the stacking interactions in various solvents are still very large. For example, in acetone, the stacking energies of the ten methylated nucleobases range between -15 and -44 kJ mol^{-1} . Since even in protein-like environments the interactions of alkylated nucleobases are greater than neutral bases, stacking interactions are a viable way for DNA repair enzymes to recognize and target damaged sites.

X-ray crystal structures show that interactions between nucleobase substrates and active site amino acids in DNA repair enzymes do not always involve parallel arrangements of the molecular planes [2]. Indeed, many T-shaped interactions also exist. Unfortunately, however, even less is known about the influence of cationic charge on T-shaped binding strengths than stacking interactions. Therefore, our group is currently investigating T-shaped interactions between methylated nucleobases and aromatic amino acids. Preliminary results for 3-methyladenine indicate that methylation, or the cationic charge, changes the preferred T-shaped orientation between the amino acid and base [62]. For example, the strongest T-shaped interaction involving a histidine edge occurs not when the most acidic N-H bond is directed towards adenine (Figure 9.11) but when the most basic edge (N lone pair) is directed towards 3-methyladenine (Figure 9.14). Therefore, our calculations reveal ways to identify the alkylation (or protonation) state of nucleobases in DNA-protein systems.

Our calculations also show that methylation significantly strengthens T-shaped contacts, where the interactions between 3-methyladenine and the amino acids are up to -70 kJ mol^{-1} . This suggests that these contacts are sometimes stronger than the corresponding stacking interactions. Furthermore, these contacts are much larger than the corresponding interactions with natural nucleobases. Owing to their calculated magnitude, we conclude that T-shaped interactions can also play an important role in selectively binding and removing alkylated nucleobases over their natural counterparts, and therefore are crucial to consider when attempting to

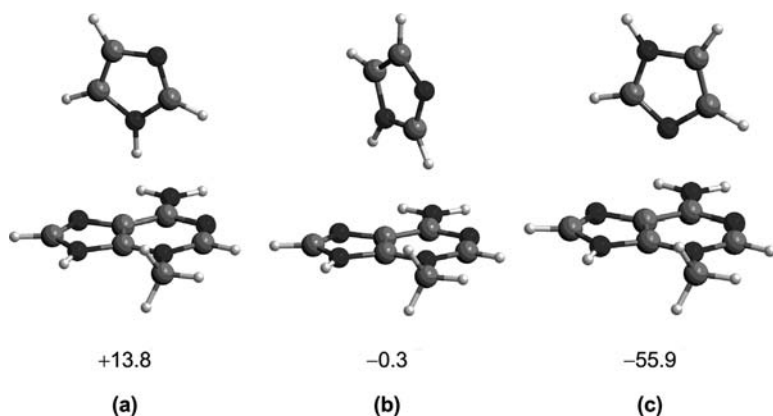


Figure 9.14 Structures and MP2/6-31G* (0.25) interaction energies (kJ mol^{-1}) for His(edge): 3MeA(face) dimers where (a) the N-H bond of His is directed towards 3-methyladenine ($\theta = 1$); (b) the N-H bond of His is bridging 3-methyladenine ($\theta = B$); (c) the lone pair of His is directed towards 3-methyladenine ($\theta = 3$).

understand biological processes. These results are currently being extended by considering the ten most common methylated nucleobases, as well as solvation effects on the magnitude of the interactions.

9.5.2

Cation– π Interactions Involving Charged Aromatic Amino Acids

Owing to its pK_a (6.1), histidine can appear in both neutral and protonated forms within active sites [61]. Since the protonation state of His cannot be directly determined from crystal structures, Rooman's group compared the stacking interactions of neutral and cationic histidine [61]. Indeed, by overlaying optimized monomers onto crystal structure coordinates, the average [MP2/6-31G(2d(0.8,0.2),p)] interaction energy between protonated histidine and adenine was determined to be approximately -32 kJ mol^{-1} , where the strongest interaction was found to be approximately -50 kJ mol^{-1} . Therefore, A: His⁺ cation– π interactions can significantly stabilize DNA–protein complexes. Indeed, these interactions are much stronger than the corresponding interactions between (neutral) His and adenine (Section 9.4.1). However, upon consideration of environmental effects using IEF-PCM at the HF/6-31G(2d(0.8,0.2),p) level, the average interaction energy decreases. Nevertheless, although the difference is drastically reduced, the protonated complexes are more stable than the neutral complexes in protein-like environments. This suggests that His⁺ cation– π interactions are likely involved in biological functions, which is supported by the sheer number of reported histidine contacts.

9.5.3

Cation– π Interactions Involving Charged Non-aromatic Amino Acids

Close contacts between nucleobases and cationic amino acid side chains (Lys or Arg) are frequently observed in protein crystal structures [86]. Since adenine is commonly used as an inhibitor (ligand) building block [17], there is an abundance of structural data for adenine in the form of ATP, ADP, AMP and ANP bound to proteins [86]. Indeed, scans of the protein data bank for crystal structures with a resolution of 2.5 Å or better identified 68 non-redundant “adenine” bound structures, where 48 (or 59%) of these structures involved cation– π interactions between adenine and Lys or Arg [86]. It was found that Lys typically occupies the adenine site near N7, while Arg lies above or below the base such that the guanidinium group is parallel to the π -system. Using representative contacts (Figure 9.15) and quantum mechanical [MP2/6-31+G(d)] calculations [86], the magnitudes of these interactions in crystal structure orientations were calculated to be -48.7 and $-36.6 \text{ kJ mol}^{-1}$ for Lys and Arg, respectively. Although the interactions decrease in water (-16.7 and -7.7 kJ mol^{-1} for Lys and Arg, respectively, using SM5.42R), this study shows that these interactions are significant and that positively charged residues can play an important role in biological processes such as recognition and binding.

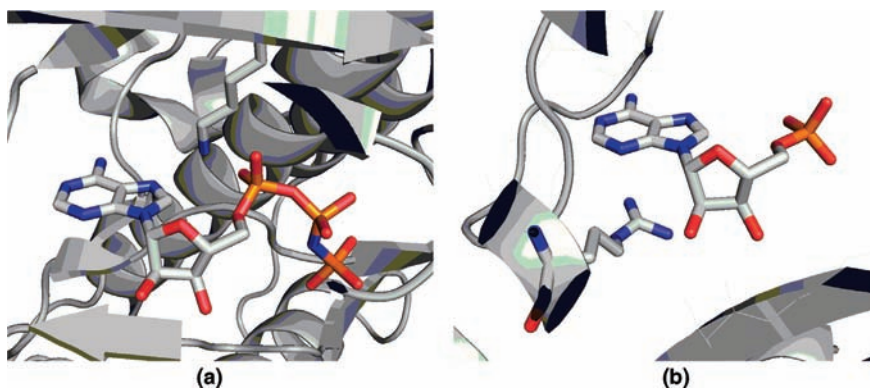


Figure 9.15 Representative examples of adenine interactions with (a) Lys (PDB code: 1IA9, [95]) and (b) Arg (PDB code: 12AS, [96]) identified by Mao *et al.* in Reference [86].

In addition to adenine interactions with Lys or Arg, interactions involving positively charged Asn and Gln residues have also been studied. Specifically, 55 non-redundant crystal structures have been identified that contain adenine cation- π interactions, where 38 of these interactions involve Arg, 7 Lys, 6 Asn and 6 Gln [87]. In nearly all of these contacts, the molecular planes of the amino acid and adenine are parallel; only nine interactions involve tilting of one molecule with respect to the other by more than 45° . MP2/6-31G(2d(0.8,0.2)) interaction energies obtained by overlaying optimized geometries onto crystal structures range between -5.9 and -23.4 kJ mol $^{-1}$ [87]. Furthermore, the average gas-phase free energies of binding were evaluated (by including zero-point vibrational, thermal and entropy corrections) to be -25.9 kJ mol $^{-1}$ for Lys, -37.6 kJ mol $^{-1}$ for Arg, and -18.8 kJ mol $^{-1}$ for Asn/Gln. Although solvation effects decrease the binding strengths, these adenine interactions are stabilizing in protein environments, where the interaction energies are approximately -6 kJ mol $^{-1}$ for Arg and Asn/Gln, but only -0.8 to $+10$ kJ mol $^{-1}$ for Lys. These calculations allow us to conclude that Arg, Lys, Asn and Gln cation- π interactions are common in DNA-protein systems due to their stabilizing nature.

Interactions between charged (non-aromatic) amino acid side chains and nucleobases other than adenine have also been scrutinized [87]. The most common interactions occur between Lys or Arg and the purine bases, while interactions involving C or T are very rare [87]. Quantum chemical [MP2/6-31G(2d(0.8,0.2))] calculations reveal that all interactions are strong. For example, the most commonly found A: Arg association has an average interaction energy of -23.4 kJ mol $^{-1}$. Guanine interactions are also very stabilizing; select overlaid structures yield binding strengths of -43.5 kJ mol $^{-1}$ for G: Arg, -16.3 kJ mol $^{-1}$ for G: Lys and -5.4 kJ mol $^{-1}$ for G: Asn. Thus, several different nucleobase-amino acid cation- π interactions can contribute to the stability of DNA-protein complexes.

LMP2/6-31G(d,p)//HF/6-31G(d,p) calculations have been completed on intermolecular interactions between cationic Lys, Asp/Glu or Arg and the natural

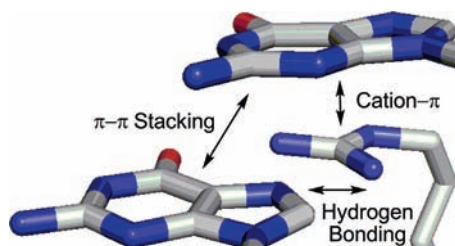


Figure 9.16 Representative example of a stair motif between guanine, guanine and arginine (PDB code: 1TC3 [97]) identified by Rooman *et al.* in References [88–90].

nucleobases that contain at least two contacts [74]. These interactions are extremely stabilizing, where the most favorable G: Lys interactions are about 45 kJ mol^{-1} more stable than G: Arg contacts. Interestingly, however, G: Lys contacts are only the third most common contact in crystal structures, while G: Arg contacts are frequently observed. Therefore, factors other than interaction energies account for the relative frequency of natural interactions.

9.5.4

Simultaneous Cation- π and Hydrogen-Bonding Interactions (DNA–Protein Stair Motifs)

Close examination of X-ray crystal structures reveals a reoccurring DNA–protein binding mode that involves two natural nucleobases (A, C, G or T) stacked in the orientation found within B-DNA, and a positively (Arg or Lys) or partially (Asn, Gln) charged amino acid hydrogen bonded to one of the bases and stacked with respect to the other [88–90]. Indeed, examination of 52 crystal structures with resolution better than $2.5 \text{ \AA}$ revealed 77 so-called “stair motifs” [88, 89]. The name originates from their structural resemblance to a stair, where the tread involves hydrogen bonding between the amino acid and one nucleobase and the riser involves π - π stacking between the nucleobases and cation- π interactions between the amino acid side chain and one nucleobase (Figure 9.16).

To gain a greater appreciation for the forces that stabilize stair motifs, the pairwise, as well as total, interaction energies have been evaluated using MP2/6-31G*(0.2) single-point calculations on crystal structure geometries [89]. The most favorable pairwise interaction was determined to be the G: Lys hydrogen bond ($-154.7 \text{ kJ mol}^{-1}$), the most stabilizing stacking interaction was found for G: C ($-40.0 \text{ kJ mol}^{-1}$) and the most favorable cation- π interaction was found for G: Arg ($-54.3 \text{ kJ mol}^{-1}$). Interestingly, through closer examination of stair motifs involving two guanine nucleobases and Arg [90], it was revealed that the sum ($-192.6 \text{ kJ mol}^{-1}$) of the three MP2/6-31G(2d(0.8,0.2),p) pairwise interactions [G: Arg cation- π interaction ($-56.8 \text{ kJ mol}^{-1}$), G: G stacking ($-12.5 \text{ kJ mol}^{-1}$) and G: Arg hydrogen bond ($-123.3 \text{ kJ mol}^{-1}$)] is greater than the total interaction energy calculated in the

presence of all three components ($-175.1 \text{ kJ mol}^{-1}$). This anticooperative behavior in the gas phase was found to be balanced by environmental effects, where inclusion of solvation leads to the anticipated result that these interactions are cooperative (i.e., the true interaction between all three components is stronger than the sum of the individual (pairwise) interactions).

In summary, the frequent appearance of stair motifs in DNA–protein interfaces suggests that these structures are important despite their currently unknown role. The calculations discussed above provide evidence that they play a stabilizing role. However, they may also play a specificity role since the calculated cation– π interactions depend on the type of nucleobase, as well as amino acid, involved. Furthermore, stair motifs may have a structural role since their presence requires very specific DNA conformations. Additional research is required to fully understand these systems.

9.6

Conclusions

The nature of DNA and protein interactions is extremely varied – surveys of crystal structures have revealed many different types of hydrogen-bonding, stacking, T-shaped and cation– π contacts between DNA and protein components. The observation of a larger number of unique DNA–protein contacts, and a current lack of understanding of their role in biological processes, provide testimony that a greater understanding of these interactions on a molecular level is needed. The examples discussed in the present chapter illustrate how quantum chemical studies using high-levels of theory and small model systems can provide clues about the nature of these interactions. The calculations show that a range of contacts are found in nature due to their significant strength, and therefore their stabilizing nature. Indeed, even stacking and T-shaped interactions between aromatic components are stronger than initially anticipated, and therefore are likely more important than previously conjectured. The calculations also hint that several interactions can contribute to the stability of DNA–protein complexes, and therefore a variety of contacts should be considered when attempting to elucidate the roles of these associations in biological processes. Although future work is required to fully understand the entire scope of DNA–protein interactions, the recent computational studies discussed in the present chapter have proven to be extremely valuable.

References

- 1 Gromiha, M.M., Siebers, J.G., Selvaraj, S., Kono, H., and Sarai, A. (2005) *Gene*, **364**, 108–113.
- 2 Berti, P.J. and McCann, J.A.B. (2006) *Chem. Rev.*, **106**, 506–555.
- 3 Stivers, J.T. and Jiang, Y.L. (2003) *Chem. Rev.*, **103**, 2729–2759.
- 4 Höglund, A. and Kohlbacher, O. (2004) *Proteome Sci.*, **2**, 3.
- 5 Ptashne, M. (1967) *Nature*, **214**, 232–234.
- 6 Bartsevich, V.V., Miller, J.C., Case, C.C., and Pabo, C.O. (2003) *Stem Cells*, **21**, 632–637.
- 7 Luscombe, N.M. and Thornton, J.M. (2002) *J. Mol. Biol.*, **320**, 991–1009.

- 8 Nadassy, K., Wodak, S.J., and Janin, J. (1999) *Biochem.*, **38**, 1999–2017.
- 9 Pailard, G. and Lavery, R. (2004) *Structure*, **12**, 113–122.
- 10 Matthews, B.W. (1988) *Nature*, **335**, 294–295.
- 11 Pabo, C.O. and Nekludova, L. (2000) *J. Mol. Biol.*, **301**, 597–624.
- 12 Sarai, A. and Kono, H. (2004), Chapter 7. in *Compact Handbook of Computational Biology* (eds A.K. Konopka and J.C. Crabbe), Marcel Dekker Inc., New York, USA.
- 13 Hogan, M.E. and Austin, R.H. (1987) *Nature*, **329**, 263–266.
- 14 Olson, W.K., Gorin, A.A., Lu, X.J., Hock, L.M., and Zhurkin, V.B. (1998) *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 11163–11168.
- 15 Luscombe, N.M., Laskowski, R.A., and Thornton, J.M. (2001) *Nucleic Acids Res.*, **29**, 2860–2874.
- 16 Černý, J. and Hobza, P. (2007) *Phys. Chem. Chem. Phys.*, **9**, 5291–5303.
- 17 Biot, C., Buisine, E., and Rooman, M. (2003) *J. Am. Chem. Soc.*, **125**, 13988–13994.
- 18 Jayaram, B. and Jain, T. (2004) *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 343–361.
- 19 Schwabe, J. (1997) *Curr. Opin. Struct. Biol.*, **7**, 126–134.
- 20 Tsui, V., Radhakrishnan, I., Wright, P., and Case, D. (2000) *J. Mol. Biol.*, **302**, 1101–1117.
- 21 Sim, F., Stamant, A., Papai, I., and Salahub, D.R. (1992) *J. Am. Chem. Soc.*, **114**, 4391–4400.
- 22 Sirois, S., Proynov, E.I., Nguyen, D.T., and Salahub, D.R. (1997) *J. Chem. Phys.*, **107**, 6770–6781.
- 23 Zhang, Y.K., Pan, W., and Yang, W.T. (1997) *J. Chem. Phys.*, **107**, 7921–7925.
- 24 Kabelac, M., Valdes, H., Sherer, E.C., Cramer, C.J., and Hobza, P. (2007) *Phys. Chem. Chem. Phys.*, **9**, 5000–5008.
- 25 Kabelac, M., Sherer, E.C., Cramer, C.J., and Hobza, P. (2007) *Chem.–Eur. J.*, **13**, 2067–2077.
- 26 Tauer, T.P. and Sherrill, C.D. (2005) *J. Phys. Chem. A*, **109**, 10475–10478.
- 27 Hobza, P., Selzle, H.L., and Schlag, E.W. (1994) *J. Am. Chem. Soc.*, **116**, 3500–3506.
- 28 Sinnokrot, M.O. and Sherrill, C.D. (2006) *J. Phys. Chem. A*, **110**, 10656–10668.
- 29 Waller, M.P., Robertazzi, A., Platts, J.A., Hibbs, D.E., and Williams, P.A. (2006) *J. Comput. Chem.*, **27**, 491–504.
- 30 Lee, E.C., Kim, D., Jurečka, P., Tarakeshwar, P., Hobza, P., and Kim, K.S. (2007) *J. Phys. Chem. A*, **111**, 3446–3457.
- 31 Šponer, J., Leszczynski, J., and Hobza, P. (1996) *J. Phys. Chem.*, **100**, 5590–5596.
- 32 Hobza, P. and Šponer, J. (1999) *Chem. Rev.*, **99**, 3247–3276.
- 33 Šponer, J., Jurečka, P., and Hobza, P. (2004) *J. Am. Chem. Soc.*, **126**, 10142–10151.
- 34 Cysewski, P. and Czyżnikowska-Balcerak, Z. (2005) *J. Mol. Struct. (THEOCHEM)*, **757**, 29–36.
- 35 Matta, C.F., Castillo, N., and Boyd, R.J. (2006) *J. Phys. Chem. B*, **110**, 563–578.
- 36 Jurečka, P., Šponer, J., Černý, J., and Hobza, P. (2006) *Phys. Chem. Chem. Phys.*, **8**, 1985–1993.
- 37 Šponer, J., Jurečka, P., Marchan, I., Javier Luque, F., Orozco, M., and Hobza, P. (2006) *Chem.–Eur. J.*, **12**, 2854–2865.
- 38 Wheaton, C.A., Dobrowolski, S.L., Millen, A.L., and Wetmore, S.D. (2006) *Chem. Phys. Lett.*, **428**, 157–166.
- 39 Rutledge, L.R., Wheaton, C.A., and Wetmore, S.D. (2007) *Phys. Chem. Chem. Phys.*, **9**, 497–509.
- 40 Jensen, F. (2007) *Introduction to Computational Chemistry*, 2nd edn, John Wiley and Sons, Ltd, Chichester, UK, pp. 225–227.
- 41 Boys, S.F. and Bernardi, F. (1970) *Mol. Phys.*, **19**, 553–566.
- 42 Halkier, A., Helgaker, T., Jorgensen, P., Klopper, W., Koch, H., Olsen, J., and Wilson, A.K. (1998) *Chem. Phys. Lett.*, **286**, 243–252.
- 43 Shi, Z., Olson, C.A., and Kallenbach, N.R. (2002) *J. Am. Chem. Soc.*, **124**, 3284–3291.
- 44 Grimme, S. (2004) *J. Comput. Chem.*, **25**, 1463–1473.
- 45 Johnson, E.R. and Becke, A.D. (2006) *Chem. Phys. Lett.*, **432**, 600–603.
- 46 Seifert, G. (2007) *J. Phys. Chem. A*, **111**, 5609–5613.
- 47 Sousa, A.F., Fernandes, P.A., and Ramos, M.J. (2007) *J. Phys. Chem. A*, **111**, 10439–10452.
- 48 Kendall, R.A. and Früchtl, H.A. (1997) *Theor. Chem. Acc.*, **97**, 158–163.

- 49 Jurečka, P., Nachtigall, P., and Hobza, P. (2001) *Phys. Chem. Chem. Phys.*, **3**, 4578–4582.
- 50 Rutledge, L.R., Campbell-Verduyn, L.S., Hunter, K.C., and Wetmore, S.D. (2006) *J. Phys. Chem. B*, **110**, 19652–19663.
- 51 Rutledge, L.R., Campbell-Verduyn, L.S., and Wetmore, S.D. (2007) *Chem. Phys. Lett.*, **444**, 167–175.
- 52 Rutledge, L.R., Durst, H.F., and Wetmore, S.D. (2008) *Phys. Chem. Chem. Phys.*, **10**, 2801–2812.
- 53 Vaupel, S., Brutschy, B., Tarakeshwar, P., and Kim, K.S. (2006) *J. Am. Chem. Soc.*, **128**, 5416–5426.
- 54 Bendová, L., Jurečka, P., Hobza, P., and Vondrášek, J. (2007) *J. Phys. Chem. B*, **111**, 9975–9979.
- 55 Mishra, B.K. and Sathyamurthy, N. (2007) *J. Phys. Chem. A*, **111**, 2139–2147.
- 56 Ringer, A.L., Figgs, M.S., Sinnokrot, M.O., and Sherrill, C.D. (2006) *J. Phys. Chem. A*, **110**, 10822–10828.
- 57 Gervasio, F.L., Chelli, R., Marchi, M., Procacci, P., and Schettino, V. (2001) *J. Phys. Chem. B*, **105**, 7835–7846.
- 58 Scheiner, S., Kar, T., and Pattanayak, J. (2002) *J. Am. Chem. Soc.*, **124**, 13257–13264.
- 59 Gil, A., Branchadell, V., Bertran, J., and Oliva, A. (2007) *J. Phys. Chem. A*, **111**, 9372–9379.
- 60 Tsuzuki, S., Mikami, M., and Yamada, S. (2007) *J. Am. Chem. Soc.*, **129**, 8656–8662.
- 61 Cauët, E., Rooman, M., Wintjens, R., Lievin, J., and Biot, C. (2005) *J. Chem. Theory Comput.*, **1**, 472–483.
- 62 Rutledge, L.R. and Wetmore, S.D. (2008) *J. Chem. Theory Comput.*, **4**, 1768–1780.
- 63 Rutledge, L.R., Durst, H.F., and Wetmore, S.D. (2009) *J. Chem. Theory Comput.*, **5**, 1400–1410.
- 64 Cossi, M., Barone, V., Cammi, R., and Tomasi, J. (1996) *Chem. Phys. Lett.*, **255**, 327–335.
- 65 Coulocheri, S.A., Pigis, D.G., Papavassiliou, K.A., and Papavassiliou, A.G. (2007) *Biochimie*, **89**, 1291–1303.
- 66 Mandel-Gutfreund, Y., Scheueller, O., and Margalit, H. (1995) *J. Mol. Biol.*, **253**, 370–382.
- 67 Rozas, I., Alkorta, I., and Elguero, J. (2004) *J. Phys. Chem. B*, **108**, 3335–3341.
- 68 Alkorta, I. and Elguero, J. (2003) *J. Phys. Chem. B*, **107**, 5306–5310.
- 69 Tang, K., Sun, H., Zhou, Z., and Wang, Z. (2008) *Int. J. Quantum Chem.*, **108**, 1287–1293.
- 70 Pabo, C.O. and Sauer, R.T. (1992) *Annu. Rev. Biochem.*, **61**, 1053–1095.
- 71 Pelmenchikov, A., Yin, X., and Leszczynski, J. (2000) *J. Phys. Chem. B*, **104**, 2148–2153.
- 72 Shelkovsky, V.S., Stepanian, S.G., Galetich, I.K., Kosevich, M.V., and Adamowicz, L. (2002) *Eur. Phys. J. D*, **20**, 421–430.
- 73 Cheng, A.C., Chen, W.W., Fuhrmann, C.N., and Frankel, A.D. (2003) *J. Mol. Biol.*, **327**, 781–796.
- 74 Cheng, A.C. and Frankel, A.D. (2004) *J. Am. Chem. Soc.*, **126**, 434–435.
- 75 Rozas, I., Alkorta, I., and Elguero, J. (2005) *Org. Biomol. Chem.*, **3**, 366–371.
- 76 Schlund, S., Mladenovic, M., Janke, E.M.B., Engels, B., and Weisz, K. (2005) *J. Am. Chem. Soc.*, **127**, 16151–16158.
- 77 Hunter, K.C., Millen, A.L., and Wetmore, S.D. (2007) *J. Phys. Chem. B*, **111**, 1858–1871.
- 78 Dabkowska, I., Gutowski, M., and Rak, J. (2002) *Pol. J. Chem.*, **76**, 1243–1247.
- 79 Dabkowska, I., Rak, J., and Gutowski, M. (2002) *J. Phys. Chem. A*, **106**, 7423–7433.
- 80 Dabkowska, I., Gutowski, M., and Rak, J. (2005) *J. Am. Chem. Soc.*, **127**, 2238–2248.
- 81 Baker, C.M. and Grant, G.H. (2007) *Biopolymers*, **85**, 456–470.
- 82 Mao, L., Wang, Y., Liu, Y., and Hu, X. (2004) *J. Mol. Biol.*, **336**, 787–807.
- 83 Copeland, K.L., Anderson, J.A., Farley, A.R., Cox, J.R., and Tschumper, G.S. (2008) *J. Phys. Chem. B*, **112**, 14291–14295.
- 84 Cysewski, P. (2008) *Phys. Chem. Chem. Phys.*, **10**, 2636–2645.
- 85 Peterson, E.J., Choi, A., Dahan, D.S., Lester, H.A., and Dougherty, D.A. (2002) *J. Am. Chem. Soc.*, **124**, 12662–12663.
- 86 Mao, L., Wang, Y., Liu, Y., and Hu, X. (2003) *J. Am. Chem. Soc.*, **125**, 14216–14217.
- 87 Biot, C., Buisine, E., Kwasigroch, J.M., Wintjens, R., and Rooman, M. (2002) *J. Biol. Chem.*, **277**, 40816–40822.

- 88 Rooman, M., Lievin, J., Buisine, E., and Wintjens, R. (2002) *J. Mol. Biol.*, **319**, 67–76.
- 89 Wintjens, R., Biot, C., Rooman, M., and Lievin, J. (2003) *J. Phys. Chem. A*, **107**, 6249–6258.
- 90 Biot, C., Wintjens, R., and Rooman, M. (2004) *J. Am. Chem. Soc.*, **126**, 6220–6221.
- 91 Zhou, G., Somasundaram, T., Blanc, E., Parthasarathy, G., Ellington, W.R., and Chapman, M.S. (1998) *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 8449–8454.
- 92 Sack, S., Muller, J., Marx, A., Thormahlen, M., Mandelkow, E.M., Brady, S.T., and Mandelkow, E. (1997) *Biochemistry*, **36**, 16155–16165.
- 93 Schmitt, E., Moulinier, L., Fujiwara, S., Imanaka, T., Thierry, J.C., and Moras, D. (1998) *EMBO J.*, **17**, 5227–5237.
- 94 Berry, M.B. and Phillips, G.N. Jr (1998) *Proteins*, **32**, 276–288.
- 95 Yamaguchi, H., Matsushita, M., Nairn, A.C., and Kuriyan, J. (2001) *Mol. Cell*, **7**, 1047–1057.
- 96 Nakatsu, T., Kato, H., and Oda, J. (1998) *Nat. Struct. Biol.*, **5**, 15–19.
- 97 van Pouderoyen, G., Ketting, R.F., Perrakis, A., Plasterk, R.H., and Sixma, T.K. (1997) *EMBO J.*, **16**, 6044–6054.

10

The Virial Field and Transferability in DNA Base-Pairing

Richard F.W. Bader and Fernando Cortés-Guzmán

10.1

A New Theorem Relating the Density of an Atom in a Molecule to the Energy

According to the theorem of Hohenberg and Kohn, molecules such as octane and hexane whose electron density distributions are shown in Figure 10.1, possess different external potentials and, therefore, possess different electron densities [1]. However, a chemist knows that the methyl and methylene groups make additive, transferable contributions to the molecular properties in this series of molecules. Figure 10.1 makes clear that transferability is a consequence of corresponding groups possessing transferable electron densities despite differing external potentials and that the densities of the groups determine their properties. Such observations have led to a new empirical theorem governing the density [2]:

The density of an atom in a molecule determines its contribution to the energy and to all other properties of the total system.

This theorem and its consequences on the pairing of DNA bases are subjects of this chapter.

The theorem of Hohenberg and Kohn, since it applies to a closed system with a fixed number of electrons, is of no relevance to the prediction and understanding of the role of a functional group in chemistry. Functional groups are the carriers of chemical information from one system (molecule or crystal) to another and they are necessarily open systems. They contribute additive contributions to all properties, contributions that vary from being simply characteristic of the group to being perfectly transferable. Since different chemical systems are characterized by different external potentials, the external potential cannot account for the characteristic properties observed for functional groups, the cornerstone of experimental chemistry.

Transferability is the result of the transferability of an atom's electron density and this is paralleled by the transferability of the *virial field*, the total electronic potential energy density [2, 3]. The aim of this chapter is to demonstrate that finding the basis for the existence of functional groups has implications beyond accounting for

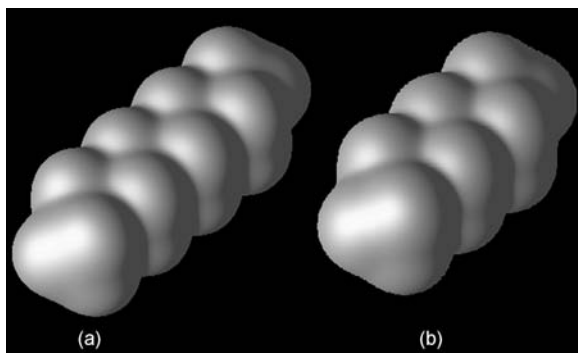


Figure 10.1 The 0.001 au density envelopes of (a) octane and (b) hexane molecules. The methyl and methylene groups are clearly discernible, a consequence of the C|C interatomic surfaces intersecting the density

envelope. No one viewing this picture can deny the existence of atoms in molecules nor question the role of the electron density as the vehicle for the transmission of chemical information.

transferability of properties. A point of particular importance is that only the total virial field is of consequence, its individual contributions having no final bearing on the properties of an atom in a molecule. Thus, finding that interactions between molecules, such as those encountered in hydrogen bonding between DNA base pairs [4], lead to significant changes in the electrostatic potential removed from the site of the interaction, are called into question as such changes are in general compensated for by other changes leaving the virial field and hence the atoms removed from the reaction site, little changed. This result questions the use of models that treat less than the entire interaction in discussions of intermolecular interactions.

As first demonstrated in 1972, the properties of an atom in a molecule are determined by its electron density distribution, the change in its contributions to the properties on transfer between systems reflecting the degree of change in its density, with perfect transferability being the limit of perfect transferability of the atom's electron density [3]. Of course, all of these statements are predicated upon an atom being a proper open quantum system [5] with properties defined by the Heisenberg equation of motion [6, 7]. Anyone unfamiliar with the variational derivation of this statement from Schwinger's principle of stationary action [8] or with its underlying physics may use their knowledge of Schrödinger's equation to derive this result for themselves in a heuristic manner [9].

Everything that follows hinges on the observation that all measurable properties of atoms or functional groupings of atoms are recovered by the physics of proper open systems [10, 11]. Thus while the theorem of Hohenberg and Kohn states that the density determines the energy of a closed system, the corresponding theorem of importance to chemistry requires a new theorem [2]:

The density of a proper open system determines its contribution to the energy and to all other properties of the total system.

It is this property and the observed transferability of the density that accounts for the additive, characteristic properties observed for functional groups that are found despite the unavoidable changes in the external potential – the attractive contribution to the virial – that occur on the transfer of groups between different systems. The present study demonstrates that the same cancellation of the attractive and repulsive contributions to the virial occur during chemical change, as exemplified in DNA bases-pairing, with the chemistry being determined by the changes in the virial field.

The predictions of quantum mechanics are unique and, thus, so is the ability of proper open systems [5] to recover the experimental properties of atoms in molecules [9]. It is, for example, well documented [10, 11] that the partitioning of the density proposed in the “stockholder” method of overlapping atoms and introduced into density functional theory by Nalewajski and Parr [12] is incapable of recovering the measured properties of functional groups. Thus, the recent proposal by Gázquez *et al.* [13] to use this approach to account for the chemistry of functional groups is of questionable value, as it will never recover their experimentally determined properties, which is the ultimate goal of a scientific investigation. Their criticism that the absence of overlap between the atomic densities and its replacement by the zero-flux partitioning surface in QTAIM (quantum theory of atoms in molecules) will result in an inadequate inter-atomic density to describe chemical bonding confuses the physical property of interest – the density in this case – with the model or method used to determine it. QTAIM is based on the total density, not on any imagined atomic contributions, and the simple existence of molecules is proof that the density employed in QTAIM is sufficient to describe chemical bonding.

10.2

Computations

The starting geometries of the base pairs were obtained from Hobza and Sponer [14] and optimized using the 6-311++G(2p,2d) basis set in calculations using the hybrid functional B3LYP given in Gaussian 03 [15]. DFT calculations give stabilization energies for the DNA base pairs close to those obtained from MP2 calculations, with differences lying in the range -0.9 to 1.4 kcal mol⁻¹ (1 kcal = 4.184 kJ). For the CC base pair, for example, the stabilization energies are -18.8 and -17.5 kcal mol⁻¹ for MP2 and B3LYP, respectively [16]. The bond and atomic properties were calculated using AIMALL [17] and AIM2000 [18] programs, the latter also being used in the construction of the diagrams. The differences between the sums of the integrated atomic energies and the molecular values are less than 0.5 kcal mol⁻¹. The small differences indicate that the errors in the atomic integrations are negligible.

10.3

Chemical Transferability and the One-Electron Density Matrix

In a 1995 paper entitled “Chemistry and the nearsighted nature of the one-electron density matrix” [19] it was argued that chemistry is a consequence of the

near-sightedness of $\Gamma^{(1)}(\mathbf{r}, \mathbf{r}')$, since this matrix determines the electron density and, through the virial theorem, all of the mechanical properties of an atom in a molecule. There is an additional important observation, bolstering this statement: as previously demonstrated, all necessary physical information is contained in the expansion of $\Gamma^{(1)}(\mathbf{r}, \mathbf{r}')$ up to second-order with regard to both the diagonal and off-diagonal terms [6, 20]. The diagonal terms yield the density $\varrho(\mathbf{r})$, the gradient vector field of the density $\nabla\varrho(\mathbf{r})$ that determines structure and structural stability [21] and the dyadic $\nabla\nabla\varrho(\mathbf{r})$ that determines the critical points in the density. The trace of the dyadic yields the Laplacian of the density $\nabla^2\varrho(\mathbf{r})$, the bridge that provides a homeomorphic mapping of the information determining the spatial pairing of electrons [22, 23] contained in the second-order density matrix. The off-diagonal terms yield the current density $\mathbf{J}(\mathbf{r})$, the stress tensor $\vec{\sigma}(\mathbf{r})$ that determines the Ehrenfest force and energy densities [20] and the divergence of the current $\nabla\cdot\mathbf{J}(\mathbf{r})$, the field that determines the critical points in $\mathbf{J}(\mathbf{r})$ [24].

10.3.1

The Virial Field

The derivation of the Ehrenfest and virial theorems obtained from the Heisenberg equation of motion for an open system for the generators $\mathbf{p}$ and $\mathbf{r} \cdot \mathbf{p}$, respectively, has been reviewed on several occasions [6]. The commutator term for the momentum operator $(i/\hbar)[H, \mathbf{p}]$ yields $-\nabla_r \hat{V}$, which determines the force exerted on the electron at $\mathbf{r}$ by the remaining electrons and by the nuclei, all in fixed positions. Taking the expectation value of this force by summing over all spins followed by the integration of all the electronic coordinates save those denoted by the position vector $\mathbf{r}$, an operation denoted by $N \int d\tau'$, one obtains an expression for $\vec{\mathcal{F}}(\mathbf{r})$, the force exerted on an electron at position $\mathbf{r}$ by the *average distribution* of the remaining electrons and by the rigid nuclear framework – a “dressed density” – giving the force exerted on the electron density.

The physics of an open system defines a corresponding “dressed” density distribution for every measurable property, one whose integration over an atomic basin yields the atom’s additive contribution to that property. A dressed density distribution for some particular property accounts for the corresponding interaction of the density at some point in space with the remainder of the molecule [25].

Such dressed densities are clearly important in the discussion of the effect of distant contributions to the transferability of atomic properties. This physics is summarized in Equation 10.1:

$$\vec{\mathcal{F}}(\Omega) = \int_{\Omega} d\mathbf{r} N \int d\tau' \{ \psi^* (-\nabla_r \hat{V}) \psi \} = \int_{\Omega} d\mathbf{r} \vec{\mathcal{F}}(\mathbf{r}) = - \oint dS(\Omega; \mathbf{r}_s) \vec{\sigma}(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r}) \quad (10.1)$$

which introduces the quantum stress tensor $\vec{\vec{\sigma}}(\mathbf{r})$, the vehicle for condensing the many-particle interactions in the potential energy operator $\hat{V}$ into a real-space expression, the local expression for the Ehrenfest force being expressible as $\vec{F}(\mathbf{r}) = -\nabla \cdot \vec{\vec{\sigma}}(\mathbf{r})$, as made clear by the surface term in Equation 10.1. The stress tensor, defined in terms of the one-electron density matrix, is given in Equation 10.2:

$$\vec{\vec{\sigma}}(\mathbf{r}) = N(\hbar^2/4m)\{(\nabla\nabla + \nabla'\nabla') - (\nabla\nabla' + \nabla'\nabla)\}\Gamma^{(1)}(\mathbf{r}, \mathbf{r}')|_{\mathbf{r}=\mathbf{r}} \quad (10.2)$$

The open system expectation value of the commutator for the virial operator $\hat{G}(\mathbf{r}) = \hat{\mathbf{r}} \cdot \hat{\mathbf{p}}$ yields $2T(\Omega) + \mathcal{V}_b(\Omega)$, which is twice the atom's electronic kinetic energy plus $\mathcal{V}_b(\Omega)$, the virial of the Ehrenfest force exerted over the basin of the atom. In a stationary state these contributions are balanced by $\mathcal{V}_s(\Omega)$, the virial of the Ehrenfest force acting over the surface of the atom. Expressing by $\mathcal{V}(\Omega)$ the total virial for atom Ω , the virial theorem for a stationary state may be stated as:

$$-2T(\Omega) = \mathcal{V}(\Omega) = \mathcal{V}_b(\Omega) + \mathcal{V}_s(\Omega) \quad (10.3)$$

The virials of the Ehrenfest force exerted over the basin and the surface of the atom expressed in terms of the stress tensor with the origin for the coordinate $\mathbf{r}$ placed at the nucleus of atom Ω are given in Equations 10.4 and 10.5a:

$$\mathcal{V}_b(\Omega) = -\int_{\Omega} d\mathbf{r} \mathbf{r}_{\Omega} \cdot \nabla \cdot \vec{\vec{\sigma}}(\mathbf{r}) = \int_{\Omega} d\mathbf{r} \mathbf{r}_{\Omega} \cdot \vec{F}(\mathbf{r}) \quad (10.4)$$

$$\mathcal{V}_s(\Omega) = \oint dS(\Omega, \mathbf{r}_s) \mathbf{r}_{\Omega} \cdot \vec{\vec{\sigma}}(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r}) \quad (10.5a)$$

The virial $\mathcal{V}$ of the Ehrenfest force acting on the electrons over the entire system, obtained from Equation 10.4 with $\Omega = \mathbb{R}^3$, equals the total potential energy V and the virial of the Feynman forces on the nuclei, the result expressed in Equation 10.5b:

$$\mathcal{V} = V_{\text{en}} + V_{\text{ee}} + V_{\text{nn}} - \sum_{\alpha} \mathbf{X}_{\alpha} \cdot \mathbf{F}_{\alpha} = V - \sum_{\alpha} \mathbf{X}_{\alpha} \cdot \mathbf{F}_{\alpha} \quad (10.5b)$$

The virial relations obtained for the total system are recovered for an open system and thus one obtains the usual statements of the virial theorem when applied to an atom in a molecule:

$$T(\Omega) = -E(\Omega) + W(\Omega) \text{ and } \mathcal{V}(\Omega) = 2E(\Omega) - W(\Omega) \quad (10.6)$$

where $W(\Omega)$ is the atomic contribution of the virial of the external (Feynman) forces acting on the nuclei. Each theorem obtained from the Heisenberg equation can be stated in a local form, the local form of the virial theorem being:

$$(\hbar^2/4m)\nabla^2 Q(\mathbf{r}) = 2G(\mathbf{r}) + \mathcal{V}(\mathbf{r}) \quad (10.7)$$

where the virial field $\mathcal{V}(\mathbf{r})$ may be expressed as:

$$\mathcal{V}(\mathbf{r}) = -\mathbf{r} \cdot \nabla \cdot \vec{\vec{\sigma}}(\mathbf{r}) + \nabla \cdot (\mathbf{r} \cdot \vec{\vec{\sigma}}(\mathbf{r})) = \text{Tr} \vec{\vec{\sigma}}(\mathbf{r}) \quad (10.8)$$

Integration of Equation 10.7 over an atom yields the atomic virial theorem, Equation 10.3.

The virial field $\mathcal{V}(\mathbf{r})$ is a dressed density distribution of particular importance [26]. It describes the energy of interaction of an electron at some position $\mathbf{r}$ with all of the other particles in the system, averaged over the motions of the remaining electrons. When integrated over all space it yields the total potential energy of the molecule, including the nuclear energy of repulsion and for a system in electrostatic equilibrium, with $\mathcal{V} = V$, it equals twice the molecule's total energy. The virial field condenses all of the electron–electron, electron–nuclear and nuclear–nuclear interactions described by the many-particle wavefunction into an energy density that is distributed in real space. The electronic energy density is defined as:

$$E_e(\mathbf{r}) = \mathbf{G}(\mathbf{r}) + \mathcal{V}(\mathbf{r}) = -K(\mathbf{r}) \quad (10.9)$$

The electronic energy $E_e(\Omega)$ equals the total energy $E(\Omega)$ in Equation 10.6 in the absence of external forces with $W(\Omega) = 0$.

10.3.2

Short-Range Nature of the Virial Field and Transferability

An atomic self-consistent potential contains the long-range e–n and e–e Coulombic interactions. In describing the energy changes that arise when atoms approach one another, a new interaction is introduced, the repulsion between the nuclei. Because the interatomic repulsions between the electrons (V_{ee}) and between the nuclei (V_{nn}) are both approximately one-half the magnitude of the inter-atomic e–n attractive interaction (V_{en}), the resulting difference between the repulsive and attractive interactions yields the relatively small change in energy accompanying the interactions between atoms [27]. Thus, it is necessary to include the nuclear–nuclear contribution to the energy changes resulting from atomic interactions or from the relative vibrational displacements of the nuclei, to obtain the net field, the field that is short-ranged compared to that determined by just the e–n and e–e interactions.

The virial field $\mathcal{V}(\mathbf{r})$, because it includes all contributions to the potential exerted at a point in space, is the most short-range possible description of the potential interactions in a many-electron system.

It is the near balance in these attractive and repulsive contributions making up the virial field that leads to the transferability of an atom's charge distribution and its properties and to the concept of a functional group as the carrier of chemical information.

The transferability of the electron density distributions and properties of functional groups, particularly those comprising the building blocks of biological molecules, is well documented [11, 28]. A theoretical study has tabulated the transferability of the atoms comprising all of the genetically encoded amino acids [29]. The standard deviations in the energies of the three groups comprising the main chain group $-[\text{C}_\alpha\text{H}(\text{NH}_2)\text{COOH}]$ common to all amino acids and comprising the back bone of a protein are 0.05% or less, being 0.04% for the $-[\text{C}_\alpha\text{H}]$ group bonded to the side chains making up the 24 different residues. Luger and Dittrich have provided

a detailed study of the comparison of the theoretically determined atomic and bond critical point properties of the amino acids with those obtained by experimental X-ray diffraction studies [30]. Their overall conclusion from the study of the peptide bond is that very reproducible atomic properties for contributing atoms can be determined if the chemical environment in the crystal is comparable.

Of equal importance is the demonstration that the paralleling degree of transferability of the virial field is found despite gross changes to the individual changes in the potential energy contributions to that field [6]. This is the crucial observation in the role of a functional group as the carrier of chemical information and accounts for the persistence of the properties of a group despite changes in its bonded neighbors. The property was first detailed in the second paper that dealt with the properties of the virial field using the example of the similarity of the H atom densities in BeH and BeH₂, demonstrating that the kinetic energy $T(H)$, the atomic virial $V(H)$ and hence the energy $E(H)$ of a hydrogen atom in BeH changed by less than 10 kcal mol⁻¹ when Be|H was transformed into H|Be|H, despite large but compensating changes in the individual external contributions to its potential energy [31]. The large changes in the separate contributions of V_{en} , V_{ee} and V_{nn} to the virials of the transferable methyl and methylene groups of the linear hydrocarbons have been previously detailed [2]. New examples of the paralleling behavior of the virial field and the density that occur despite gross changes in the external potential are provided here from the study of the energy changes incurred on the formation of DNA base pairs.

10.4

Changes in Atomic Energies Encountered in DNA Base Pairing

There have been numerous computational studies of DNA base pairing, many concentrating on the role of hydrogen bonding in the pairing process [16]. Reviews of previous work are provided by Popelier and Joubert [4] and by Parthasarathi *et al.* [32]. Particular attention has been paid to the relation between the energy of base pairing and the number of hydrogen bonds. However, as pointed out by Popelier and Joubert, simply counting hydrogen bonds in a complex or folding pattern does not necessarily provide insight into stability, a point made earlier by Gellman and coworkers [33, 34]. Jorgensen and coworkers introduced the so-called secondary interaction hypothesis (SIH) in an attempt to better reconcile the base pairing strength with the hydrogen bonded structure [35–37]. SIH addresses this problem by invoking short-range “cross interactions” between the frontier atoms of each base pair.

Electrostatic models are frequently employed in the study of interactions encountered in biological molecules. Gadre and Pudlik have reviewed their application to DNA base pairing in a paper where they introduce the electrostatic potential for intermolecular complexation through a mapping of the topography of the molecular electrostatic potentials [38]. Their method uses a point charge model with the charges chosen to fit the electrostatic potential. Kosov and Popelier have demonstrated that when the molecular electrostatic potential is expressed in terms of atomic contributions it may be rigorously expanded in terms of QTAIM electrostatic multipole

moments [39]. Popelier and Joubert [4] have carried the calculation of the electrostatic energy of interaction between 27 DNA base pairs to its practical limit by using the program ORIENT to calculate the intermolecular electrostatic energy using all multipole–multipole interactions up to R^{-6} , with the atomic multipole moments calculated from QTAIM. They have demonstrated the ability of their electrostatic approach, when augmented with simple repulsion terms, to recover the energies and geometries of the DNA base-pairs, through a comparison of the predictions with the supermolecule calculations at the MP2 and B3LYP levels of theory [40]. They conclude that the electrostatic description dominates DNA base-pair energies and geometries. In a companion study [4], they reach the important conclusion that – contrary to the widespread belief in the predominance of the importance of hydrogen bonding and the SIH [35] – the interaction energy is found to involve important contributions from long-range electrostatic interactions. They find that the atomic partitioning of the electrostatic interaction energy contains many substantial contributions between distant atoms, an observation coupled with the finding that base pairs with similar interaction energies are not stable for the same reasons in terms of the atomic partitioning of the electrostatic energy.

Parthasarathi *et al.* [32] have analyzed the interactions between base pairs using the atomic and bond path properties obtained from QTAIM together with an analysis based on the DFT derived electronegativity and hardness descriptors introduced by Parr and Pearson [41–43]. They give the molecular graphs calculated using the single point MP2/6-31G*(0.25) level of calculation for 28 base pairs. The molecular graphs, in addition to recovering all of the anticipated chemical structures of the individual bases, determine all of the bonded intermolecular interactions resulting from the pairing of the bases. These intermolecular interactions include the anticipated NH|B (acid|base) hydrogen bonded interactions with the acid N–H an amino NH_2 group, or a ring amino NH group, and B an imino N or keto oxygen atom. There are several instances in which a C–H group serves as the acid with a keto O or imino N serving as the base. All of the base pairs possess two N–H|B interactions with B = imino N or keto O atoms. GCWC (WC = Watson–Crick) possesses two N–H|O interactions in addition to the single N–H|N interaction. ATWC, in addition to the single N–H|N interaction, possesses a N–H|O interaction and a CH|O interaction.

Parthasarathi *et al.* [32] choose to describe the hydrogen bonded interactions with C–H| acting as the acid as “secondary” although they differ from the N–H|B hydrogen bonded interactions only in strength, not in any of the properties that characterize hydrogen bonding in QTAIM. They also find two weak inter-pair bonded interactions between the oxygen atoms of T and C in TC1 and TC2. The structure of TC2 is recovered in the present calculations in the molecular graphs shown in Figure 10.2 for the 23 base pairs reported here.

Parthasarathi *et al.* [32] determined the atomic and bonded properties of the hydrogen atoms involved in hydrogen bonding. The QTAIM characterization of hydrogen bonding is well established and has been amply illustrated [44, 45]. The principal properties used in the characterization of hydrogen bonding are given in terms of the properties of the hydrogen atom and its bond path. Hydrogen bonding causes a transfer of electron density from the basin of the H atom primarily to A of

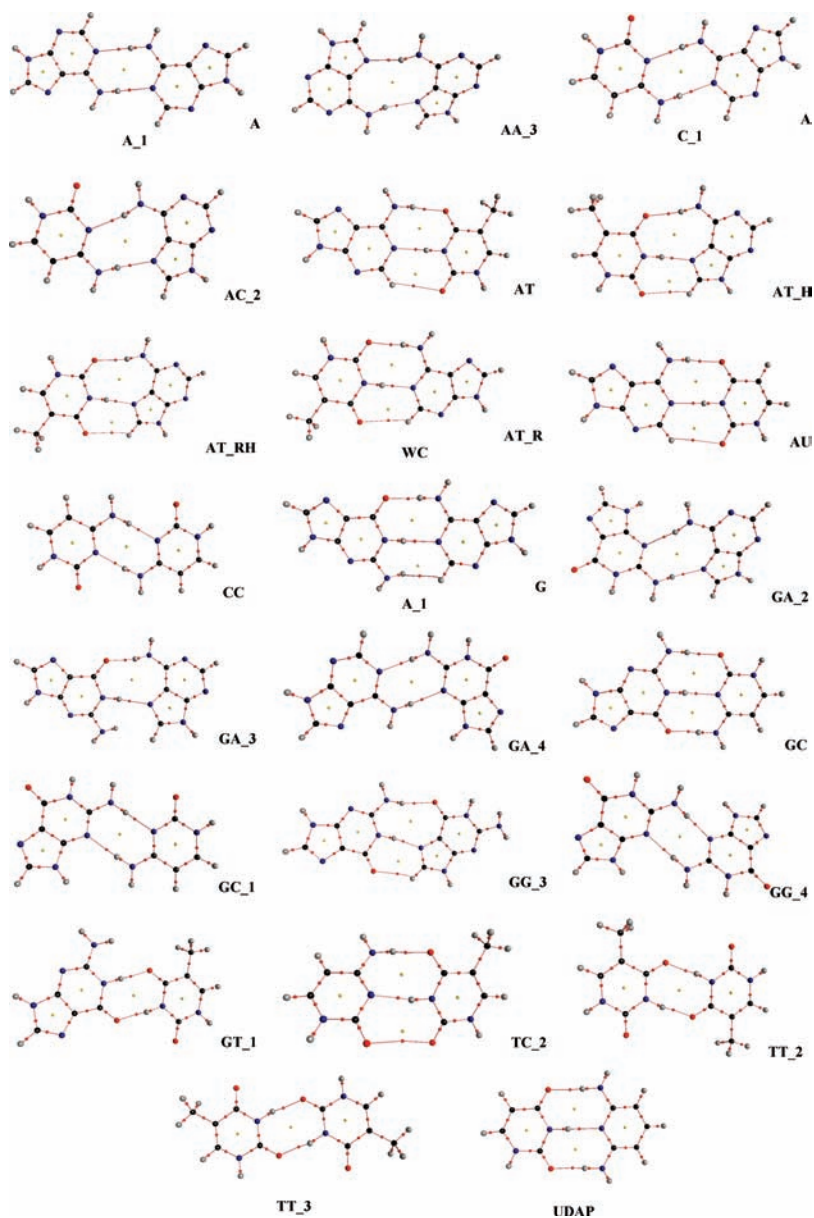


Figure 10.2 Molecular graphs for 23 of the most commonly considered DNA base pairs identified by their monomeric members. The configurations are identified by the abbreviations with WC denoting Watson–Crick

and H and RH denoting Hoogsteen and its reverse. The atom colors are H (white), C (black), N (blue) and O (red); bond critical points are denoted by red dots and ring critical points by yellow dots.

HA, the amino N atom in the case of the base pairs. The imino N and keto O atoms serving as the base atoms B receive smaller amounts of electronic charge. Because of the loss of density, there is an accompanying decrease in the stability of H. The interaction is further characterized by the mutual penetration of the outer densities of the H and B atoms to yield bonded radii less than their van der Waals nonbonded radii, which is determined by the 0.001 au density envelope, the strength of the interaction being paralleled by the degree of penetration of the van der Waals envelopes [44]. Thus the volume of H is significantly decreased. The penetration of the density of the H atom also results in a decrease in its dipolar polarization. Parthasarathi *et al.* recover these characteristics in the properties of the hydrogen bonded H atoms in the base pairs. They plot the change in the electron populations and the energies of the hydrogen atoms involved in hydrogen bond formation in a given base pair versus the total interaction energy. They find a scattering of points along a diagonal, indicating that the total interaction energy becomes increasingly negative as the charge and energy loss of the hydrogen atoms increase. They conclude that in addition to the electrostatic interaction, which appears to be the dominant contribution to hydrogen bonding at the HF level, other interactions are responsible for the stabilization of DNA base pairs.

QTAIM enables the determination of all atomic properties and their change. It is important in the investigation of the stability of DNA base pairing to consider the contributions from all atoms and not just those involved in the formation of inter-pair interactions. To assess the importance of “long-range” interactions, we ultimately view the energy changes in terms of the changes in the virial field, to better understand how the final energy changes are determined in terms of the separate internal and external contributions to the atom’s virial. It is clear from Equation 10.6 that for a system in electrostatic equilibrium $W(\Omega) = 0$, and hence $V(\Omega) = 2E(\Omega)$.

10.4.1

Dimerization of the Four Bases A, C, G and T

We begin with a study of the dimerization of each of the four bases, specifically AA1, CC, GG4 and TT2. All of the base dimers possess two N–H|B interactions: N–H denoting either one H atom of NH₂ the N being denoted by N(2), or the single H bonded to an amino N in a ring structure, the N being denoted by N(1); B is either an imino N or a keto oxygen atom. The first three have N(2)–H| as the acid and an imino N atom as the base. TT2 employs N(1)–H| as the acid and a keto oxygen as the base. Figure 10.3 shows the molecular graphs and atomic numberings. The major atomic contributions to the pairing of each base are made clear in the bar graphs giving the change in $N(\Omega)$ and $E(\Omega)$ for every atom in the dimer (Figure 10.4). These are seen to be, primarily, the atoms involved in hydrogen bonding.

Table 10.1 lists the calculated energy changes (ΔE_T in kcal mol^{–1}) for dimer formation. The stability incurred by the presence of the amino N(2)–H|N bond paths in AA1 is, computationally, the same as that from the presence of the N(1)–H|O bond paths in TT2. Table 10.1 gives the changes in the energies of the three atoms directly involved in hydrogen bonding; the column headed ΔE_H gives twice their sum

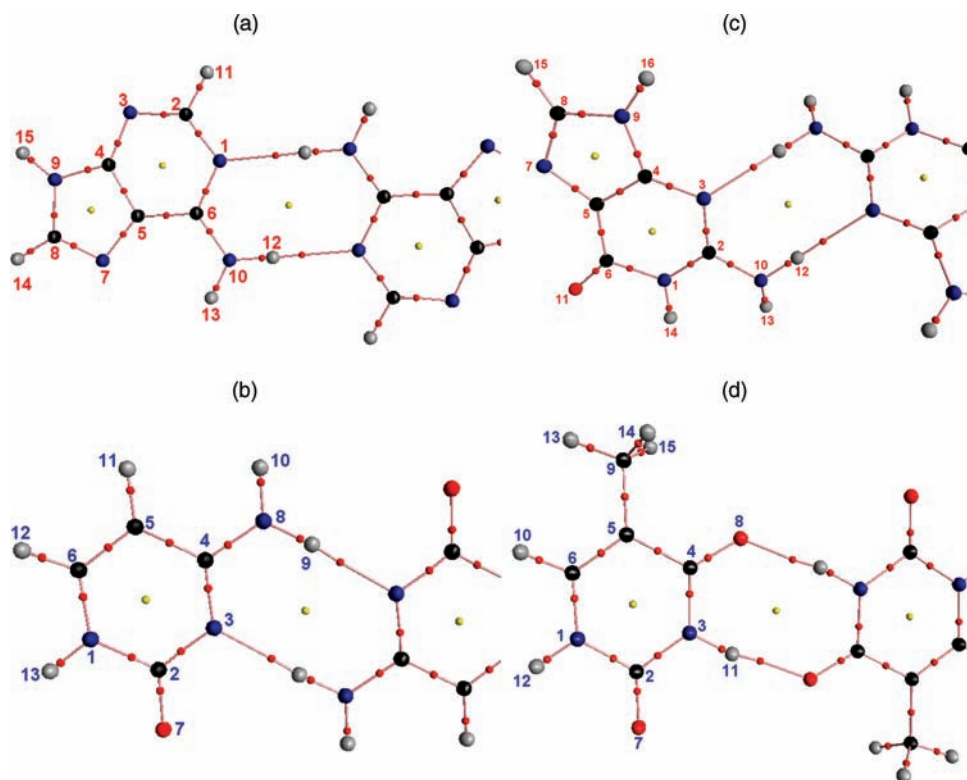


Figure 10.3 Molecular graphs and atomic numbering schemes for the DNA base dimers AA1 (a), CC (b), GG4 (c) and TT2 (d) (see Figure 10.2 for color scheme).

for the formation of the dimer. In the first three cases, those in which nitrogen is the base atom, the hydrogen bonding contributions exceed the total change in energy. The keto O hydrogen bonding results in a 20 kcal mol^{-1} increase in the total energy, primarily as a result of N1 being considerably less stabilized than N2 in acting as the base atom coupled with an only slight stabilization of the keto O as base.

The hydrogen bonding occurring in DNA base-pairing yields an eight-membered ring structure, with the ring apices consisting of ring C atoms. The energies of these ring carbon atoms, listed under E_C in Table 10.1, increase in the formation of each dimer, being most destabilized in CC and AA1 and least in TT2. These energy increases parallel their loss of electron density, $0.2 e$ for AA1 and CC, decreasing to $0.007 e$ and $0.003 e$ for GG4 and TT2, respectively. Addition of the changes in the apical carbon energies to those for the six hydrogen bonded atoms gives the energy change incurred by the formation of the hydrogen bonded ring of atoms (ΔE_r , Table 10.1). This addition decreases the stability resulting from hydrogen bond formation in the first three members, yielding a net destabilizing contribution in the

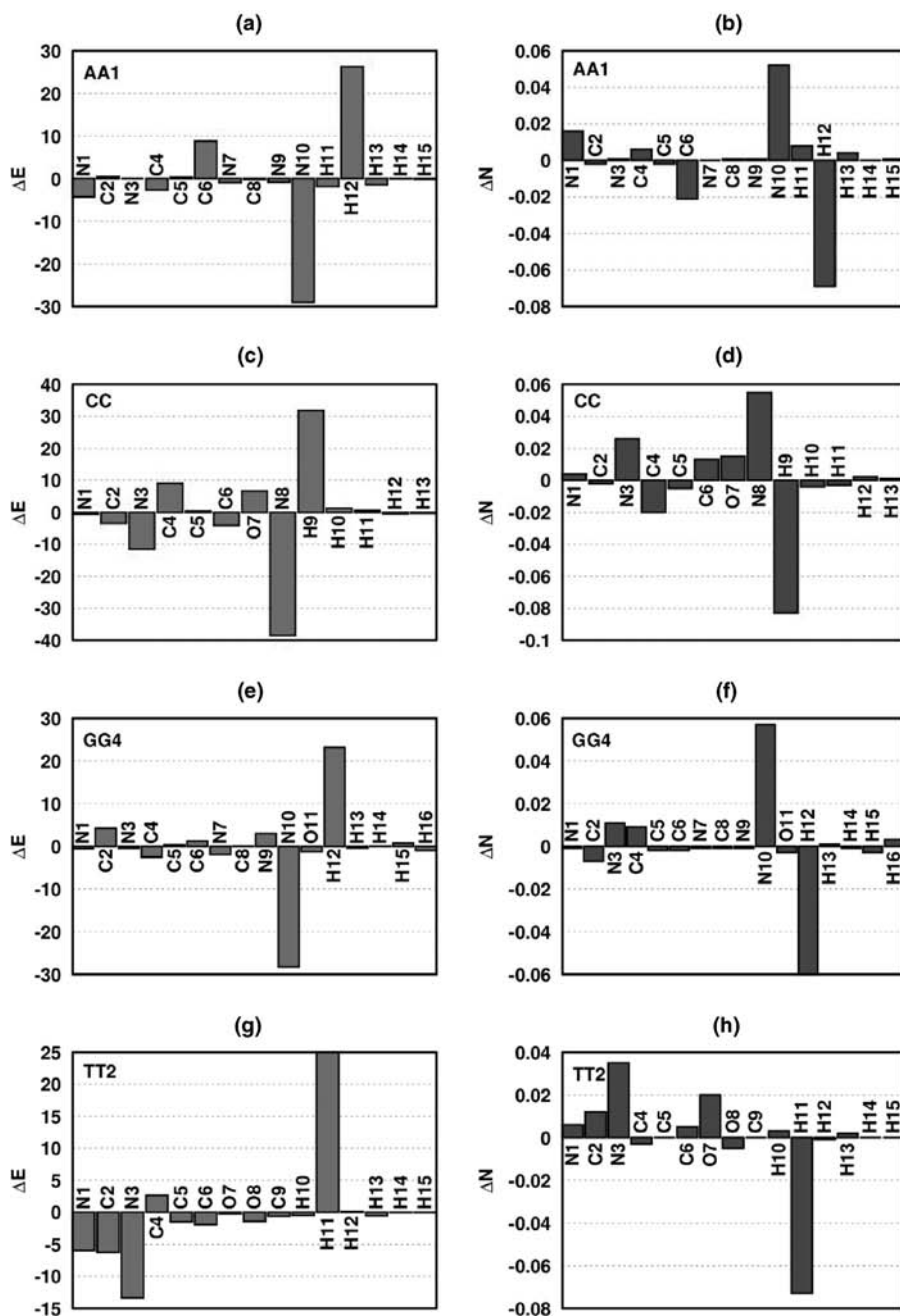


Figure 10.4 Bar graphs of the changes in atomic energies [(a), (c), (e) and (g)] and atomic populations [(b), (d), (f) and (h)] for the DNA base dimers.

Table 10.1 Atomic contributions to base pairing in dimer formation (energies in kcal mol⁻¹).

Base pair	$\Delta E(\Omega)$ acid		$\Delta E(\Omega)$ base N(O)	ΔE_H	ΔE_T	ΔE_C	ΔE_r	ΔE_d
	N	H						
AA1	-29.0	+26.3	-4.3	-14.0	-10.3	+8.8	+3.6	-13.9
CC	-38.5	+31.8	-11.6	-36.6	-17.5	+9.1	-18.4	+0.9
GG4	-28.3	+23.2	-0.5	-11.2	-7.8	+4.2	-2.8	-5.0
TT2	-13.4	+25.0	-1.4(O)	+20.4	-10.2	+2.7	+25.8	-36.0

case of AA1. For TT2, where the hydrogen bonding itself is destabilizing, the addition of the C atom energy changes increases the instability to 26 kcal mol⁻¹.

The missing energy—required to account for the energy of formation ΔE_T —must be found in the energy changes of the remaining atoms. This “energy deficit,” denoted ΔE_d ($=E_T - E_r$) in Table 10.1, is near zero for CC, the dimer with the strongest hydrogen bonding, ΔE_H , and largest for TT2, the only dimer with destabilizing hydrogen bonding. The energy deficit is thus a measure of the spreading of the perturbative effects of the formation of the hydrogen bonded rings into the remaining atoms of the system.

10.4.2

Energy Changes in CC

The dimer CC has the most stabilizing hydrogen bonding and the smallest energy deficit, ΔE_d . There are, however, non-negligible but compensating energy changes (kcal mol⁻¹) from three atoms within the cytosine ring: +6.7 for keto O7, -3.5 for its bonded carbon C2 and -4.1 for carbon C6 once removed. The remaining difference is made up by the other atoms, all of which, with the exception of H13 with $\Delta E(H) = 1.3$ kcal mol⁻¹, contribute less than |1.0| kcal mol⁻¹. Thus, in this case of strong hydrogen bonding, there are significant but canceling contributions from three atoms of the cytosine ring.

10.4.3

Energy Changes in AA1

The dimer AA1 has the next most stabilizing hydrogen bond formation, but one that is less than half of that for CC and it exhibits a significant energy deficit of -14 kcal mol⁻¹. Examination of the bar graph of energy changes shows that only three atoms make significant contributions to ΔE_d : the carbon atom C4 common to both rings of adenine and the second of the two hydrogen atoms, H13, bonded to an amino N(2) and a ring hydrogen H11. All three are stabilized and account for -12 of the -14 kcal mol⁻¹ deficit. The remaining ring atoms, each with contributions of less than |1| kcal mol⁻¹ to ΔE_d , make up the remaining 2 kcal mol⁻¹ difference.

10.4.4

Energy Changes in GG4

The dimer GG4 has a hydrogen bonding energy 3 kcal mol^{-1} less than AA1 but a considerably smaller energy deficit equal to 5 kcal mol^{-1} . Atoms N7 and N9 of the five-membered ring are stabilized and destabilized by 2 and 3 kcal mol^{-1} , respectively, and their contributions together with the stabilization of C4 linking the two rings contribute -3 kcal mol^{-1} to ΔE_d . The remaining atoms contribute energy changes of $\sim |1| \text{ kcal mol}^{-1}$ or less to attain the reported ΔE_d . As in the above two examples, three atoms other than those involved in the formation of the hydrogen bonded ring contribute significant amounts to the energy of formation of the base pair.

10.4.5

Energy Changes in TT2

The hydrogen bonding resulting from N(1)–H as acid and keto O as base, the interaction N(1)–H|O, is not only less effective than that from N(2)–H|N but is in fact destabilizing to the extent of 10 kcal mol^{-1} . The bar graph for TT2 in Figure 10.4 makes clear that while the energy changes for H and N1 dominate this interaction, the destabilizing increase in $\Delta E(\text{H})$ prevails. The keto O serving as the base actually undergoes a small loss of density and, correspondingly, only a small stabilization. The energy deficit arises from the transfer of significant density, $0.01 e$, to both C2, the carbon linked to the base atom N(1)3, and to its bonded neighbor N1. Each of these atoms undergoes an energy decrease of 6 kcal mol^{-1} , contributing -24 of the missing $-36 \text{ kcal mol}^{-1}$ to ΔE_d . The remaining stabilization comes from C5 and C6 of the thymine ring, contributing -1.5 and $-1.9 \text{ kcal mol}^{-1}$, respectively. The remaining atoms all exhibit energy changes of less than 1 kcal mol^{-1} . Clearly, the 50 kcal mol^{-1} increase in the energy of the hydrogen bonded H atoms dominates the interaction and the stabilizing contributions come from the atoms of the thymine ring, as well as the acid atom N1 of N(1)–H.

10.5

Energy Changes in the WC Pairs GC and AT

The analysis of the stability (ΔE_T) of these two primary base pairs follows that employed in the analysis of the dimers: the determination of the energy of hydrogen bonding ΔE_H , given in these cases by the separate sums of the three atomic energy changes $\Delta E(\text{A})$, $\Delta E(\text{H})$ and $\Delta E(\text{B})$; ΔE_r = the addition of the energies of the apical C atoms forming the hydrogen bonded rings and ΔE_d , the contribution to the energy of formation ΔE_T , from the remaining atoms. These energy changes are given in Table 10.2 for GC and in Table 10.3 for AT. The total interaction energy ΔE_T is given in the table for each base pair along with the contribution from each base. Figure 10.5 shows the molecular graphs and Figure 10.6 gives the bar graphs showing the changes in the atomic populations and energies.

Table 10.2 Atomic contributions to base pairing in GC (energies in kcal mol⁻¹);^{a)} $\Delta E_T = -24.7$, $\Delta E_T(G) = -18.9$ and $\Delta E_T(C) = -5.8$ kcal mol⁻¹.

$\Delta E(\Omega)$ acid		$\Delta E(\Omega)$ base	ΔE_H	ΔE_C	ΔE_T	ΔE_d
N(2)	H13(G)	O				
-47.5	+ 33.1	+ 6.8	-7.6	+ 1.0		
N(1)	H14(G)	N				
-29.3	+ 24.8	-15.8	-20.3		-27.3	+ 2.7
N(2)	H9(C)	O				
-39.3	+ 37.3	+ 6.3	+ 4.3	-4.7		
			$\Sigma \Delta E_H = -23.6$			

a) Hydrogen bonded interaction is identified by the number assigned to the H atom.

There is a transfer of $0.035 e$ from C to G in forming the GC pair and the energy of G is stabilized by -18.9 kcal mol⁻¹ compared to the -5.8 kcal mol⁻¹ stabilization for C. A similar amount of charge, $0.037 e$, is transferred from A into T in the formation of AT, causing a 0.8 kcal mol⁻¹ increase in the energy of A and a decrease of 13.0 kcal mol⁻¹ in T. The strongest of all six H-bonded interactions is the central N(1)–H|N interaction in GC, the weakest is C–H|O in AT. As anticipated, C–H is the weakest acid, the charge on the H atom in unbound AH being $+0.04 e$ compared to those on the unbound amino hydrogens, whose charges are ten times more positive. The values of ΔE_H do not correlate with the extent to which the bonded H atom is destabilized, this value being greatest for N(2)–H–O in GC, which yields a destabilizing increase in ΔE_H of $+4$ kcal mol⁻¹. ΔE_H does correlate with the stabilization of the acid N(1) and the most stabilized N(2) gives the most negative ΔE_H . Clearly, a keto O serving as base is much less stabilizing than an imino N atom, the energy change for O being destabilizing in every case, with its instability increasing with increasing stability of the N atom of the acid. These results for ΔE_H parallel those found for the dimers – the strongest hydrogen bonding arises from the

Table 10.3 Atomic contributions to base pairing in AT (energies in kcal mol⁻¹);^{a)} $\Delta E_T = -12.2$, $\Delta E_T(A) = +0.8$ and $\Delta E_T(T) = -13.0$ kcal mol⁻¹.

$\Delta E(\Omega)$ acid		$\Delta E(\Omega)$ Base	ΔE_H	ΔE_C	ΔE_T	ΔE_d
N(2)	H12(A)	O				
-25.9	+ 24.4	+ 0.7	-0.8	+ 1.2		
N(1)	H11(T)	N				
-13.7	+ 29.8	-24.7	-8.6		-4.9	-7.3
C	H11(A)	O				
-0.9	+ 7.1	+ 0.1	+ 6.3	-3.0		
			$\Sigma \Delta E_H = -3.1$			

a) Hydrogen bonded interaction is identified by the number assigned to the H atom.

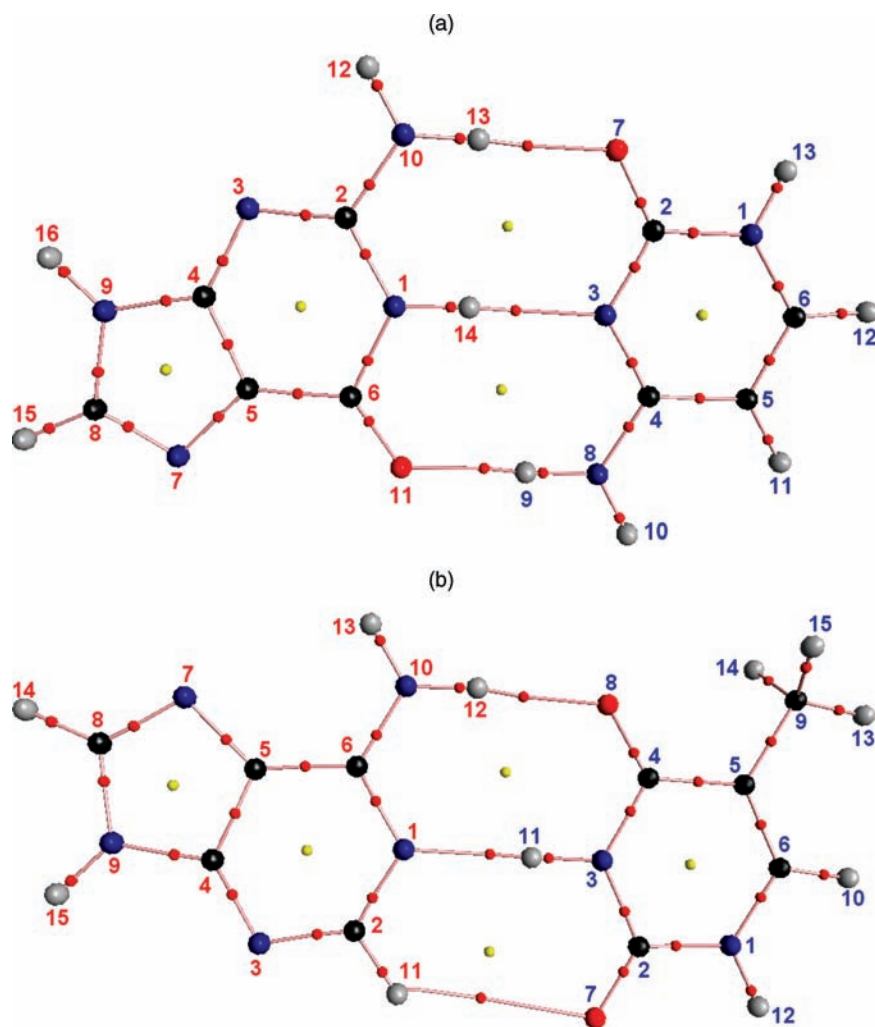


Figure 10.5 Molecular graphs and atomic numbering for the WC base pairs GC (a) and AT (b) (see Figures 10.2 for color scheme).

N(2)–H|N interaction in CC and the weakest and destabilizing interaction from N(1)–H|O with the keto O as base in TT2. The energy of hydrogen bond formation is most stabilizing in the case of GC, where $\Delta E_H = -23.6 \text{ kcal mol}^{-1}$ compared to a stabilization of $-3.1 \text{ kcal mol}^{-1}$ for AT.

Grunenberg has calculated, at both DFT and MP2 levels of theory, the “bond strengths” of the inter-residue interactions in both base pairs using the method of compliance constraints [46]. These constraints are independent of the coordinate system and provide a measure of the displacement of an internal coordinate resulting from a unit force acting on it. He found the same two extremes in H-bonded interaction energies reported here, reproducing as well the considerable gap between

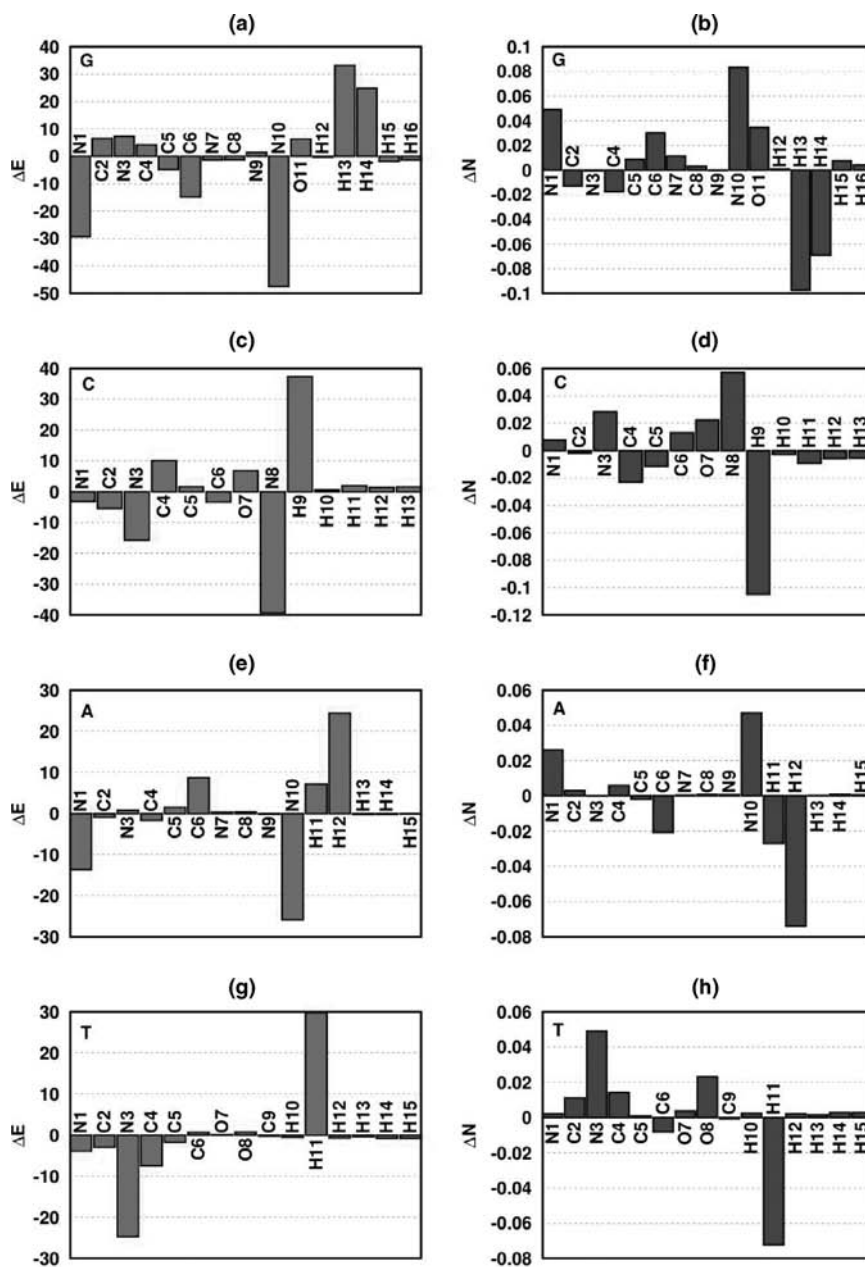


Figure 10.6 Bar graphs of the changes in atomic energies [(a), (c), (e) and (g)] and atomic populations [(b), (d), (f) and (h)] for the individual bases in the WC base pairs GC and AT.

the energy of the most stable $N(1)-H|N$ and the next most stable interaction. The ordering of the next two is interchanged from that found here, but at MP2 their constraint values are identical to within $0.01 \text{ Å mdyne}^{-1}$. The ordering of the electrostatic energies of the three hydrogen bonds found by Popelier and Joubert for GC are not in agreement with either of these findings [4]. They find the most stable bonding as given by $N(1)-H(14)|N$ in the QTAIM analysis to be the weakest and the weakest, that from $N(2)-H(9)|O$, to be the strongest.

Bader and Carroll [44] have demonstrated that the strength of hydrogen bonding, ΔE_H , parallels the extent of the mutual penetration of the van der Waals radii of the H and B atoms, where the van der Waals radii are identified with the nonbonded radii of the H and B, as defined by the 0.001 au density envelope. The hydrogen bonds of GC and AT provide further examples of the importance of the penetration effect in determining ΔE_H . The sum of the changes in the radii of the H and B atoms, given first in each case and denoted by Δr , are given for each H bonded interaction followed by the individual contributions from the H and B atoms. The results are listed in order of decreasing penetration, all in au. Considering the degree of penetrations for GC first, one has (i) $H14(G)$, $\Delta r = 2.62$; $\Delta r(H) = 1.29$, $\Delta r(N) = 1.33$; (ii) $H13(G)$, $\Delta r = 1.96$; $\Delta r(H) = 1.03$, $\Delta r(N) = 0.93$; (iii) $H9(C)$, $\Delta r = 1.34$; $\Delta r(H) = 1.13$, $\Delta r(O) = 0.21$. The decreasing extent of penetration follows the decrease in hydrogen bonding strength, with the largest gap between H14 and H13, as found for ΔE_H . For AT one finds: (i) $H11(T)$, $\Delta r = 2.42$; $\Delta r(H) = 1.33$, $\Delta r(N) = 1.09$; (ii) $H12(A)$, $\Delta r = 1.92$; $\Delta r(H) = 0.98$, $\Delta r(O) = 0.94$; (iii) $H11(A)$, $\Delta r = 0.78$; $\Delta r(H) = 0.27$, $\Delta r(O) = 0.51$. Here again, the decreasing extent of penetration parallels the decrease in ΔE_H . One notes the small penetration of the H atom bonded to C in the $C-H|O$ interaction. Clearly, $C-H$ is a "hard" acid.

Addition of the energies of the apical carbon atoms yields the energies of the hydrogen bonded ring formation, ΔE_r . Unlike the case of the dimers where the apical atoms are destabilizing, one finds one carbon stabilizing and the other destabilizing in the mixed base pairs. Considering GC first (Table 10.2), one finds the $C(2)$ atom of G is destabilizing by $6.5 \text{ kcal mol}^{-1}$, while $C(2)$ of C is stabilizing by $-5.5 \text{ kcal mol}^{-1}$ to give an overall deficit of $+1.0 \text{ kcal mol}^{-1}$. In the lower ring, $C(6)$ of G is stabilized by $-14.8 \text{ kcal mol}^{-1}$, while $C(4)$ is destabilized by $10.1 \text{ kcal mol}^{-1}$. The four apical atoms thus contribute $-3.7 \text{ kcal mol}^{-1}$ to the energy of formation of GC and when added to the energy of formation of the hydrogen bonding, $\Delta E_H = 23.6 \text{ kcal mol}^{-1}$, yield an energy of ring formation of $\Delta E_r = -27.3 \text{ kcal mol}^{-1}$ and an energy deficit of $+2.7 \text{ kcal mol}^{-1}$.

The bar graph of the energy changes for G in CG, Figure 10.6, shows that atoms N3, C4 and C5 contribute amounts in excess of $\pm 4 \text{ kcal mol}^{-1}$ to the energy deficit, with the energies of the remaining atoms of the two ring systems of guanine decreasing by less than 2 kcal mol^{-1} . The remaining three heavy atoms of cytosine, with energies ranging from -3.4 to $+1.6 \text{ kcal mol}^{-1}$, together with the three hydrogen atoms contribute $-0.1 \text{ kcal mol}^{-1}$ to ΔE_d . This gives $\Delta E_d = +2.3 \text{ kcal mol}^{-1}$ (compared to $+2.7 \text{ kcal mol}^{-1}$ calculated from data in table).

For AT (Table 10.3), the energy of $C(6)$ in A increases by $8.7 \text{ kcal mol}^{-1}$ while that of $C(4)$ in T decreases by $7.5 \text{ kcal mol}^{-1}$. The second ring of AT consists of only seven

atoms and the contribution of C(2) of T to ring formation is $-3.0 \text{ kcal mol}^{-1}$ to yield a total contribution from the carbons of $-1.8 \text{ kcal mol}^{-1}$, which when added to the energy of hydrogen bonding, $\Delta E_{\text{H}} = -3.1 \text{ kcal mol}^{-1}$, yields an energy of ring formation $\Delta E_{\text{r}} = -4.9 \text{ kcal mol}^{-1}$ and an energy deficit of $-7.3 \text{ kcal mol}^{-1}$. The remaining atoms on A contribute $-0.3 \text{ kcal mol}^{-1}$, 1.5 from C5 and -1.7 from C4, with the other atoms all contributing less than $|1| \text{ kcal mol}^{-1}$. The only contributions from the remaining atoms on T are from N1, C5 and C6, which contribute $-5.0 \text{ kcal mol}^{-1}$; the hydrogen atoms, all with energy changes of less than $|1| \text{ kcal mol}^{-1}$, contribute the remaining $-2.0 \text{ kcal mol}^{-1}$.

10.6

Discussion

Clearly, from the bar graphs for changes in the atomic populations and energies, the principal contributions to the charge transfers and energy changes accompanying base pairing occur for the atoms directly involved in the formation of the hydrogen bonded rings, with the major contributions coming from the N–H acid group. In the cases of dimer formation, the hydrogen bonding energy (ΔE_{H}) is stabilizing, its magnitude exceeding the total energy of formation (ΔE_{T}) with the exception of TT2 where it is destabilizing. In the formation of the WC base pairs, $E_{\text{H}} \sim \Delta E_{\text{T}}$ for GC, while in the formation of AT the ΔE_{H} is a quarter of the value of ΔE_{T} . From these examples it is clear that hydrogen bonding with a keto oxygen as base is less stabilizing than that with an imino N and further that the C–H acidic group yields an interaction that is overall destabilizing. Thus the largest energy deficits (ΔE_{d}) – the greatest perturbations of atoms other than those involved in the formation of the hydrogen bonded rings – are from the molecules TT2 and AT, the base pairs that have the weakest hydrogen bonded interactions. In both these examples, the energy deficit is stabilizing and in the case of TT2 it exceeds three times the magnitude of ΔE_{T} .

While some of the atoms other than those involved in the formation of hydrogen bonding undergo significant energy changes, they are of small magnitude compared to the atomic contributions obtained by Popelier and Joubert [4] in their electrostatic analysis of the energy of base pairing express. Popelier *et al.* [47] have demonstrated that the atom–atom contributions to the electrostatic energy of interaction can be determined by means of a spherical tensor multipole expansion followed by six-dimensional integration over two atomic basins where the atomic moments are determined for the isolated base molecules. Their atom–atom electrostatic energy of interaction thus includes the repulsion between the two atomic nuclei, the attraction of each nucleus for the electron density of the other atom and the repulsion between their electron density distributions.

They report the base pair GC in detail, finding an electrostatic interaction energy of $-27 \text{ kcal mol}^{-1}$ and an average of 34 kcal mol^{-1} for the absolute value the atom–atom interactions. An electrostatic atom–atom interaction energy, computed from the moments of the unperturbed base molecules, is not comparable to the atomic energy changes reported here, as the latter represent the interaction of each atom with all of

the atoms in the complex. The atomic energies are found to be of much smaller magnitude than the atom–atom electrostatic energies; the average contribution of a single atom of C to ΔE_T being $-0.4 \text{ kcal mol}^{-1}$ and of G being $-1.2 \text{ kcal mol}^{-1}$. They find substantial energy changes to occur beyond 7 au, wherein a “cumulative energy profile” exhibits alternating regions dominated by either attractive or repulsive interactions, which can reach values of more than 16 times that of the total interaction energy. They state that these findings lend credence to “the strong long-range electrostatic interactions observed in condensed matter.” This conclusion is at variance with the atomic energy changes incurred on base pairing computed by the physics of an open system.

The following discussion makes clear that while individual changes in the attractive and repulsive contributions to the potential energy changes that determine an atom’s virial and energy can be substantial they largely cancel to yield changes that, with the exception of the primary atoms of hydrogen bonding (the atoms of the acidic N–H| group), are in general less than the overall change in the interaction energy.

10.6.1

Attractive and Repulsive Contributions to the Atomic Virial and its Short-Range Nature

The contributions to the virial are obtained as the expectation values of the corresponding electrostatic terms in the Hamiltonian: namely V_{en} the electron–nuclear attractive energy, V_{ee} the electron–electron repulsion energy and V_{nn} the nuclear–nuclear repulsion energy, the latter two being usefully grouped together to give V_r , the repulsive contribution to the potential energy, as determination of their separate contributions is computationally demanding. We consider the case of the atoms in GC in detail, as they are representative. Tables 10.4 and 10.5 give the changes in the energies and populations of each atom, $\Delta N(\Omega)$ and $\Delta E(\Omega)$, together with the change in the electron–nuclear potential energy $\Delta V_{en}(\Omega)$, which is the change in the value of the external potential brought on by hydrogen bonding, and the change in the repulsive contributions $\Delta V_r(\Omega)$ for all of the atoms in GC and AT, respectively. Also listed, is the separate change in the internal contribution to the electron–nuclear potential energy, denoted by $\Delta V_{en}^o(\Omega)$. This is the change in the attractive interaction of the nucleus of atom Ω with the electron density in the atomic basin of Ω , resulting from the loss or gain of electron density by atom Ω , a quantity that will clearly parallel the transferability of the atom’s density and hence its energy.

The atoms in each base are arranged in the approximate order of increasing magnitude in their energy change. The smallest energy changes in C are for the hydrogen atoms, those peripheral to the cytosine ring and the second H of the N(2) acid. All have $\Delta E(\Omega)$ lying between 0 and 2 kcal mol^{-1} . Since their energies are nearly conserved so are their electron densities and the changes in $\Delta V_{en}^o(\Omega)$ are correspondingly small, being slightly in excess of twice $\Delta E(\Omega)$. What is to be contrasted with these small resultant changes in atomic energies and virials are the very large separate and opposing contributions from $\Delta V_{en}(\Omega)$ and $\Delta V_r(\Omega)$, of the order of 2×10^3 to $3 \times 10^3 \text{ kcal mol}^{-1}$. Thus even in cases where the atomic densities and

Table 10.4 Changes in $E(\Omega)$, $N(\Omega)$, $V_{\text{en}}(\Omega)$ and $V_r(\Omega)$ on H bonding for GC (energies in kcal mol⁻¹).

GC	Ω	$\Delta E(\Omega)$	$\Delta V_{\text{en}}(\Omega)$	$\Delta V_r(\Omega)$	$\Delta V_{\text{en}}^{\circ}(\Omega)$	$\Delta N(\Omega)$
G	H12	-0.31	-2114.05	2113.44	-0.84	0.001
	H15	-1.87	-2013.33	2009.62	-3.96	0.008
	H16	-1.34	-1335.70	1333.04	-2.75	0.004
	C8	-1.28	-11 973.45	11 971.70	-5.80	0.003
	N7	-1.34	-21 172.80	21 171.31	-18.38	0.011
	N9	1.48	-19 487.23	19 491.37	5.95	0.000
	C4	4.21	-15 066.37	15 075.58	38.78	-0.017
	C5	-4.83	-18 136.23	18 127.40	-22.55	0.009
	O11	6.25	-41 295.43	41 309.52	-112.94	0.035
	C2	6.51	-18 456.05	18 469.85	31.24	-0.013
	N3	7.25	-24 903.77	24 919.42	16.79	-0.001
	C6	-14.77	-19 596.90	19 568.23	-90.95	0.030
	H14	24.79	-2694.57	2744.04	51.95	-0.069
	N1	-29.34	-38 630.40	38 573.04	-152.65	0.049
	H13	33.12	-2117.89	2183.98	68.91	-0.098
	N10	-47.49	-35 836.76	35 743.18	-255.35	0.083
C	H10	0.59	-2748.15	2749.32	1.32	-0.003
	H12	1.37	-2883.40	2886.12	2.93	-0.006
	H13	1.55	-2039.60	2042.68	3.23	-0.005
	H11	1.97	-3308.12	3312.05	4.40	-0.009
	C5	1.63	-23 232.31	23 235.13	18.11	-0.012
	N1	-3.19	-31 711.77	31 704.77	-31.05	0.008
	C6	-3.38	-20 099.31	20 092.12	-30.29	0.013
	C2	-5.46	-21 829.81	21 818.49	-15.41	-0.002
	O7	6.81	-48 398.04	48 410.76	-75.40	0.022
	C4	10.09	-24 456.61	24 476.31	55.22	-0.023
	N3	-15.75	-47 383.77	47 351.72	-124.89	0.028
	H9	37.34	-2611.18	2685.68	78.29	-0.105
	N8	-39.32	-43 955.09	43 875.99	-183.19	0.057

energies are only slightly perturbed the external potential undergoes large changes, indicating that the external potential is not a gauge of the chemical energy changes encountered in a chemical reaction. The same behavior is found for the cases of perfect or near perfect transferability, wherein the atomic density and energy changes, while vanishingly small, are accompanied by very large changes in the external potential. Thus the external potential is not the determining potential in chemistry as has been suggested by Prodan and Kohn [48]. All of the above H atoms undergo a loss of density on dimer formation, Table 10.4 and Figure 10.6, and hence the increase in $\Delta V_{\text{en}}^{\circ}(\Omega)$, while small, dominates their energy change. The corresponding hydrogens in G, H12, H16 and H15, Table 10.4, gain small amounts of charge and their energies and virial contributions change in the opposite direction with both $\Delta V_{\text{en}}^{\circ}(\Omega)$ and $\Delta E(\text{H}) < 0$. The nitrogen atom N7, of the five membered-ring of G, has only a small energy decrease despite a gain of 0.011 e , a consequence of a relatively large increase in the repulsive contribution.

Table 10.5 Changes in $E(\Omega)$, $N(\Omega)$, $V_{\text{en}}(\Omega)$ and $V_r(\Omega)$ on H bonding for AT (energies in kcal mol⁻¹).

AT	Ω	$\Delta E(\Omega)$	$\Delta V_{\text{en}}(\Omega)$	$\Delta V_r(\Omega)$	$\Delta V_{\text{en}}^{\circ}(\Omega)$	$\Delta N(\Omega)$
A	H15	-0.10	-1436.56	1436.35	-0.25	0.000
	H14	-0.20	-2160.58	2160.17	-0.47	0.001
	H13	-0.28	-2356.02	2355.46	-0.45	0.000
	N9	-0.19	-21 783.60	21 782.58	-2.24	0.001
	N7	0.29	-23 412.99	23 412.94	-1.53	0.000
	C8	0.33	-13 248.72	13 248.95	-1.14	0.001
	N3	0.80	-28 091.41	28 092.37	-1.81	0.000
	C2	-0.87	-22 661.30	22 659.14	-5.94	0.003
	C5	1.53	-19 757.05	19 759.67	5.40	-0.002
	C4	-1.71	-17 037.53	17 033.69	-14.82	0.006
	H11	7.14	-4318.72	4332.95	14.77	-0.027
	C6	8.68	-22 131.21	22 148.10	49.57	-0.021
	N1	-13.66	-42 971.03	42 943.14	-114.78	0.026
	H12	24.44	-2610.73	2659.50	52.24	-0.074
	N10	-25.87	-38 774.56	38 722.31	-142.96	0.047
T	H13	-0.41	-2688.04	2687.23	-0.84	0.002
	H10	-0.57	-2615.57	2614.44	-1.27	0.002
	H12	-0.64	-1803.12	1801.86	-1.35	0.002
	O7	0.14	-39 603.13	39 605.02	-9.75	0.004
	C9	-0.21	-17 883.82	17 884.21	2.49	-0.001
	C6	0.65	-17 642.25	17 644.36	17.08	-0.008
	O8	0.71	-44 675.75	44 678.79	-83.36	0.023
	H14	-0.73	-2999.45	2998.00	-1.52	0.003
	H15	-0.73	-2999.44	2998.00	-1.52	0.003
	C5	-1.76	-20 693.52	20 690.82	-3.76	0.001
	C2	-2.96	-18 361.50	18 356.39	-25.31	0.011
	N1	-3.87	-27 349.00	27 342.46	-6.43	0.002
	C4	-7.48	-20 884.99	20 870.86	-45.37	0.014
	N3	-24.70	-41 081.73	41 033.61	-132.85	0.049
	H11	29.77	-2710.80	2770.23	59.96	-0.072

The next set of atoms from cytosine is those that make the major contributions to ΔE_d , namely, N1, C5 and C6 of the cytosine ring. C5 loses density and has both $\Delta E(\text{C})$ and $\Delta V_{\text{en}}^{\circ}(\text{C}) > 0$, while N1 and C6 gain density and their energy changes are stabilizing. The changes in $\Delta V_{\text{en}}(\Omega)$ and $\Delta V_r(\Omega)$ are again in the range of 2×10^3 to 3×10^3 kcal mol⁻¹ but the change in the total virial reflects the small change in their energies and in atomic charge distributions. The corresponding atoms N3, C4 and C5 of the six-membered ring of G behave in a corresponding manner with $\Delta V_{\text{en}}^{\circ}(\Omega) > 0$ for N3 and C4 that lose density and $\Delta V_{\text{en}}^{\circ}(\Omega) < 0$ for C5 that gains density. The two carbons of C forming the apices of the hydrogen bonded rings, C2 and C4, gain and lose electron density, respectively, and correspondingly their energies and internal contributions to the e-n interaction energy are stabilizing and destabilizing. The corresponding atoms C2 and C6 of G mimic the atoms C4 and C2 of cytosine, respectively.

The largest changes in the attractive and repulsive contributions to the virial, of absolute order of 4×10^3 to 5×10^3 kcal mol⁻¹, are for the atoms directly involved in hydrogen bonding, N3, O7, N8 and H9 in C and H13, N10, N1, H14 and O11 in G. Their large values of $|\Delta V_{\text{en}}^{\text{o}}(\Omega)|$ indicate that the hydrogen bonding incurs substantial changes in the atoms' density distributions. The destabilization of the keto oxygens of C and G are of particular interest. These atoms gain electron density, $\Delta N(\text{O}) = 0.022$ and $0.035 e$, respectively, resulting in stabilizing decreases of -75 and -113 kcal mol⁻¹ in the attractive interaction of the O nucleus with its atomic density, but appeal to Table 10.4 shows that the repulsive contribution to their virials exceeds the total attractive one and the oxygens are destabilized despite the transfer of electronic charge to their basins. Thus the dominant increase in the repulsive over attractive interactions incurred when a keto oxygen serves as base accounts for the weak resultant hydrogen bonding. Interestingly, the same O7 of cytosine gains density and is destabilized to an almost equal extent in the dimer CC, where it does not participate in hydrogen bonding.

The behavior of a keto O as base is to be contrasted with that of an imino N. The transfer of $0.023 e$ to the imino N3 atom results in a substantial decrease in $V_{\text{ne}}^{\text{o}}(\text{N})$, equal to -125 kcal mol⁻¹, and while there is an accompanying increase in the repulsive contribution to the virial, the attractive interactions prevail and $\Delta E(\text{N}) = -16$ kcal mol⁻¹. By far the most stabilizing decrease in $\Delta V_{\text{en}}^{\text{o}}(\Omega)$ of the atoms in C, equal to -183 kcal mol⁻¹, is for the N8 atom of the N(2) base group, which gains $0.057 e$ on hydrogen bonding. The repulsive contribution to the virial also increases, reducing the final energy change -39 kcal mol⁻¹, a quantity of the same order of magnitude as ΔE_{T} . The corresponding N(2) atom in G behaves in a similar but more exaggerated manner, a consequence of a larger gain in electron density, with $\Delta N = 0.083 e$. It is the increase in electronic charge and subsequent stabilization of these two N atoms in the GC pair that give the only stabilizing contribution to the hydrogen bonded interactions with a keto O serving as the base atom. The amino N(1) atom of guanine behaves in a similar manner to the N(2) acidic N atoms, but undergoes a smaller increase in population, equal to $0.049 e$, and consequently a smaller stabilizing internal interaction energy, $\Delta V_{\text{en}}^{\text{o}}(\text{N})$. The resulting hydrogen bonding is stronger than with the N(2) acids because of an imino N rather than a keto O serving as the base.

The three hydrogen bonded H atoms all lose substantial electron density, with the N(2) hydrogen atoms losing approximately $0.03 e$ more than the hydrogen bonded to N(1). Paralleling the loss of density, all three undergo a destabilization in the internal e-n interaction, with $\Delta V_{\text{en}}^{\text{o}}(\text{H})$ increasing from 50 to 80 kcal mol⁻¹. This loss in density causes the increases in the repulsive contribution to the virials of the H atoms to exceed, in magnitude, the decreases in the attractive e-n interactions with the excesses in $\Delta V_{\text{r}}(\text{H})$ exceeding those for $\Delta V_{\text{en}}^{\text{o}}(\text{H})$ by amounts equal to twice the $\Delta E(\text{H})$. Thus, the hydrogen bonded H atoms are destabilized by both a loss of density and by an increase in the repulsive contributions to their virials.

The corresponding hydrogens in AT, Table 10.5, behave in a similar manner, as do all of the atoms detailed here, to GC. This is a reflection of the property of QTAIM atoms: they maximize transferability not only in their static properties but

also in the changes in these properties when undergoing corresponding chemical interactions.

In summary, each of the acidic atoms N(1), N(2) and the base atoms N, O, together with the hydrogen atoms directly involved in hydrogen bonding, behave in similar and understandable manners on base pairing. The dominant repulsive contribution to the atomic virial for a keto oxygen that occurs despite a gain in electron density accounts for its destabilizing contribution to the energy of hydrogen bonding. In a comparison of the WC base pairs, GC with the strongest hydrogen bonding is the more stable of the two. The situation, however, is not so straightforward for the dimer base-pairs. While the dimer CC is the most stable and has the most stable hydrogen bonding, TT2 with destabilizing hydrogen bonding has an overall stability the same as that for AA1, which has the second largest hydrogen bonding energy of the four dimers. This leaves GG4 as the least stable and also possessing the least stabilizing hydrogen bonding. The stability of TT2 that is achieved despite the destabilizing hydrogen bonded interaction is entirely a result of the stabilization of the atoms of the thymine ring, particularly atoms N1, an amino N(1) base and C2 the carbon of the keto group. These are the same atoms that contribute significantly to the energy deficit in AT. Thus, the overall stabilizing effect of the presence of adenine in base pairing comes from the same ring atoms in both TT2 and AT, stabilizations that compensate for the destabilizing use of the keto O in hydrogen bond formation. Thus QTAIM can identify the stabilizing and destabilizing interactions generic to a given base. In this regard we disagree with the conclusion of Popelier and Joubert [4] in their study of the “Elusive atomic rationale for DNA base pair stability. However, in general simple rules to rationalize the pattern of energetic stability across naturally occurring base pairs in terms of subsets of atoms remains elusive.” Our somewhat cursory investigation of the changes in the atomic energies and populations on base-pairing could be extended to include a tabulation of the changes in atomic properties for the 27 natural base pairs and include other properties, such as changes in the localization/delocalization index, a measure of the importance of the role of the exchange density in the base-pairing. It provides a measure of the extent of changes in the delocalization of the electrons over the unsaturated rings of the DNA bases incurred by pairing. With this catalogue of atomic contributions determined by the quantum mechanics of an open system, one could hope to obtain a predictive empirical approach to the understanding of the stability of base-pairing.

10.6.2

Can One Go Directly to the Virial Field?

The changes in the attractive and repulsive contributions to the energy of an atom encountered during chemical change largely cancel to yield the resultant virial that determines the atom's energy. This observation begs the question as to why one does not attempt to determine the change in the virial directly rather than proceeding through the laborious procedure of calculating individual contributions that largely cancel one another out. The virial field is homeomorphic with the electron density,

yielding the same structure diagram [49]. A plot of the virial field looks like a plot of the density, exhibiting the same critical points.

The virial field $\mathcal{V}(\mathbf{r})$ may be calculated through a calculation of the kinetic energy density [6]. Equation 10.8 relates $\mathcal{V}(\mathbf{r})$ to the trace of the stress tensor $\mathcal{V}(\mathbf{r}) = \text{Tr} \vec{\sigma}(\mathbf{r})$, and the trace may in turn be expressed in terms of the kinetic energy densities $K(\mathbf{r})$, the Schrödinger formulation and $G(\mathbf{r})$, its more useful positive definite form, Equation 10.10 [50]:

$$\text{Tr} \vec{\sigma}(\mathbf{r}) = -K(\mathbf{r}) - G(\mathbf{r}) = -2G(\mathbf{r}) - L(\mathbf{r}) \quad (10.10)$$

The quantity $L(\mathbf{r})$ is determined by the Laplacian of the electron density (Equation 10.11):

$$L(\mathbf{r}) = -(\hbar^2/4m)\nabla^2 \varrho(\mathbf{r}) \quad (10.11)$$

Thus the two quantities of interest are the Laplacian of the electron density $\varrho(\mathbf{r})$ and the positive definite form of the kinetic energy density $G(\mathbf{r})$:

$$G(\mathbf{r}) = (\hbar^2/2m)(\nabla \cdot \nabla')\Gamma^{(1)}(\mathbf{r}, \mathbf{r}')|_{\mathbf{r}=\mathbf{r}'} \quad (10.12)$$

both of which are determined by the one-electron density matrix.

In classical mechanics it is possible to state the principle of least action in terms of a variation of the kinetic energy T employing a generalized variation of the action integral denoted by the symbol Δ [51]. In this generalized procedure, the variations are not required to vanish as the time end points and there may be a variation in the coordinates at the time end points. This generalization of the variation principle is that employed by Schwinger in his development of the principle of stationary action in which the time and the state vector are both varied at the time end points [8]. In classical mechanics the generalized variation yields the following statement of the principle of least action:

$$\Delta \int_{t_1}^{t_2} p_i \dot{q}_i dt = 0 \quad (10.13)$$

If the generalized coordinates do not involve the time explicitly, the kinetic energy T is a quadratic function of the velocities $\dot{q}_i$ and, providing the potential energy is not velocity dependent, the principle of least action may be expressed in terms of the generalized variation of the kinetic energy T as given in Equation 10.14 [51]:

$$\Delta \int_{t_1}^{t_2} T dt = 0 \quad (10.14)$$

Just as Schwinger's generalized variation of the action could be extended to include a variation of the time-like (the evolving boundaries of an open system) as well as the space-like boundaries of a system [7, 52], so could the generalized expression given in Equation 10.14 be extended to the variation of an open system. It is not clear how one could implement the classical principle in quantum mechanics, but it is always useful to have knowledge of a possible classical analogue in searching for a new path in quantum mechanics.

Modeling of the changes in the density and kinetic energy as a function of perturbations is not unrelated – both being derivable from a modeling of the first-order density matrix. Tsirelson has reviewed recently the attempts to treat the use of the electron density in the modeling of energetic quantities from experimentally obtained densities, giving particular attention to the approximate expressions for the kinetic and potential energy densities [53]. Other explicit functionals for relating G to the density have been put forward by Ayers [54], Perdew and Constantin [55] and Garcia-Aldea and Avarelllos [56]. Formulations of implicit functionals relating G to the density derived within the Kohn–Sham approximation have been given by Wu and Yang [57] Yang, Ayers and Wu [58] and Colonna and Savin [59].

The restatement of the interaction of molecules in terms of the virial field may suggest that the approach will provide no replacement of the electrostatic and orbital models that have been used in the past to predict the course of a reaction, for example, the approach of a positively charged hydrogen atom towards the negatively charged base. This, however, is not the case. The Laplacian of the electron density has been shown to provide an operational model of the Lewis acid–base model of atomic interactions, with a local charge concentration on a base aligning with local charge depletion on the acid [60]. This description has been applied, for example, to hydrogen bonding, where it has been demonstrated that the geometries of the hydrogen bonded complexes of HF with various bases are in good agreement with the calculated angles [61]. The angle is predicted by aligning the $(3, +3)$ critical point found on the nonbonded side of the H atom of HF, denoting a local charge depletion, with the $(3, -3)$ critical point, denoting a local charge concentration, on the recipient atom of the base. Clearly, recalling the homeomorphic relation between the topology of the electron density and that of the virial field [49], one should investigate the Laplacian of the virial field and determine if it plays a complementary role in predicting the approach of reactant molecules to that provided by the Laplacian of the density. A $(3, -3)$ critical point in $\mathcal{V}(\mathbf{r})$ will denote a local region of maximally low potential energy density and a $(3, +3)$ critical point a local region of relatively high potential energy. Does the preferred course of a chemical reaction correspond to the mating of two such extrema in the Laplacian of $\mathcal{V}(\mathbf{r})$? A study is presently underway to address these ideas.

Notably, in giving a paramount role to the kinetic energy one must be prepared to counter the many false arguments that are prevalent in the existing literature regarding its role in chemistry. All involve the invoking of imaginary states that violate the theorems of quantum mechanics to achieve a desired result: that the kinetic energy decreases rather than increases on bonding and that the release of excess kinetic energy stabilizes a system when in fact, in the absence of external forces, $\Delta E = -\Delta T$.

Acknowledgment

We express our thanks to UNAM-DGSCA for their generous support in supplying the necessary computer time.

References

- 1 Hohenberg, P. and Kohn, W. (1964) *Phys. Rev. Sect. B*, **136**, 864.
- 2 Bader, R.F.W. (2008) *J. Phys. Chem. A*, **112**, 13717–13728.
- 3 Bader, R.F.W. and Beddall, P.M. (1972) *J. Chem. Phys.*, **56**, 3320–3329.
- 4 Popelier, P.L.A. and Joubert, L. (2002) *J. Am. Chem. Soc.*, **124**, 8725–8729.
- 5 Bader, R.F.W. (1994) *Phys. Rev.*, **B49**, 13348–13356.
- 6 Bader, R.F.W. (1990) *Atoms in Molecules: A Quantum Theory*, Oxford University Press, Oxford UK.
- 7 Bader, R.F.W. and Nguyen-Dang, T.T. (1981) *Adv. Quantum Chem.*, **14**, 63–124.
- 8 Schwinger, J. (1951) *Phys. Rev.*, **82**, 914–927.
- 9 Bader, R.F.W. (2007) *J. Phys. Chem. A*, **111**, 7966–7972.
- 10 Bader, R.F.W. and Matta, C.F. (2004) *J. Phys. Chem. A*, **108**, 8385–8394.
- 11 Matta, C.F. and Bader, R.F.W. (2006) *J. Phys. Chem. A*, **110**, 6365–6371.
- 12 Nalewajski, R.F. and Parr, R.G. (2000) *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 8879.
- 13 Gázquez, J.L., Cedillo, A., Gómez, B., and Vela, A. (2006) *J. Phys. Chem. A*, **110**, 4535–4537.
- 14 Hobza, P. and Sponer, J. (1999) *Chem. Rev.*, **99**, 247.
- 15 Frisch, M.J., Trucks, G.W., Schlegel, H.B. et al. (2004) Gaussian 03 Revision E.01, Gaussian, Inc., Wallingford CT.
- 16 Sponer, J., Leszczynski, J., and Hobza, P. (1996) *J. Phys. Chem.*, **100**, 1965–1974.
- 17 Keith, T. (2008) AIMALL.
- 18 Biegler-König, F.W., Schönbohm, J., and Bayles, D. (2001) *J. Comput. Chem.*, **22**, 545–559.
- 19 Bader, R.F.W. (1995) *Int. J. Quantum Chem.*, **56**, 409–419.
- 20 Bader, R.F.W. (1980) *J. Chem. Phys.*, **73**, 2871–2883.
- 21 Bader, R.F.W., Nguyen-Dang, T.T., and Tal, Y. (1981) *Rep. Prog. Phys.*, **44**, 893–948.
- 22 Fradera, X., Austen, M.A., and Bader, R.F.W. (1999) *J. Phys. Chem. A*, **103**, 304–314.
- 23 Bader, R.F.W. and Heard, G.L. (1999) *J. Chem. Phys.*, **111**, 8789–8798.
- 24 Keith, T.A. and Bader, R.F.W. (1993) *J. Chem. Phys.*, **99**, 3669–3682.
- 25 Bader, R.F.W. (1998) *Can. J. Chem.*, **76**, 973–988.
- 26 Bader, R.F.W. (2003) *The Fundamentals of Electron Density, Density Matrix and Density Functional Theory of Atoms, Molecules and the Solid State*, Kluwer Academic Publishers, Dordrecht, pp. 185–193.
- 27 Bader, R.F.W. (1970) *An Introduction to the Electronic Structure of Atoms and Molecules*, Clarke Irwin & Co Ltd, Toronto, Canada (available on line at: www.chemistry.mcmaster.ca/faculty/bader/aim/).
- 28 Bader, R.F.W., Matta, C.F., and Martin, F.J. (2003) Chapter 7, *Quantum Medicinal Chemistry* (eds P. Carloni and F. Alber), Wiley-VCH Verlag GmbH, Weinheim, pp. 201–231.
- 29 Matta, C.F. and Bader, R.F.W. (2003) *Proteins: Struct., Funct. Genet.*, **52**, 360–399.
- 30 Luger, P. and Dittrich, B. (2007) Chapter 12, in *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design* (eds C.F. Matta and R.J. Boyd), Wiley-VCH Verlag GmbH, Weinheim, pp. 317–339.
- 31 Bader, R.F.W., Beddall, P.M., and Peslak, J. (1973) *J. Chem. Phys.*, **58**, 557–566.
- 32 Parthasarathi, R., Amutha, R., Subramanian, V., Balachandran, U.N., and Ramasami, T. (2004) *J. Phys. Chem. A*, **108**, 3817–3828.
- 33 Gardner, R.R. and Gellman, S.H. (1995) *J. Am. Chem. Soc.*, **117**, 10411–10412.
- 34 Yang, J. and Gellman, S.H. (1998) *J. Am. Chem. Soc.*, **120**, 9090–9091.
- 35 Jorgensen, W.L. and Pranata, J. (1990) *J. Am. Chem. Soc.*, **112**, 2008–2010.
- 36 Jorgensen, W.L. and Severance, D.L. (1991) *J. Am. Chem. Soc.*, **113**, 209–216.
- 37 Pranata, J., Wiershke, S.G., and Jorgensen, W.L. (1991) *J. Am. Chem. Soc.*, **113**, 2810–2819.
- 38 Gadre, S.R. and Pudlik, S.S. (1997) *J. Phys. Chem. B*, **101**, 3298–3303.
- 39 Kosov, D.S. and Popelier, P.L.A. (2000) *J. Phys. Chem. A*, **104**, 7339–7345.
- 40 Joubert, L. and Popelier, P.L.A. (2002) *Phys. Chem. Chem. Phys.*, **4**, 4353.

- 41 Parr, R.G. and Pearson, R.G. (1983) *J. Am. Chem. Soc.*, **105**, 7512.
- 42 Pearson, R.G. (1987) *J. Chem. Educ.*, **64**, 561.
- 43 Parr, R.G. (1985) *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 6723.
- 44 Carroll, M.T. and Bader, R.F.W. (1988) *Mol. Phys.*, **65**, 695–722.
- 45 Koch, U. and Popelier, P.A.L. (1995) *J. Phys. Chem.*, **99**, 9747.
- 46 Grunenberg, J. (2004) *J. Am. Chem. Soc.*, **126**, 16310–16311.
- 47 Popelier, P.A.L., Joubert, L., and Kosov, D.S. (2001) *J. Phys. Chem. A*, **105**, 8254–8261.
- 48 Prodan, E. and Kohn, W. (2005) *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 11635–11638.
- 49 Keith, T.A., Bader, R.F.W., and Aray, Y. (1996) *Int. J. Quantum Chem.*, **57**, 183–198.
- 50 Bader, R.F.W. and Preston, H.J.T. (1969) *Int. J. Quantum Chem.*, **3**, 327–347.
- 51 Goldstein, H. (1965) *Classical Mechanics*, Addison-Wesley, Reading, MA.
- 52 Bader, R.F.W., Srebrenik, S., and Nguyen Dang, T.T. (1978) *J. Chem. Phys.*, **68**, 3680–3691.
- 53 Tsirelson, V.G. (2007) Chapter 10, in *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design* (eds C.F. Matta and R.J. Boyd), Wiley-VCH Verlag GmbH, Weinheim, pp. 259–283. The potential energy density defined in Equation (2) of this chapter is not the virial field as it is so identified in Equation (3). The expression in Equation (2) lacks the virial of the Feynman forces exerted on the density that determine the nuclear–nuclear repulsive contribution.
- 54 Ayers, P.W. (2005) *J. Chem. Sci.*, **117**, 441–454.
- 55 Perdew, J.P. and Constantin, L.A. (2007) *Phys. Rev. B*, **75**, 155109.
- 56 Garcia-Aldea, D. and Alvarellos, J.E. (2007) *J. Chem. Phys.*, **127**, 144109.
- 57 Wu, Q. and Yang, W.T. (2003) *J. Chem. Phys.*, **118**, 2498.
- 58 Yang, T., Ayers, P.W., and Wu, Q. (2004) *Phys. Rev Lett.*, **92**, 146404.
- 59 Colonna, F. and Savin, A. (1999) *J. Chem. Phys.*, **110**, 2828.
- 60 Bader, R.F.W., MacDougall, P.J., and Lau, C.D.H. (1984) *J. Am. Chem. Soc.*, **106**, 1594–1605.
- 61 Carroll, M.T., Chang, C., and Bader, R.F.W. (1988) *Mol. Phys.*, **63**, 387–405.

11

An Electron Density-Based Approach to the Origin of Stacking Interactions

Ricardo A. Mosquera, María J. González Moa, Laura Estévez, Marcos Mandado, and Ana M. Graña

11.1

Introduction

Stacking interactions, as well as other noncovalent interactions usually included within this term, are considered among the most important factors involved in chemical and biological recognition [1–3]. They are fundamental for the architecture and stabilization of DNA molecules, the crystal packing of aromatic molecules, the formation of the tertiary structure of proteins, the control in the enzyme–nucleic acids recognition regulating gene expression, intercalation of drugs into DNA, and so on. Therefore, considerable attention has been paid to π – π stacking and related interactions in the chemical literature. The most important findings hitherto obtained on this topic have been reviewed recently in a themed issue of *Physical Chemistry Chemical Physics* edited by Hobza [4].

It is generally accepted that stacking and hydrogen bonding play leading roles in determining the structure of biomacromolecules and supramolecular systems [2, 5]. Whereas the origin of hydrogen bonding has been analyzed in detail and extensively, the opposite is really true about the set of noncovalent interactions of aromatic, pseudoaromatic or conjugated subunits that are many times denoted as “stacking interactions” [4]. The combination of the computational levels required for obtaining an accurate description of stacking complexes and the size of these systems explain why the origin of this kind of noncovalent interactions is not so well known yet. At this point, we warn that it is too soon to obtain a detailed and final description of how they take place from an electronic point of view. Nevertheless, the reliability of the results provided by recent kinetic-optimized DFT functionals [6, 7] lead us to believe that a first approach to this matter can be written now. To this end, detecting electronic trends associated to stacking interactions and proposing a rough starting hypothesis about its electronic origin, we carry out diverse electron density analysis with the quantum theory of atoms in molecules (QTAIMs) [8, 9], which is considered among the most reliable tools of modern electron density analysis. Previously,

QTAIM has provided insight into the electronic origin of several chemical basic features [8–10] such as approximate transferability [11–15], diverse conformational preferences [16, 17], hydrogen bonding [18–21], strain energy [12, 22], characterization of intermolecular interactions [23], and so on.

After providing details of the computational techniques used, we summarize the results obtained in the study of some model compounds of charge-transfer complexes (quinhydrone and methyl gallate–caffeine adduct), homocomplexes (catechol crystalline structure) and C–H/ π interaction (benzene complexes with methane, acetylene and trichloromethane). After reporting the results for the first model we show the results obtained in the analysis of three combinations of B-DNA base-pair steps.

11.2

Computational Method

To achieve the above detailed objectives we propose to analyze the electron density provided by kinetic-optimized DFT functionals [6, 7, 24–32], like MPW1B95 [25], within the context of QTAIM [8–10, 33, 34]. The performance of these functional was checked by comparing the results obtained with experimental magnitudes (when available), or CCSD computed quantities (when the calculation was viable). In most cases (when the contrary is not stated in the text) we carried out single-point calculations with the Gaussian series of programs [35] on X-ray diffraction geometries taken from Cambridge database, or on geometries carefully obtained by computational methods in the case of DNA base-pair steps [36].

As QTAIM has been extensively reviewed [8–10, 33, 34], we restrict ourselves to introducing the nomenclature used henceforth. Thus, we remind the reader that between every couple of nuclei connected by a bond the electron density, $\rho(\mathbf{r})$, displays a (3,–1) singular point that is known as a bond critical point (BCP) and whose coordinates will be denoted by $\mathbf{r}_c$. The set of points connecting the BCP and the two bonded nuclei is known as the *bond path*. Among the properties computed at the BCP, the electron density, $\rho(\mathbf{r}_c)$, its Laplacian, $\nabla^2\rho(\mathbf{r}_c)$, and the value of the total energy density, $H(\mathbf{r}_c)$, play a fundamental role in describing the interatomic interactions. Higher values of $\rho(\mathbf{r}_c)$ indicate stronger bonds for the same pair of atoms, whereas positive values of $\nabla^2\rho(\mathbf{r}_c)$ and $H(\mathbf{r}_c)$ have been related generally to interactions between closed shells, in contrast to negative values, which are usually indicative of covalent bonds [23]. Application of the zero-flux condition (central to QTAIM and given by Equation 11.1) defines a set of surfaces, orthogonal to $\nabla\rho(\mathbf{r})$ and surrounding each nucleus, which joined with a contour where $\rho(\mathbf{r}_c)$ vanishes (usually 10^{-5} au) allows the definition of the atomic basin, Ω . Integration of the proper density function over such basins provides atomic properties. Some of these quantities are of interest for our purposes here: electron atomic population, $N(\Omega)$, obtained by (11.2) and its associated value of atomic charge, $q(\Omega)$, calculated by (11.3); atomic energy, $E(\Omega)$, obtained by multiplying the integrated value of the kinetic energy electron density function (11.4), $K(\Omega)$, by $(1 + \gamma)$, where γ is the molecular virial ratio

(ideally -2); the integrated value of the $L(\mathbf{r})$ function given by (11.5), $L(\Omega)$, which should be zero for a perfectly determined basin; the first moment of the atomic electron density, $\mu(\Omega)$, given by (11.6) with its module and components; the elements of the matrix of the atomic electron quadrupole moment, $Q_{ij}(\Omega)$, especially $Q_{zz}(\Omega)$ when z represents an axis that is orthogonal to the ring of an aromatic system, which is given by (11.7); and, finally, the atomic Shannon entropy of the electron distribution, $Sh(\Omega)$, obtained by (11.8):

$$\nabla \varrho(\mathbf{r}) d\mathbf{n} = 0 \quad (11.1)$$

$$N(\Omega) = \int_{\Omega} \varrho(\mathbf{r}) d\mathbf{r} \quad (11.2)$$

$$q(\Omega) = Z_{\Omega} - N(\Omega) \quad (11.3)$$

$$K(\mathbf{r}) = \frac{-N}{4} \int \cdots \int [\psi^* \nabla^2 \psi + \psi \nabla^2 \psi^*] d\mathbf{r}_2 \cdots d\mathbf{r}_N d\phi_1 \cdots d\phi_N \quad (11.4)$$

$$L(\Omega) = \frac{-1}{4} \nabla^2 \varrho(\mathbf{r}) \quad (11.5)$$

$$\mu(\Omega) = - \int_{\Omega} \mathbf{r}_{\Omega} \varrho(\mathbf{r}) d\mathbf{r} \quad (11.6)$$

$$Q_{zz}(\Omega) = - \int_{\Omega} (3z_{\Omega}^2 - r_{\Omega}^2) \varrho(\mathbf{r}) d\mathbf{r} \quad (11.7)$$

$$Sh(\Omega) = - \int_{\Omega} \varrho(\mathbf{r}) \ln |\varrho(\mathbf{r})| d\mathbf{r} \quad (11.8)$$

In each molecule, the QTAIM electron density analysis was performed with the AIMPACK [37] package of programs and AIM2000 [38]. Taking into account the small magnitude of electron density modifications involved in stacking complexation, it is crucial to check the accuracy of atomic integrations. This task was performed using standard criteria [12]. Thus, summations of $N(\Omega)$ and $E(\Omega)$ for each molecule reproduce total electron populations and electronic molecular energies below 10^{-3} au and 2 kJ mol^{-1} , respectively. No atom was integrated with absolute values of the $L(\mathbf{r})$ function [8, 9] larger than 10^{-3} au. The $N(\Omega) + L(\Omega)$ approximation [16, 39] was used to improve the accuracy of atomic electron populations. This approximation usually leads to a reduction in the difference between the total number of electrons in the molecule and the summation of $N(\Omega)$ values by one order of magnitude.

11.3

Charge-Transfer Complexes: Quinhydrone

Charge transfer (CT) between monomers has been frequently used to explain the stability of π -stacking heteromolecular complexes. The benzoquinone–hydroquinone complex (quinhydrone) is a simple and well known example, where CT between

Table 11.1 Stacking energies, ΔE , (kcal mol⁻¹) for quinhydrone.

	HF ^{a)}	B3LYP ^{a)}	MPWB1K ^{a)}	MPW1B95 ^{a)}	MPW1B95 ^{b)}	BH&H ^{a)}	MP2 ^{a)}	SCS-MP2
ΔE	4.76 ^{c)}	2.09 ^{c)}	-0.22	-2.80	-2.25	-6.11	-6.89 ^{c)}	-7.61

a) 6-311 + + G(2d,2p) 6d.

b) AUG-cc-pVTZ.

c) Includes BSSE correction.

the electron donor (hydroquinone) and the electron acceptor (quinone) has been generally assumed as the primary source for complex stabilization. In-plane intermolecular hydrogen bonds provide additional stability in solution and solid state [40]. Nevertheless, a previous computational study on quinhydrone using the MP2/6-31G (d) level and NBO analysis was unable to confirm the leading role of CT in the complex stabilization [41]. This led us to perform a QTAIM-based computational study on this system [42].

The fact that reliable experimental geometry [43] and stabilization energy [44] are available for the quinhydrone complex provide a good reference point for comparing the results provided by diverse computational levels. Thus, single point calculations for the crystal geometry of the adduct [43] were carried out using 6-311 + + G(2d, 2p) 6d basis set at HF, MP2, B3LYP and MPW1B95 and MPWB1K Truhlar's density functionals [25]. Stacking energies (Table 11.1), obtained from molecular energies computed for the adduct and quinone and hydroquinone isolated molecules (both constrained to their geometry in the adduct),¹⁾ reveal that HF and B3LYP, as could be expected from previous studies on other systems [2, 45–47], give rise to an unstable complex. Conversely, MP2 level calculations overestimate the stabilization energy, as also happens to DNA base pairs [24] and other systems [48]. This failure is not solved in this case by using Grimme's correction [49], as the quinhydrone complex results shown an even greater stabilization than with standard MP2 calculations. The hybrid BH&H level [50] produces a stacking energy close to that of MP2. Finally, MPW1B95 and MPWB1K give rise to stable quinhydrone with not so high stacking energy. In fact, the MPW1B95/6-311 + + G(2d, 2p) 6d value for the stacking energy agrees with the experimental data (2.8 ± 0.1 kcal mol⁻¹) (1 kcal 4.184 kJ) [44] and that obtained with the same functional and the AUG-cc-pVTZ basis set differs from the experimental value by less than 0.6 kcal mol⁻¹ (Table 11.1).

QTAIM analysis of the electron density obtained for quinhydrone at any of the computational levels here considered reveals the presence of four intermolecular

¹⁾ Even though the pairs in the crystal do not necessarily correspond to the lowest energy arrangements in the gas-phase adduct it will work as a good estimation for the purpose of this case. Counterpoise correction for basis set superposition error (BSSE) was only performed at the HF, MP2 and B3LYP levels, since MPW1B95 and MPWB1K functionals were developed in

such a way that they give reasonable results for noncovalent interactions both with and without counterpoise corrections, and the developers pointed out that they should be useable without the need of counterpoise corrections, especially when the basis is triple zeta quality or better (as is the case here).

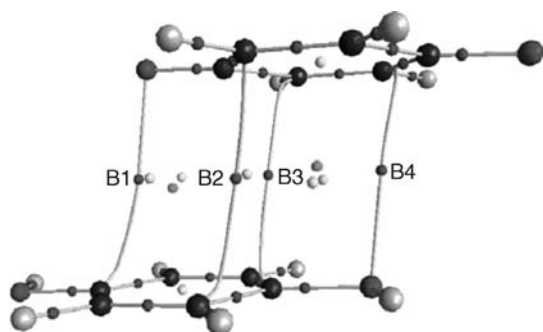
Table 11.2 Main properties (in au)^{a)} of the intermolecular BCPs (Figure 11.1) of quinhydrone.

BCP	$10^3 \varrho(r_c)$	$10^3 \nabla^2 \varrho(r_c)$	$10^3 H(r_c)$	R	ΔR
B1	7.0	21.0	1.2	3.375	0.331
B2	7.4	21.3	1.2	3.242	0.565
B3	6.4	22.6	1.1	3.171	0.231
B4	4.4	15.0	0.7	3.423	0.115

a) Internuclear distances, R , and differences between bond paths lengths and R , ΔR , in Å.

BCPs. Two of them correspond to $C \cdots C$ weak interactions and two to $C \cdots O$ ones, with the former showing higher density at the BCPs. All of them exhibit similar $\varrho(r_c)$ (between 4×10^{-3} and 7×10^{-3} au) and $H(r_c)$ (around $1 \cdot 10^{-3}$ au) values to those found in previous QTAIM work on stacking interactions in DNA bases [51]. We also observe quite large differences between bond path lengths and internuclear distances (Table 11.2). Although we had reported negative $\nabla^2 \varrho(r_c)$ values [42], this is not correct because the values presented in the paper were taken directly from the AIM2000 output, which provides $L(r_c)$ values, that is $-\nabla^2 \varrho(r_c)/4$ [8, 9]. Therefore, actual $\nabla^2 \varrho(r_c)$ values are positive (and four times larger), as usually found in π - π complexes [51–53]. If we exclude the B3LYP electron density (where the intermolecular bond path at the oxygen of hydroquinone is connected to the ipso carbon of quinone), all the electron distributions provide the same chemical graph (Figure 11.1).

The charge transfer that takes place within the complex has been also measured from the electron density, $\varrho(r)$, described by Truhlar's functionals. To accomplish this, the variations experienced by QTAIM atomic electron population, $\Delta N(\Omega)$, between the isolated molecules and the adduct, were obtained with good accuracy. Analysis of $\Delta N(\Omega)$ values also allows us to test the reliability of Mulliken's overlap and orientation principle, according to which the geometry of CT complexes is conditioned by obtaining the maximum overlap of the filled donor molecular orbital (HOMO) and the vacant acceptor orbital (LUMO) [54].

**Figure 11.1** Quinhydrone molecular graph (obtained with AIM2000 [38]), indicating the nomenclature for intermolecular bond paths.

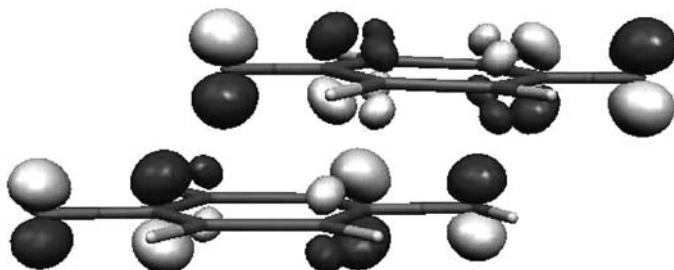


Figure 11.2 Side view (± 0.1 contours) of Kohn–Sham HOMO obtained for hydroquinone monomer and LUMO for quinone monomer at the MPW1B95/6-311 + + G(2d,2p) 6d level in the crystal geometry of quinhydrone.

QTAIM analysis indicates that the adduct formation is accompanied by some noticeable modifications of the atomic properties of the monomers (Table 11.3). Thus, there is an electron population transfer of 0.046 au from hydroquinone to quinone at the MPW1B95 level, confirming the CT character traditionally assigned to this adduct [55]. Analysis of the $\Delta N(\Omega)$ values (Figure 11.1) shows that the atoms experiencing the highest electron density loss are the carbons connected by a bond path to an oxygen of the other molecule. In addition, all the hydrogens belonging to the donor molecule lose electron density, while all the hydrogens in the acceptor molecule gain electron density.

Figure 11.2 shows the HOMO and LUMO calculated for hydroquinone and quinone monomers, respectively. As can be observed, in the case of hydroquinone the carbon atoms trans to hydroxyl do not participate in HOMO, while all the carbons participate in the LUMO of quinone. This would explain why bonding does not

Table 11.3 MPW1B95/6-311 + + G(2d,2p) 6d selected atomic properties in quinone (Q) and hydroquinone (H) monomers or variations experienced upon formation of their adduct (QH) (in au).

	$q(\Omega)$	$10^3 \Delta N(\Omega)$	$10^2 \Delta \mu_z(\Omega)$	$Q_{zz}(\Omega)$	$10^2 \Delta Q_{zz}(\Omega)$	$10^2 \Delta Sh(\Omega)$
Q(O)	0.951	13	−5.2	−0.137	0.4	0.5
Q(C) (carbonyl C)	−0.022	−3	−6.7	−1.765	10.9	−0.8
Q(C) (ortho C)	0.073	−2	0.7	−2.951	3.2	1.6
Q(H)	−1.055	9	0.9	−0.287	0.2	1.8
H(O)	0.440	−2	1.1	−1.235	1.8	−0.3
H(C) (ipso C)	−0.026	−3	5.2	−3.102	24.0	−1.6
H(C) (Z-ortho C)	0.034	−1	0.0	−3.227	15.3	−0.7
H(C) (E-ortho C)	0.575	−6	−0.3	−3.225	15.1	−0.7
H(H) (hydroxylic H)	0.028	−4	−0.3	0.003	0.0	−0.5
H(H) (Z-ortho H)	0.046	−3	0.2	−0.301	0.5	0.7
H(H) (E-ortho H)	−1.097	−4	0.1	−0.303	0.6	0.6

always occur between atoms that are in the same vertical alignment. Although some of the atoms displaying the largest $\Delta N(\Omega)$ also present a significant HOMO-LUMO overlap and participate in the intermolecular bond paths, other atoms show significant $\Delta N(\Omega)$ (Table 11.3). Therefore, Mulliken's overlap and orientation principle should be combined with a reorganization of electron density within each monomer to obtain the final atomic populations in the complex.

The six-center delocalization indices, Δ_6 , [56] calculated for each of the C_6 rings in the complex and monomers indicate a noticeable reduction of the local aromaticity of the C_6 ring of hydroquinone upon complex formation. Thus, it goes from 0.0218 au in the monomer to 0.0201 au in the complex. In contrast, the C_6 quinone ring (substantially less aromatic) displays the same Δ_6 value (0.0018 au) in both cases. Overall, the formation of quinhydrone complex is accompanied by a loss of electron density and electron delocalization in the hydroquinone ring. In contrast, the electron density gained by quinone in the complex is not reflected by the increase of any delocalization index within this unit. In addition, noticeable electron delocalization appears between both monomers in the complex.

The modifications experienced by $N(\Omega)$ are not large enough to alter the basic description of each monomer in the adduct. Therefore, quinone shows a strong positive charge on the carbonylic carbon, of around +1 au at any computational level, while the oxygen atoms have even stronger negative charges (Table 11.3). This extra charge is donated by the hydrogens (0.058–0.073 au depending on the computational level). In all three DFT levels, the results are similar and we have only found some differences at the MP2 level, where the atoms with the highest charge present even higher charge (carbonylic carbons and oxygens) [42], which is as usual when comparing MP2 and DFT results [57]. Hydroquinone also displays similar behavior in the three levels of calculation. In this case, the carbons bonded to the oxygens display around 0.5 au of positive charge, the hydroxylic hydrogens have a charge of 0.6 au and the oxygens bear a similar negative charge to those of quinone.

Although, on first thought, we did not pay attention to the variations experienced by other properties, further work has made us to reexamine this system and look at the variations of other integrated properties (see following sections). In particular, we notice that the variations experienced by the zz element (z being the axis perpendicular to C_6 rings) in the tensor of the atomic electronic quadrupole moment, $\Delta Q_{zz}(\Omega)$, are always positive (Table 11.3).

11.4

π - π Interactions in Hetero-Molecular Complexes: Methyl Gallate-Caffeine Adduct

The significant electron density transfer in the quinone-hydroquinone complex may lead to the consideration that such transfer could be present in all π - π complexes formed by monomers bearing a substantially different structure. Methyl gallate-caffeine may be taken as an example. It is also an example of the stacking complexes considered for explaining the antioxidative activity of polyphenols [58], the structure of which is well known [59, 60], which unfortunately is far from being a general case.

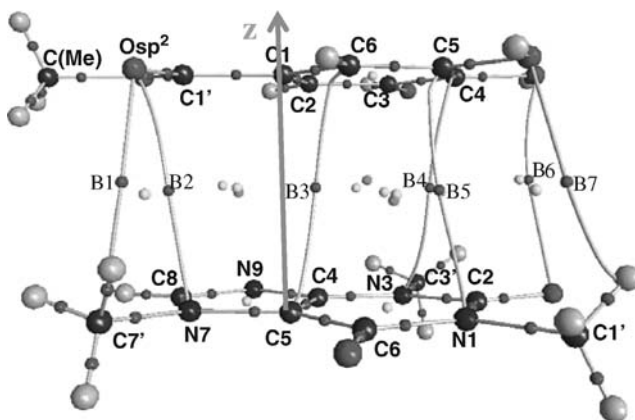


Figure 11.3 Methyl gallate–caffeine face to face adduct graph (obtained with AIM2000 [38]) indicating atom numbering (bold face) and the nomenclature for intermolecular bond paths.

Because of the good performance of the MPW1B95/6-311 + + G(2d, 2p) 6d level in quinhydrone [42] we also employed it to study the methyl gallate–caffeine π – π adduct. The geometry of this adduct was extracted from the crystal structure [60], where every methyl gallate (MG) molecule is surrounded in its plane by three molecules of caffeine. Reciprocally, every caffeine molecule is surrounded by three coplanar MG units. The in-plane intermolecular structure is due to three different kinds of hydrogen bonds, where MG always acts as H-donor: O3–H3 $\cdots$ O4, O4–H4 $\cdots$ N7 and O5–H5 $\cdots$ O2. Planes are displaced to allow face to face π – π stacking, where every caffeine molecule is stacked between two MG and vice versa. Thus, we have carried out single point calculations for the system formed by one caffeine molecule and one of its closest out of plane MG neighbors (Figure 11.3), and for both monomers in the geometry of this adduct.

QTAIM analysis of the MPW1B95/6-311 + + G(2d,2p) 6d electron density of the adduct reveals seven intermolecular BCPs. In accordance with previous findings in other stacking complexes [51–53], all of them display positive $\nabla^2 \rho(\mathbf{r}_c)$ values (Table 11.4). In all cases $\rho(\mathbf{r}_c)$, $\nabla^2 \rho(\mathbf{r}_c)$ and $H(\mathbf{r}_c)$ values are smaller than those obtained in quinhydrone, even for the C7'–H7' $\cdots$ O=C interaction (denoted as B1) where the interatomic distance is smaller than any of those associated to bond paths in quinhydrone. This can be interpreted as the π – π interactions in MG–caffeine being smaller than in quinhydrone.

Integration of the electron density within each atomic basin in the complex indicates that global CT between monomers is really small: 0.007 au from MG to caffeine. We also observe that, while in quinhydrone, all the atoms of one monomer display negative $\Delta N(\Omega)$ values, both monomers display atoms with positive and negative variations in this MG–caffeine adduct, all of them of very small amount (Table 11.5). The MG monomer shows atoms with the largest positive variations though. Despite small $\Delta N(\Omega)$ values that vary in a complicated fashion, we observe certain common trends for higher moments of the electron density, like $\mu(\Omega)$ and

Table 11.4 Main properties (in au) of the intermolecular BCPs (Figure 11.3) found for the adduct of caffeine and methyl gallate.

BCP	$10^3 \rho(r_c)$	$10^3 \nabla^2 \rho(r_c)$	$10^3 H(r_c)$	R (Å)
B1	4.8	19.5	0.9	2.849
B2	4.7	17.4	0.8	3.352
B3	5.3	14.9	0.8	3.501
B4	4.1	13.6	0.7	3.587
B5	3.8	13.4	0.7	3.571
B6	2.5	10.8	0.6	3.575
B7	3.2	13.0	0.7	3.062

$Q(\Omega)$ matrix, or statistics descriptors like $Sh(\Omega)$ (Table 11.5) that could be taken as indicators of the participation of electrostatic interactions in this complex. Thus:

- 1) The electron distribution of nearly all the basins of both monomers are more ordered (meaning closer to a uniform distribution) after the formation of the adduct, as indicated by negative $\Delta Sh(\Omega)$. The only exceptions are atoms that are

Table 11.5 Change of selected atomic properties in the formation of methyl gallate-caffeine adduct computed from MPW1B95/6-311 + + G(2d,2p) 6d electron densities (all values in au multiplied by 10^3).

Caffeine						Methyl gallate					
Ω	$\Delta N(\Omega)$	$\Delta Sh(\Omega)$	$\Delta Q_{zz}(\Omega)$	$\Delta \mu_z(\Omega)$	$\Delta \mu(\Omega)$	Ω	$\Delta N(\Omega)$	$\Delta Sh(\Omega)$	$\Delta Q_{zz}(\Omega)$	$\Delta \mu_z(\Omega)$	$\Delta \mu(\Omega)$
N1	2	-2	71	-108	8	C1	-1	-6	139	4	146
C2	-2	-3	26	-19	554	C1'	5	-1	45	-1	157
N3	1	-2	64	-12	5	C2	-4	-7	198	-11	104
C4	-3	-6	88	19	-399	C3	2	-3	89	36	-665
C5	-4	-6	127	-2	-325	C4	-6	-10	167	62	65
C6	3	0	22	-79	-368	C5	-1	-6	109	-97	-568
N7	-3	-4	114	49	-8	C6	3	-8	197	23	-543
C8	-2	-6	92	0	201	C(Me) ^{a)}	5	4	-11	-84	-559
N9	8	-1	142	54	-108	H2	-5	-8	20	-48	43
O2	-6	-5	105	81	-257	H3	-1	0	1	238	-260
O6	4	1	-13	-128	-216	H4	0	2	0	-5	-15
H8	7	2	22	83	10	H5	-1	-5	5	-27	-28
C1' ^{a)}	4	-9	5	-40	-223	H6	-2	-10	28	-34	16
C3' ^{a)}	6	-12	56	168	-4	O1'sp ²	11	-3	141	-7	240
C7' ^{a)}	-7	-28	34	-121	399	O1'sp ³	-1	-1	-5	-29	121
						O3	2	-1	38	4	-162
						O4	-9	-5	54	25	80
						O5	-1	-5	92	-57	146

a) $\Delta N(\Omega)$ and $\Delta Sh(\Omega)$ values refer to the whole methyl group.

- in the furthest positions with regard to the other monomer. The summation of $\Delta Sh(\Omega)$ values is -0.080 au in caffeine and -0.074 au in MG.
- 2) Atomic electronic dipole moments vary significantly for many basins. These variations give rise to noticeable modifications of the z-component (see Figure 11.3 for z-axis definition) of the dipole moment of the monomers. This effect takes place through the contribution due to polarization of atomic distributions [obtained as $\Sigma \Delta \mu_z(\Omega)$] in MG (0.060 au) and caffeine (-0.089 au) from the monomer to the adduct.
 - 3) The zz element of the electronic quadrupoles of all the basins, except those furthest away from the other monomer (OCH_3 group of MG and O6 in caffeine), increase upon complexation. This indicates that adduct formation has been accompanied by a certain flattening of the prolate spheroid representing the electron density analogue of π population, which, therefore, is more concentrated towards the corresponding π nodal plane in each monomer.

11.5

π - π Interactions between DNA Base Pair Steps

Noncovalent interactions among base heterocycles are among the key contributions to the structure and dynamics of nucleic acids [61]. In fact, whereas hydrogen bonding (HB) is responsible for complementary base pairing and the puckering of the sugar moiety determines the type of DNA (A, B or Z) [62], important geometry properties, like the diameter of the helix and the number of residues per turn, are influenced by stacking between neighboring pairs of bases [63, 64]. Thus, stacking complexes between base DNA pair steps are a biologically interesting example of systems to test if they also follow the above reported electron reorganization trends, which would reinforce them as a starting hypothesis for describing the electron origin of π - π stacking.

The same analysis carried out for the two complexes indicated in the previous sections was extended to 3 of 16 possible duplexes of DNA base-pair steps: AT-AT, GC-GC and GC-AT (A = adenine, C = cytosine, G = guanine, T = thymine). According to this nomenclature, the adduct formed by the bases of the $5'-G \cdots C-3'/5'-A \cdots T-3'$ duplex (where the slash separates first and second steps, which will be represented as 1 and 2 henceforth) is represented by GC-AT. The geometries selected for these systems correspond to the most common form of DNA (DNA-B) [65] and were taken directly from the idealized pair-base steps proposed in a recent computational study [36]. The electron density obtained from MPW1B95/6-311 + + G(2d,2p) 6d single points of adduct and monomers (now composed of two bases attached by HB) were analyzed subsequently with QTAIM.

These systems had been previously analyzed in detail with QTAIM [51], providing interesting information about both base-pairing HB and stacking interactions, on the exclusive basis of BCP properties without reporting results on atomic properties. In addition, the study was carried out at a different computational level and on different geometries, which were extracted from the structure determined

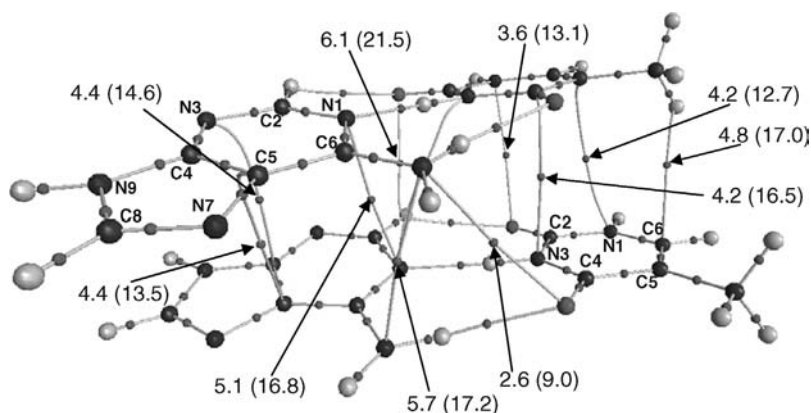


Figure 11.4 AT-AT adduct graph (obtained with AIM2000 [38]) indicating atom numbering (bold face) and $q(\mathbf{r}_c)$ and $\nabla^2 q(\mathbf{r}_c)$ (in parenthesis) values (both in au multiplied by 10^3) for intermolecular bond paths.

by high-resolution spectroscopy for the $d(\text{CATGGCCATG})_2$ DNA decamer [66]. Therefore, these systems also allow us to test the sensibility of BCPs of stacking interactions to modification of geometries and computational levels.

According to our study, the AT-AT complex (Figure 11.4) has six HB BCPs and ten BCPs associated to stacking interactions. The six HB BCPs and associated bond paths confirm once more the presence of a third HB in the AT pair [51, 67–69]. The HB not usually described in textbooks is established between C2–H in adenine and O=C2 in thymine and displays, as previously observed [51, 69], much smaller $q(\mathbf{r}_c)$ and $\nabla^2 q(\mathbf{r}_c)$ (Table 11.6). We also observe the well known dependency of BCPs properties with internuclear distance [70–72]. The differences in the geometry of both base-pair steps provide changes in bond properties as large as 0.006 au for $q(\mathbf{r}_c)$ or 0.02 au for $\nabla^2 q(\mathbf{r}_c)$. In contrast, no important difference is observed in BCP properties when comparing the values of adduct and the corresponding isolated pair of bases. Thus, differences are below 10^{-4} au in $q(\mathbf{r}_c)$ and 10^{-3} au in $\nabla^2 q(\mathbf{r}_c)$.

Using the same terms employed in Matta *et al.*'s paper [51], eight of the stacking interactions correspond to intrastrand interactions (four between A molecules, and four between T molecules) and two mirror interstrand interactions: N6(A1) $\cdots$ O4(T2) and O2(T1) $\cdots$ H2(A2). The values are comparable in all cases to those obtained for quinhedrone and MG–caffeine complexes. Despite the differences in geometry and computational level, both the interactions and their properties are in good agreement with those previously reported by Matta *et al.* The only differences are (i) the assignation of one of the interactions: C5(A1) $\cdots$ C5(A2) in our study and C4(A1) $\cdots$ C5(A2) in theirs and (ii) another intrastrand interaction, C6(A1) $\cdots$ C6(A2), was found in the previous paper.

The descriptions of stacking and HB BCPs for the GC-GC adduct (Figure 11.5) given by the previous study and ours are in total coincidence, even for individual values of $q(\mathbf{r}_c)$ and $\nabla^2 q(\mathbf{r}_c)$. For HB BCPs we observe differences between values

Table 11.6 Main properties (in au; except internuclear distances, R , in Å) of HBs detected in QTAIM analysis of AT-AT, CG-CG and CG-AT adducts; all values computed from MPW1B95/6-311++G (2d,2p) 6d electron densities.

Adduct	HB ^{a)}	Step 1			Step 2		
		$10^3 \varrho(r_c)$	$10^3 \nabla^2 \varrho(r_c)$	R	$10^3 \varrho(r_c)$	$10^3 \nabla^2 \varrho(r_c)$	R
AT-AT	N1...H-N3	36.9	93.6	1.861	39.8	93.7	1.827
	N6-H...O=C4	19.9	67.1	2.055	13.9	46.1	2.601
	C2-H...O=C2	4.9	16.7	2.811	7.3	24.3	2.601
GC-GC	N1-H...N3	39.0	108.2	1.831	38.7	93.9	1.832
	C6=O...H-N4	32.2	101.6	1.843	30.6	102.1	1.855
	N2-H...O=C2	30.6	100.2	1.848	31.4	104.7	1.838
GC-AT	N1-H...N3	26.6	80.8	2.009			
	C6=O...H-N4	28.9	93.3	1.895			
	N2-H...O=C2	24.0	80.2	1.965			
	N1...H-N3				40.4	101.4	1.819
	N6-H...O=C4				15.3	53.7	2.138
	C2-H...O=C2				6.9	22.7	2.647

a) Numbering before HB interaction (...) refers to the first base in adduct number, and that after the interaction to the second base.

reported in both studies of around 2% for $\varrho(r_c)$ and 10% for $\nabla^2 \varrho(r_c)$. To explain such excellent concordance we checked whether the interatomic distances reported in Reference [51] are exactly the same as with our geometries, finding that both GC-GC geometries are equal or very similar. Therefore, the main source of differences between the two studies should be the geometry and not the

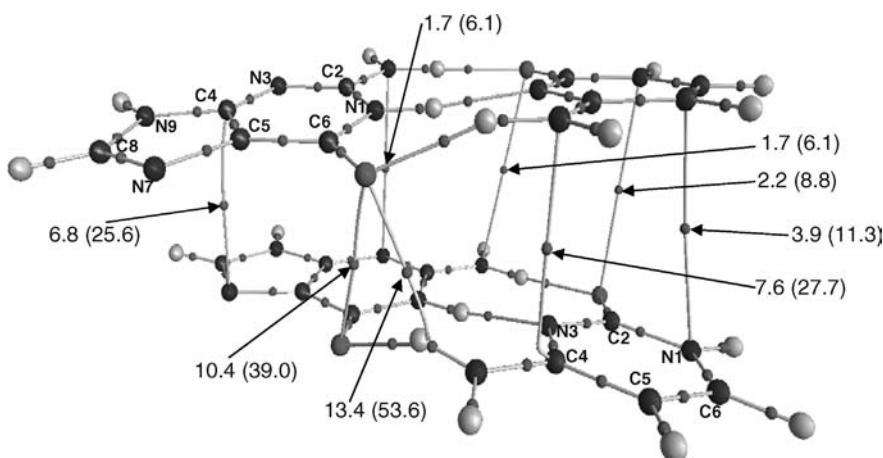


Figure 11.5 GC-GC adduct graph (obtained with AIM2000 [38]) indicating atom numbering (bold face) and $\varrho(r_c)$ and $\nabla^2 \varrho(r_c)$ (in parenthesis) values (both in au multiplied by 10^3) for intermolecular bond paths.

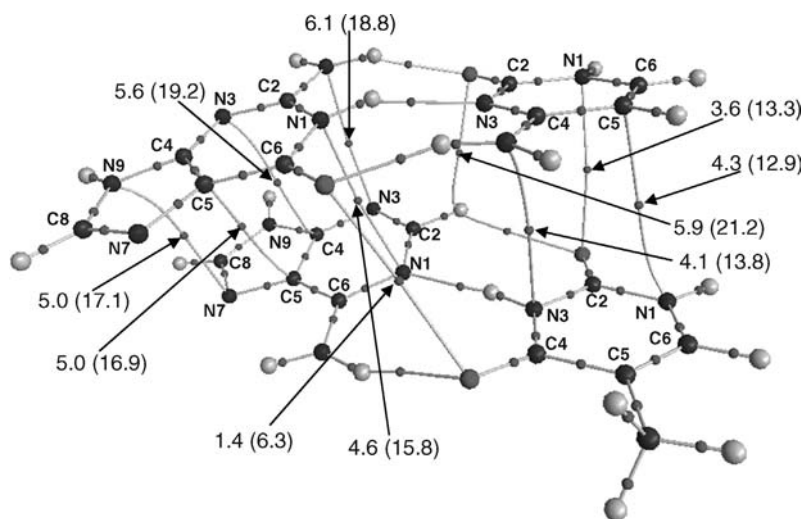


Figure 11.6 GC-AT adduct graph (obtained with AIM2000 [38]) indicating atom numbering (bold face) and $q(r_c)$ and $\nabla^2 q(r_c)$ (in parenthesis) values (both in au multiplied by 10^3) for intermolecular bond paths.

computational level (B3LYP/6-311 + + G(d,p) in Reference [51]). Finally, in this case, the strength of the three HBs is very similar and the differences between HB properties in steps 1 and 2 are much smaller (Table 11.6).

Frozen geometries for GC-AT adduct (Figure 11.6) are clearly different in both studies (Table 11.6 and Reference [51]). This produces different values for the HB BCP properties. In contrast, it does not produce any significant change in stacking description, if we exclude the fact that one intrastrand bond path more, connecting G and A, was previously reported [51].

Overall, we conclude that the two approaches for the analysis of BCP stacking properties are compatible and they basically provide the same picture; that is, several intrastrand and two interstrand bond paths connecting atoms of both base-pair steps that are close enough. Relative values of $q(r_c)$ cannot be inferred from this inter or intra character and are mainly affected by internuclear distances (Figures 11.4–11.6). Although $q(r_c)$ values assigned to stacking interactions are below those of traditional HBs, we notice that, in the duplexes studied here, some of them exceed those displayed by the BCP associated to C2–H $\cdots$ O=C2 HB, that is “the third HB” in the AT pair. Finally, all the $\nabla^2 q(r_c)$ values assigned to stacking interactions are positive.

QTAIM atomic properties were computed for the three duplexes and the corresponding base pair steps in the geometry. The results, summarized as summations of atomic properties for each base in Table 11.7, provide some conclusions:

- 1) Every base displays a small partial charge, positive for purines and negative for pyrimidines, which act as hydrogen donors in two HB of each pyrimidine–purine pair.

Table 11.7 Molecular charge and summations of the variations of selected atomic properties (in au, except $\Sigma\Delta E(\Omega)$ in kJ mol^{-1}) experienced upon duplex formation from base pair steps in AT-AT, CG-CG and CG-AT; all values computed from MPW1B95/6-311++G(2d,2p) 6d electron densities.

Duplex	Unit	$\Sigma q(\Omega)$	$\Sigma\Delta N(\Omega)$	$\Sigma\Delta E(\Omega)$	$\Sigma\Delta\mu_z(\Omega)$	$\Sigma\Delta Q_{zz}(\Omega)$	$\Sigma\Delta Sh(\Omega)$	$\Sigma\Delta\nu(\Omega)$
AT-AT	A1	0.042	-0.013	218	-0.269	1.878	-0.028	-14.6
	T1	-0.040	0.012	-278	0.624	5.362	-0.045	-4.3
	A2	0.028	0.009	-219	0.114	1.498	-0.080	-13.6
	T2	-0.030	-0.007	256	0.436	5.498	-0.042	-8.6
CG-CG	C1	0.040	0.004	57	0.100	4.237	-0.018	-6.1
	G1	-0.039	-0.005	22	-0.401	9.836	-0.035	-25.0
	C2	0.040	0.006	-36	0.484	2.463	-0.094	-13.3
	G2	-0.041	-0.005	-40	0.051	4.019	-0.033	-17.8
CG-AT	C1	0.033	-0.004	-11	0.749	7.981	-0.031	-7.9
	G1	-0.031	0.002	-20	-0.037	4.310	-0.033	-10.3
	A2	0.040	-0.007	12	-0.517	3.139	-0.076	-16.3
	T2	-0.042	0.009	4	0.643	6.418	-0.017	1.3

- Most of the charge borne by bases is due to HB, but it is modified after duplex formation in a non-negligible extent. Thus, $\Sigma\Delta N(\Omega)$ represents in AT-AT bases a fifth to a third part of the net charge.
- Electron density of the monomers is substantially polarized upon duplex formation, as indicated by $\Sigma\Delta\mu_z(\Omega)$ values (z -axis is perpendicular to nearly orthogonal to the main planes of each pair of bases). The same is true for quadrupolar moments, whose $Q_{zz}(\Omega)$ element is significantly enhanced when the pairs of bases stack. In fact, all the atoms in the three duplexes, except those hydrogens of methyl groups placed in outer disposition to the other step, display positive $\Delta Q_{zz}(\Omega)$.
- In addition, nearly all the atoms display negative $\Delta Sh(\Omega)$ and $\Delta\nu(\Omega)$ values; $\nu(\Omega)$ was computed by integrating the intersection of zero-flux surfaces and the 10^{-3} au contour of $\rho(\mathbf{r})$ [8, 9]. This indicates that the reorganization of electron density that accompanies duplex formation is concentrated mainly in the most diffuse electron density of each atomic basin.

Overall, the general electronic trends observed in the formation of other stacked complexes are also followed by the examples of DNA duplex formation considered here.

11.6

π - π Interactions in Homo-Molecular Complexes: Catechol

Two of the DNA duplexes studied above are formed by the same pairs of bases. Nevertheless they bear a different geometry that precludes symmetry and allows

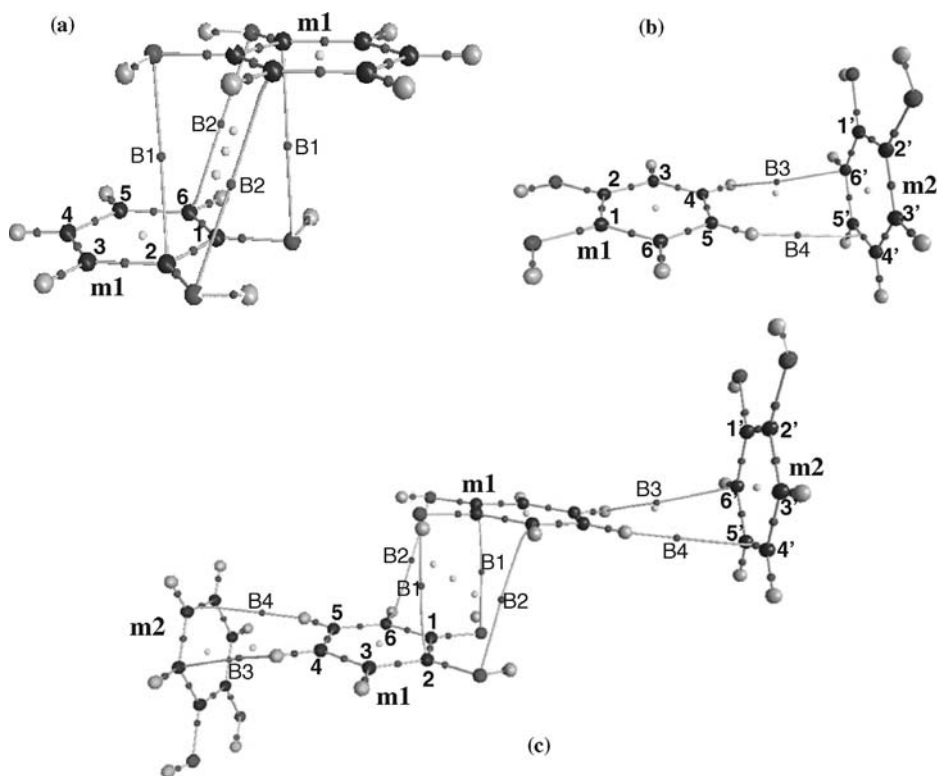


Figure 11.7 Molecular graphs (obtained with AIM2000 [38]) indicating atom numbering and nomenclature for monomers and intermolecular BCPs of face to face dimer (a), C–H/ π dimer (b) and tetramer (c) of catechol.

certain CT. The face to face (FF) dimer of catechol (Figure 11.7) displays C_i symmetry and therefore we should find strictly neutral monomers in it. Also in this case, we have carried out the same kind of calculations: single point MPW1B95/6-311++G(2d,2p) 6d calculation on the geometry of the crystal, where face to face and C–H/ π catechol dimers are present. [The geometry of crystalline catechol was obtained from Cambridge Crystallographic Data Center (CCDC).] This allows us to analyze cooperative effects, studying both dimers separately and combined in the tetramer.

The FF dimer displays four intermolecular bond paths (Figure 11.7a), which by symmetry can be reduced to two different interactions: $C2 \cdots O1$ and $O2 \cdots C6$. The C–H/ π dimer presents only two intermolecular bond paths (Figure 11.7b). Considering the similarity among all the $Q(r_c)$ values (Table 11.8), the larger number of bond paths can be invoked to justify the preference for the FF dimer, whose dimerization energy is $-11.7 \text{ kJ mol}^{-1}$, whereas that of the C–H/ π one is only -3.1 kJ mol^{-1} (both values obtained without ZPVE corrections). The molecular

Table 11.8 Main properties (in au; internuclear distances, R , in Å) of the intermolecular BCPs (Figure 11.7) found for the dimers and tetramer of catechol.

Interaction	BCP	R	Dimer $10^3 q(r_c)$	Tetramer $10^3 q(r_c)$	Dimer $10^3 \nabla^2 q(r_c)$	Tetramer $10^3 \nabla^2 q(r_c)$
π – π	B1	3.429	4.39	4.37	14.56	14.60
	B2	3.272	5.96	5.97	21.17	21.22
C–H/ π	B3	3.091	3.58	3.56	11.48	11.50
	B4	2.831	6.31	6.31	20.21	20.21

graph of the tetramer is just the superposition of those obtained for both dimers. Table 11.8 shows that we need to add one digit more when writing $q(r_c)$ and $\nabla^2 q(r_c)$ values in order to observe differences between dimers and tetramer. Consequently, cooperative effects between π – π and C–H/ π interactions in the tetramer on BCP properties are negligible.

In contrast with the symmetric FF dimer, the C–H/ π dimer is formed with a non-negligible CT (0.011 au are transferred from monomer m2 to m1). This CT remains the same in each C–H/ π unit of the tetramer. As a consequence of CT, m2 destabilizes (9 kJ mol^{-1}) to a lower extent than the stabilization gained by m1 ($-10.5 \text{ kJ mol}^{-1}$), which is also larger than that gained by each m1 unit during the formation of the FF dimer (-6 kJ mol^{-1}). Finally, the stabilization of m1 in the tetramer is -20 kJ mol^{-1} , revealing cooperative effects (-3.5 kJ mol^{-1}) between the π – π and C–H/ π interactions affecting the same monomer. In contrast, m2 experiences the same destabilization in dimer and tetramer.

Analysis of integrated properties reveals again that important changes in the polarization of basins, mainly indicated by $\Delta\mu_z(\Omega)$ and $\Delta Q_{zz}(\Omega)$ (data not shown), take place upon complex formation. This effect is more intense in the atoms connected by bond paths or which are close to them. We even observe that, although the global CT for C–H/ π formation is the same in dimer or tetramer, the evolution of the electron density is different. Thus, the subset of atoms attached to the other monomer by intermolecular bond paths, $\{\Omega^*\}$, of m2 goes from losing electron population in the dimer to gaining it in the tetramer (Table 11.9). The decrease in atomic values of the scalar first and second moments of $q(\mathbf{r})$ [denoted as $r^1(\Omega)$ and $r^2(\Omega)$, respectively] also indicates the electron density approaches, on average, the nucleus of basins and it turns to a more spherical distribution after complex formation, explaining why $\Delta Sh(\Omega)$ values also decrease (Table 11.9).

Finally, atomic volumes computed with 0.001 and 0.002 au contours, $v_1(\Omega)$ and $v_2(\Omega)$, and the electron population enclosed respectively by them, $N_1(\Omega)$ and $N_2(\Omega)$, and the electron population enclosed between both contours, $\Delta N_{12}(\Omega)$, shown in Table 11.10, clearly indicate that, in most of the atomic basins, the most diffuse part of $q(\mathbf{r})$ becomes more concentrated after complex formation, enlarging $v_2(\Omega)$ and reducing $\Delta N_{12}(\Omega)$.

Table 11.9 Variations experienced by selected integrated properties of catechol monomers (Figure 11.7) during dimer and tetramer formation (all values in au multiplied by 10^3).

	Unit	$\Sigma\Delta N(\Omega)$	$\Sigma\Delta N(\Omega^a)$	$\Sigma\Delta r^1(\Omega)$	$\Sigma\Delta r^2(\Omega)$	$\Sigma\Delta Sh(\Omega)$
FF	m1	0	5	–63	–467	–57
CH/ π	m1	11	–15	19	19	–49
	m2	–11	–6	–61	–353	–39
Tetramer	m1	11	–13	–53	–527	–112
	m2	–11	5	–66	–382	–38

a) Ω^a refers to atoms connected through intermolecular bond paths.

11.7

C–H/ π Complexes

The weak attraction between a C–H bond and π system was often described as the weakest class of conventional hydrogen bonds. Nevertheless, recently reviewed [73] theoretical and spectroscopic studies indicate that, while the electrostatic interaction is mainly responsible for the attraction in the conventional hydrogen bonds [19–21, 74], dispersion has been recognized as the major source of attraction between C–H and π units, with a very small electrostatic contribution [75]. Moreover, the directionality of C–H/ π interaction is very weak compared to conventional hydrogen bonds [73]. In this section we report results obtained in the QTAIM analysis of three usual model systems of C–H/ π complexes: (i) methane–benzene, (ii) acetylene–benzene and (iii) trichloromethane–benzene (Figure 11.8). They are compared with those obtained for usual examples of C–H $\cdots$ O hydrogen bonding and noncovalent π – π interactions.

In this case, although crystal geometries are available for all of them [76–78], the geometries of the three complexes and their monomers were completely optimized at the MPW1B95/6-311++G(2d,2p) 6d level. The methane–benzene complex and monomers were also optimized at the CCSD/6-31++G(2d,2p) 6d levels.

Table 11.10 Variations experienced by atomic volumes and related properties of catechol monomers (Figure 11.7) during dimer and tetramer formation (all values in au).

	Unit	$\Sigma\Delta\nu_1(\Omega)$	$\Sigma\Delta\nu_2(\Omega)$	$\Sigma\Delta N_1(\Omega)$	$\Sigma\Delta N_2(\Omega)$	$\Sigma\Delta N_{12}(\Omega)$
FF	m1	–11.9	7.2	0.042	0.067	–0.025
CH/ π	m1	2.2	7.4	0.035	0.042	–0.007
	m2	–4.8	4.0	0.017	0.029	–0.012
Tetramer	m1	–8.0	14.3	0.078	0.109	–0.031
	m2	–4.9	3.8	0.016	0.028	–0.012

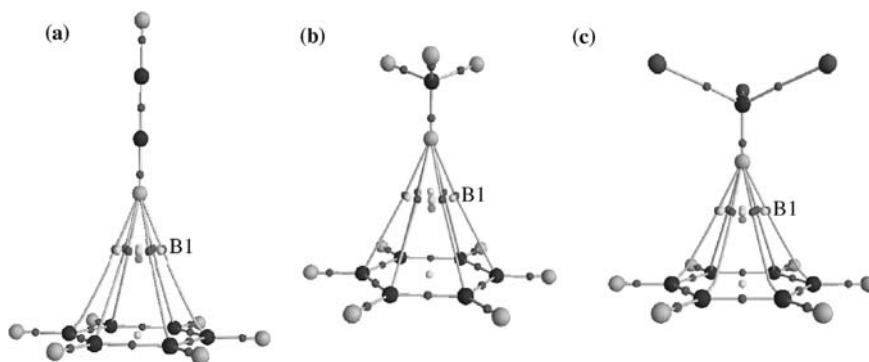


Figure 11.8 Molecular graphs (obtained with AIM2000 [38]) for C–H/ π adducts of benzene with acetylene (a), methane (b) and trichloromethane (c).

The variations experienced by QTAIM integrated properties of the monomers upon methane–benzene complex formation are significantly similar (Table 11.11), indicating that MPW1B95/6-311++G(2d,2p) 6d electron distributions are a reasonable approach to describe the formation of C–H/ π complexes. Thus, both sets of calculations indicate a small electron density transfer from benzene to methane (0.007 au), the depletion of $N(\text{H}^*)$, with H^* denoting the methane H involved in the interaction, and the increase of $N(\Omega)$ for all the atoms of the methyl group. We also notice a significant polarization of H^* , opposite to that experienced by the methyl carbon. Finally, $Q_{zz}(\Omega)$ values increase substantially for benzene carbons and H^* ; that is, all the atoms directly involved in the C–H/ π interaction. It is remarkable that all the electronic trends listed by Koch and Popelier as characteristic of hydrogen bonds [20] are shown by this system if we exclude the decrease of the hydrogen atom's volume, which has not been considered as a necessary condition because of the numerous exceptions reported [20, 79].

Looking at BCP properties, we observe that the six symmetric bond paths connecting H^* and the carbons of the benzene ring (Figure 11.8) display certainly

Table 11.11 Selected relative values of atomic properties for methane–benzene (in au multiplied by 10^3).

Ω^a	CCSD/6-311++G(2d,2p) 6d				MPW1B95/6-311++G(2d,2p) 6d			
	$\Delta N(\Omega)$	$\Delta \mu(\Omega)$	$\Delta \mu_z(\Omega)$	$\Delta Q_{zz}(\Omega)$	$\Delta N(\Omega)$	$\Delta \mu(\Omega)$	$\Delta \mu_z(\Omega)$	$\Delta Q_{zz}(\Omega)$
C(b)	0	3	7	98	3	3	3	90
H(b)	–1	0	0	0	–4	–1	0	1
H^*	–15	–19	–19	101	–19	–21	–21	109
C(m)	4	14	14	–52	13	16	16	–49
H(m)	6	2	–2	1	4	2	–2	2

a) (b) refers to benzene and (m) to methane.

Table 11.12 Main geometry features and BCP properties related to C–H/ π bonds in complexes studied here; all values in au (except R , internuclear distance, and d , distance from H* to benzene RCP, in Å).

Complex	$10^3 \varrho(r_c)$	$10^3 \nabla^2 \varrho(r_c)$	$10^3 H(r_c)$	R	d
CH ₄ –C ₆ H ₆ MPW1B95	3.8	12.96	0.63	3.108	2.781
CH ₄ –C ₆ H ₆ CCSD	3.9	13.36	0.61	3.113	2.784
C ₂ H ₂ –C ₆ H ₆ MPW1B95	3.2	11.41	0.57	3.142	2.820
C ₂ H ₂ –C ₆ H ₆ CCSD	5.3	19.43	0.85	2.886	2.527
Cl ₃ CH–C ₆ H ₆ MPW1B95	7.5	25.95	1.11	2.737	2.361

small $\varrho(r_c)$ values (Table 11.12), but – for instance – they are not smaller than those obtained for C–H $\cdots$ O bonds in the dimers of methoxymethane [80] or the acetone–benzene complex [20], and are scarcely exceeded by that found for the C–H $\cdots$ O bond found in one trimer of methanol (9.6×10^{-3} au) [81], or even by that of the F–H $\cdots$ Cl–H adduct (7.2×10^{-3} au) [19]. Moreover, $\nabla^2 \varrho(r_c)$ values are positive. They again exceed those obtained in the examples of C–H $\cdots$ O bonds given above [20, 79] and are a half of that reported for the only trimer of methanol containing C–H $\cdots$ O hydrogen bonds. Total energies at the C–H/ π and C–H $\cdots$ O BCPs are also comparable (0.9×10^{-3} au in the methanol trimer or 0.6×10^{-3} au in one of the formaldehyde dimers [82]). Overall, no significant difference is obtained when comparing the main BCP properties of C–H $\cdots$ O and C–H/ π bond paths.

Values of BCP properties increase with the acidity of H*, as can be observed on comparing the results obtained for the three complexes studied here (Table 11.12). MPW1B95 results for acetylene–benzene are an exception, probably due to the lower reliability of the computational level compared to CCSD.

Looking at the variations experienced by QTAIM atomic properties along the three complexes (Table 11.13), we notice that, as expected for hydrogen bonds, both the electronic population and the dipolar polarization of H* decrease. In contrast, $\nu(H^*)$ is only reduced for the strongest complex (C₆H₆–Cl₃CH) and employing the 0.001 au electron density envelope, $\nu_1(\Omega)$. The only component of $\mu(\Omega)$ that is substantially modified is that parallel to the symmetry axis, μ_z . $\Delta\mu_z(\Omega)$ values indicate that complex formation is accompanied by an accumulation of electron density along the H*–C bond of the non-aromatic monomer. Another significant modification of the electronic distribution due to complex formation is shown by the $Q_{zz}(\Omega)$ values of the atoms directly concerned in the intermolecular interaction (H* and benzene carbons) (Table 11.13).

The acidity of H*, and consequently the stability of the complex (Table 11.14), increases along the series CH₄ < C₂H₂ < Cl₃CH. No direct relation between acidity of H* and variation of a single property is observed in Table 11.13, if we exclude the increase of $\Delta Q_{zz}(C)$ for the carbons of benzene. The effects of this acidity sequence become clearer if we define three regions in the complex: benzene monomer (b), H* and the rest of the other monomer (R) (Table 11.14).

Table 11.13 Selected relative values of atomic properties computed with MPW1B95/6-311++G (2d,2p) 6d electron densities in compound–benzene adducts (in au multiplied by 10^3 , except $\Delta\nu(\Omega)$ values).

Compound	Ω	$\Delta N(\Omega)$	$\Delta\mu(\Omega)$	$\Delta\mu_z(\Omega)$	$\Delta Q_{zz}(\Omega)$	$\Delta Sh(\Omega)$	$\Delta\nu_1(\Omega)$	$\Delta\nu_2(\Omega)$
CH ₄	C(b)	3	3	3	90	4	0.0	1.1
H(b)	–4	–1	0	1	3	0.0	0.0	
H*	–19	–21	–21	109	–78	1.1	4.3	
C	13	16	16	–49	9	0.8	0.6	
H	4	2	–2	2	15	0.5	0.4	
C ₂ H ₂	C(b)	6	10	–5	126	7	1.0	1.7
H(b)	–8	7	0	10	22	0.8	0.8	
H*	–4	–3	–3	58	–18	4.9	5.4	
C*	2	14	14	–15	7	3.0	1.9	
C	3	7	–7	92	11	3.7	1.2	
H	10	5	–5	–11	41	1.7	0.9	
Cl ₃ CH	C(b)	6	3	–5	167	–3	–1.0	0.6
H(b)	–9	–2	1	13	–15	–0.8	–0.2	
H*	–24	–18	–18	117	–110	–7.4	0.6	
C	11	36	37	–83	7	1.0	1.0	
Cl	11	15	–20	114	3	3.0	1.8	

Thus, increasing acidity of H* results in a larger electron transfer from b. Another important factor in explaining the electron density evolution is the size of the other monomer and its associated electron density attractors. In fact, $\Delta N(R)$ increases with the summation of its atomic numbers. Summation of $\Delta E(\Omega)$ values (computed from Kohn–Sham MOs) for each of these regions reveals different origins for the stabilization of each complex. Although $\Delta N(\Omega)$ and $\Delta E(\Omega)$ values usually display a reverse relationship in many cases [83], it is not applicable in this series for diverse reasons. Thus, in CH₄–C₆H₆ the small electron transfer from benzene is not enough to destabilize it or stabilize the RH* monomer. C₂H₂–C₆H₆ has a very small number of atomic basins in R to obtain an efficient distribution for the electron density gained by these atoms, and $\Delta N(R)$ and $\Delta E(R)$ are both positive. Finally, the very large $\Delta E(\Omega)$ variations observed in Cl₃CH–C₆H₆ is a consequence of the introduction of large attractors and important electron–electron repulsions.

Table 11.14 Variations of electron population (in au multiplied by 10^3) and energy (kJ mol^{–1}) due to the formation of the C–H/ π complex denoted as R–H/C₆H₆.

Complex	$\Delta N(b)$	$\Delta N(H^*)$	$\Delta N(R)$	ΔE	$\Delta E(b)$	$\Delta E(H^*)$	$\Delta E(R)$
CH ₄ –C ₆ H ₆	–7	–19	26	–2.1	–62	24	36
C ₂ H ₂ –C ₆ H ₆	–10	–5	15	–7.8	–31	–13	37
Cl ₃ CH–C ₆ H ₆	–21	–24	45	–13.6	1048	34	–1096

11.8

Provisional Conclusions and Future Research

Our main (provisional) conclusions, provisional because more data are needed to confirm them, can be summarized in the following work hypotheses:

- 1) Kinetic-optimized DFT functionals provide electron density distributions for stacking and related complexes that lead to similar conclusions to those obtained from higher computational levels.
- 2) The formation of stacking complexes is accompanied by a significant modification of electronic polarization that is especially noticeable through $Q_{zz}(\Omega)$ values.
- 3) C–H/ π interactions cannot be clearly distinguished, from an electronic point of view, from hydrogen bonds, especially from the weak ones like C–H $\cdots$ O.

Our research will concentrate, in the near future, on testing these hypotheses by enlarging the database of atomic properties for π – π stacking and C–H/ π complexes.

Acknowledgments

Free access to computational resources of Centro de Supercomputación de Galicia (CESGA) is gratefully acknowledged. We also thank Dr Antonio Vila, Dr José Manuel Hermida-Ramón and Mr Nicolás Otero for helpful contributions.

References

- 1 Meyer, E.A., Castellano, R.K., and Diederich, F. (2003) *Angew. Chem. Int. Ed. Engl.*, **42**, 1210–1250.
- 2 Müller-Dethlefs, K. and Hobza, P. (2000) *Chem. Rev.*, **100**, 143.
- 3 Hunter, C.A., Lawson, K.R., Perkins, J., and Urch, C.J. (2001) *J. Chem. Soc., Perkin Trans. 2*, 651.
- 4 Hobza, P. (2008) *Phys. Chem. Chem. Phys.*, **10**, 2581–2583.
- 5 Černý, J. and Hobza, P. (2007) *Phys. Chem. Chem. Phys.*, **9**, 5291–5303.
- 6 Zhao, Y. and Truhlar, D.G. (2008) *Acc. Chem. Res.*, **41**, 157–167.
- 7 Zhao, Y. and Truhlar, D.G. (2007) *J. Chem. Theor. Comp.*, **3**, 289–300.
- 8 Bader, R.F.W. (1991) *Chem. Rev.*, **91**, 893.
- 9 Bader, R.F.W. (1990) *Atoms in Molecules: A Quantum Theory*, Oxford University Press, Oxford.
- 10 Matta, C.F. and Boyd, R.J. (eds) (2007) *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*, Wiley-VCH Verlag GmbH, Weinheim.
- 11 Wiberg, K.B., Bader, R.F.W., and Lau, C.D.H. (1987) *J. Am. Chem. Soc.*, **109**, 985.
- 12 Wiberg, K.B., Bader, R.F.W., and Lau, C.D.H. (1987) *J. Am. Chem. Soc.*, **109**, 1001.
- 13 Mandado, M., Vila, A., Graña, A.M., Mosquera, R.A., and Cioslowski, J. (2003) *Chem. Phys. Lett.*, **371**, 739.
- 14 Cortés-Guzmán, F. and Bader, R.F.W. (2003) *Chem. Phys. Lett.*, **379**, 183.
- 15 Bader, R.F.W., Popelier, P.L.A., and Keith, T.A. (1994) *Angew. Chem. Int. Ed. Engl.*, **33**, 620–631.
- 16 Vila, A. and Mosquera, R.A. (2007) *J. Comput. Chem.*, **28**, 1516–1530.
- 17 Cortés-Guzmán, F., Hernández-Trujillo, J., and Cuevas, G. (2003) *J. Phys. Chem. A*, **107**, 9253–9256.
- 18 Carroll, M.T., Chang, C., and Bader, R.F.W. (1988) *Mol. Phys.*, **63**, 387–405.

- 19 Carroll, M.T. and Bader, R.F.W. (1988) *Mol. Phys.*, **65**, 695–722.
- 20 Koch, U. and Popelier, P.L.A. (1995) *J. Phys. Chem.*, **99**, 9747–9754.
- 21 Grabowski, S. (ed.) (2006) *Hydrogen Bonding – New Insights*, Springer-Verlag.
- 22 Vila, A. and Mosquera, R.A. (2006) *J. Phys. Chem. A*, **110**, 11752–11759.
- 23 Bader, R.F.W. (1998) *J. Phys. Chem. A*, **102**, 7314–7323.
- 24 Zhao, Y. and Truhlar, D.G. (2005) *Phys. Chem. Chem. Phys.*, **7**, 2701–2705.
- 25 Zhao, Y. and Truhlar, D.G. (2005) *J. Phys. Chem. A*, **109**, 4209.
- 26 Zhao, Y. and Truhlar, D.G. (2005) *J. Phys. Chem. A*, **109**, 5656.
- 27 Zhao, Y., Schultz, N.E., and Truhlar, D.G. (2006) *J. Chem. Theory Comput.*, **2**, 364–382.
- 28 Zhao, Y. and Truhlar, D.G. (2006) *J. Phys. Chem. A*, **110**, 13126–13130.
- 29 Zhao, Y. and Truhlar, D.G. (2006) *Org. Lett.*, **8**, 5753–5755.
- 30 Zhao, Y. and Truhlar, D.G. (2006) *J. Chem. Phys.*, **125**, 194101.
- 31 Zhao, Y. and Truhlar, D.G. (2007) *J. Chem. Theory Comput.*, **3**, 289–300.
- 32 Zheng, J.J., Zhao, Y., and Truhlar, D.G. (2007) *J. Chem. Theory Comput.*, **3**, 569–582.
- 33 Popelier, P.L.A. (2000) *Atoms in Molecules: An Introduction*, Prentice Hall, Harlow.
- 34 Bader, R.F.W. (2005) *Monatsh. Chem.*, **136**, 819–854.
- 35 Frisch, M.J. *et al.* (2004) Gaussian 03, Revision C.02, pp. Gaussian, Inc, Wallingford CT.
- 36 Šponer, J., Jurečka, P., Marchan, I., Luque, F.J., Orozco, M., and Hobza, P. (2006) *Chem.–Eur. J.*, **12**, 2854–2865.
- 37 Bader, R.F.W. (1994) AIMPAC: A Suite of Programs for the Theory of Atoms in Molecules, McMaster University, Hamilton, Ontario, Canada.
- 38 Biegler-König, F.W., Schönbohm, J., and Bayles, D. (2001) *J. Comput. Chem.*, **22**, 545.
- 39 Graña, A.M. and Mosquera, R.A. (1999) *J. Chem. Phys.*, **110**, 6606–6616.
- 40 D’Souza, F. and Deviprasad, G.R. (2001) *J. Org. Chem.*, **66**, 4601.
- 41 Kurita, Y., Takayama, C., and Tanaka, S. (1994) *J. Comput. Chem.*, **15**, 1013.
- 42 González Moa, M.J., Mandado, M., and Mosquera, R.A. (2007) *J. Phys. Chem. A*, **111**, 1998–2001.
- 43 Sakurai, T. (1965) *Acta Crystallogr.*, **19**, 320.
- 44 Kuboyama, A. and Nagakura, S. (1955) *J. Am. Chem. Soc.*, **77**, 2644.
- 45 Tsuzuki, S., Honda, K., and Azumi, R. (2002) *J. Am. Chem. Soc.*, **124**, 12200.
- 46 Hobza, P., Šponer, J., and Reschel, T. (1995) *J. Comput. Chem.*, **16**, 1315.
- 47 Černý, J. and Hobza, P. (2005) *Phys. Chem. Chem. Phys.*, **7**, 1624.
- 48 Sinnokrot, M.O. and Sherrill, C.D. (2006) *J. Phys. Chem. A*, **110**, 10656.
- 49 Grimme, S. (2003) *J. Chem. Phys.*, **118**, 9095.
- 50 Becke, A.D. (1993) *J. Chem. Phys.*, **98**, 1372.
- 51 Matta, C.F., Castillo, N., and Boyd, R.J. (2006) *J. Phys. Chem. B*, **110**, 563–578.
- 52 Robertazzi, A. and Platts, J.A. (2006) *J. Phys. Chem. A*, **110**, 3992–4000.
- 53 Waller, M.P., Robertazzi, A., Platts, J.A., and Hibbs, D.E. (2006) *J. Comput. Chem.*, **27**, 491–504.
- 54 Mulliken, R.S. (1952) *J. Am. Chem. Soc.*, **74**, 811.
- 55 Hobza, P., Selzle, H.L., and Schlag, E.W. (1994) *J. Am. Chem. Soc.*, **116**, 3500.
- 56 Mandado, M., González Moa, M.J., and Mosquera, R.A. (2007) *J. Comput. Chem.*, **28**, 127–136.
- 57 Otero, N., González Moa, M.J., Mandado, M., and Mosquera, R.A. (2006) *Chem. Phys. Lett.*, **428**, 249.
- 58 Haslam, E. (1998) *Practical Polyphenolics: From Structure to Molecular Recognition and Physiological Action*, Cambridge University Press, Cambridge.
- 59 Martin, R., Lilley, T.H., Falshaw, C.P., Bailey, N.A., Haslam, E., Begley, M.J., and Magnolato, D. (1986) *J. Chem. Soc. Chem. Commun.*, 105–106.
- 60 Cai, Y., Martin, R., Lilley, T.H., Haslam, E., Gaffney, S.H., Spencer, C.M., and Magnolato, D. (1990) *J. Chem. Soc. Perkin Trans. 2*, 2197–2208.
- 61 Hobza, P. and Šponer, J. (1999) *Chem. Rev.*, **99**, 3247–3276.
- 62 Ghosh, A. and Bansal, M. (2003) *Acta Crystallogr. D*, **53**, 620–626.
- 63 Gorin, A.A., Zhurkin, V.B., and Olson, W.K. (1995) *J. Mol. Biol.*, **247**, 34–48.

- 64 Olson, W.K., Gorin, A.A., Lu, X.-J., Hock, L.M., and Zhurkin, V.B. (1998) *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 11163–11168.
- 65 Leslie, A.G., Arnott, S., Chandrasekaran, R., and Ratliff, R.L. (1980) *J. Mol. Biol.*, **143**, 49–72.
- 66 Dornberger, U., Flemming, J., and Fritzsche, H. (1998) *J. Mol. Biol.*, **284**, 1453–1463.
- 67 Leonard, G.A., McAuley-Hecht, K., Brown, T., and Hunter, W.N. (1995) *Acta Crystallogr. D*, **51**, 136–139.
- 68 Asensio, A., Kobko, N., and Dannenberg, J.J. (2003) *J. Phys. Chem. A*, **107**, 6441–6443.
- 69 Parthasarathi, R., Amutha, R., Subramanian, V., Nair, B.U., and Ramasami, T. (2004) *J. Phys. Chem. A*, **108**, 3817–3828.
- 70 Boyd, R.J. and Choi, S.C. (1986) *Chem. Phys. Lett.*, **129**, 62–65.
- 71 Rozas, I., Alkorta, I., and Elguero, J. (1998) *Chem. Soc. Rev.*, **27**, 163–170.
- 72 Domagala, M. and Grawoski, S. (2005) *J. Phys. Chem. A*, **109**, 5683–5688.
- 73 Tsuzuki, S. and Fujii, A. (2008) *Phys. Chem. Chem. Phys.*, **10**, 2584–2594.
- 74 Stone, A.J. (1993) *Chem. Phys. Lett.*, **211**, 101–109.
- 75 Tsuzuki, S., Honda, K., Uchimaru, T., Mikami, M., and Fujii, A. (2006) *J. Phys. Chem. A*, **110**, 10163.
- 76 Ringer, A.L., Figs, M.S., Sinnokrot, M.O., and Sherrill, C.D. (2006) *J. Phys. Chem. A*, **110**, 10822.
- 77 Tekin, A. and Jansen, G. (2007) *Phys. Chem. Chem. Phys.*, **9**, 1680.
- 78 Fujii, A., Shibasaki, K., Kazama, T., Itaya, R., Mikamia, N., and Tsuzuki, S. (2008) *Phys. Chem. Chem. Phys.*, **10**, 2836.
- 79 Vila, A. and Mosquera, R.A. (2006) *Int. J. Quantum Chem.*, **106**, 928–934.
- 80 Vila, A., Mosquera, R.A., and Hermida-Ramón, J.M. (2001) *J. Mol. Struct. (THEOCHEM)*, **541**, 149–158.
- 81 Mandado, M., Graña, A.M., and Mosquera, R.A. (2003) *Chem. Phys. Lett.*, **381**, 22–29.
- 82 Vila, A., Graña, A.M., and Mosquera, R.A. (2002) *Chem. Phys.*, **281**, 11–22.
- 83 López, J.L., Graña, A.M., and Mosquera, R.A. (2009) *J. Phys. Chem. A*, **113**, 2652–2657.

12

Polarizabilities of Amino Acids: Additive Models and *Ab Initio* Calculations

Noureddin El-Bakali Kassimi and Ajit J. Thakkar

12.1

Introduction

The polarizability of a molecule is one of its fundamental properties. It provides a measure of the volume and softness that can have qualitative uses in discussions of reactivity. The polarizability also has many quantitative uses such as the calculation of induction and dispersion coefficients for long-range interactions between molecules [1–3].

Polarizabilities can and have been measured for many molecules [4–6]. Moreover, advances in computer technology have greatly extended the reach of quantum chemistry. Polarizabilities can now be computed reliably for small and medium-sized molecules with well-established methods and widely available software [7, 8]. However, the same cannot be said for the large molecules that are often of interest in biochemistry. Hence, the development of simpler computational schemes remains important.

In this chapter we discuss additive models that allow the polarizability of a molecule to be estimated from the polarizabilities of its constituent atoms, bonds, functional groups or fragments in isolation. The long history of additive models of polarizabilities is sketched in Section 12.2, which also outlines some of the additive models that we consider later. The application of *ab initio* methods and additive models to the computation of isotropic polarizabilities of the 20 fundamental amino acids is described in Section 12.3. Some concluding remarks are made in Section 12.4. Atomic units (au) are used throughout. The correspondence between atomic units and SI units is given by $1 \text{ au of polarizability} = 1.64878 \times 10^{-41} \text{ C}^2 \text{ m}^2 \text{ J}^{-1}$.

12.2

Models of Polarizability

Attempts to express a molecular property as a weighted sum of transferable contributions from its constituent parts probably began with the mid-nineteenth century

work of Hermann Kopp [9–13]. He found that the molar volumes, and hence the molecular volumes and the closely related molecular polarizabilities, of organic liquids at their boiling points were close to additive functions of the molar volumes of their constituent elements. Moreover, Kopp observed that structural isomers had nearly identical molar volumes at their boiling points. A scholarly exposition of Kopp's work can be found in the Kopp Memorial Lecture delivered to the London Chemical Society in 1893 by Thorpe [14]. Experimental evidence to support nearly additive group contributions to molar refraction, a property even more closely related to the polarizability, was published later in the nineteenth century by Gladstone and Dale [15] and by Brühl [16–19]. Additivity was exploited as an important clue to chemical composition and later to structural assignments.

Early in the twentieth century, Eisenlohr [20] and then Silberstein [21] realized that the molecular refraction cannot be written simply as a sum of effective atomic refractions if atoms are defined solely by their atomic number. Instead, the environment of the atom must be taken into account. As Silberstein [21] stated, this was “the clearest confession of non-additivity.” The atomic environment can be accounted for by introducing different types of atoms of the same element, such as single-, double- and triple-bonded carbon atoms. Eisenlohr [20, 22] and, later, Vogel [23] worked out additive schemes for molar refraction using this approach. A different way of taking the atomic environment into account is by writing the molecular polarizability (refraction) as a sum of bond polarizabilities (refractions) as in the early to mid-twentieth century work of von Steiger [24], Smyth [25, 26], Denbigh [27], Vickery and Denbigh [28] and Vogel *et al.* [29, 30]. Yet another approach was developed extensively by Vogel [23], who represented different types of bonding with a set of prototypical groups that can be interpreted with atomic hybrids. William Shockley [31], Roberts [32] and Tessman *et al.* [33] applied additive schemes to the polarizabilities of crystals using atomic ions as the constituent units. The early work was all based on empirical analysis of experimental values, although efforts were made to understand additivity of polarizabilities using variational perturbation theory [1]. LeFèvre gives an extensive account of the additivity work done prior to 1963 in his fine review of molecular refractivity and polarizability [34].

The failure of certain optical rotation calculations led Applequist *et al.* [35] to consider a model of molecular polarizabilities in which the interactions between the induced dipoles in the atoms are explicitly accounted for. He applied the atomic dipole interaction model (ADIM), which dates back to early twentieth century theories of optical rotation formulated by Max Born [36], Oseen [37] and Gray [38]. ADIM was first considered for polarizabilities by Silberstein [21, 39], and elaborated in greater detail by Rowell and Stein [40], and by Mortensen [41]. DeVoe [42] and Birge [43] made the atomic point dipoles anisotropic. Olson and Sundberg [44] extended the ADIM model to account for charge transfer in molecules with delocalized π -electrons. Applequist [45] subsequently applied their model to aliphatic and aromatic hydrocarbons. Thole [46] improved the ADIM model by replacing the point dipole interaction by an interaction between smeared out dipoles.

Inspired by the analysis of Hirschfelder *et al.* [1], Miller and Savchik [47] put forward a very successful empirical model in which the polarizability of a molecule with N electrons is expressed as $(4/N)$ times the square of a sum of atomic hybrid components (ahcs). The overall parameterization requires a hybrid component for each hybridization (valence) state of each element. Kang and Jhon [48] showed that very similar results could be obtained by approximating the molecular polarizability as a sum of polarizabilities of atom types or atomic hybrid polarizabilities (ahp) – one for each hybridization state of each element. In a later study, Miller [49] compared the ahc and ahp models. He found the ahc model gave a slightly better fit to roughly 400 experimental polarizabilities but he based his model of anisotropic polarizabilities [50] on the ahp model for the isotropic part. Miller [50] showed how Vogel's group polarizabilities [23] could be factored into a set of ahp's because Vogel's units coincided with atoms in the usual hybridization states. No *et al.* [51] introduced the charge dependence of the effective atomic polarizability (CDEAP) model as an improvement of the ahp model. They made the atomic hybrid polarizabilities depend explicitly upon net atomic charges calculated with their modified partial equalization of orbital electronegativity (M-PEOE) method [52–55].

All the models mentioned so far were parameterized by fits to experimental polarizabilities. In this work, we apply six such models to the amino acids. These include Miller's ahc model, his parameterization of the ahp model, and a simplified version of the CDEAP model in which averaged values of the net atomic charges were used for each type of atomic hybrid. We refer to these three models as M90, KJM90 and NCJS93, respectively. The other three empirical models we use are described next. We use an unpublished model of Goedhart, referred to here as G69, in which the additive units are functional groups. It was parameterized by a fit to about a thousand liquid organic compounds containing 43 different functional groups. G69 includes unique "constitutional" corrections for steric hindrance. It has been stated [56] that Goedhart presented this work at an international seminar on gel permeation chromatography held in Monaco in October 1969. We are unaware of any published report by Goedhart on this topic and used his RLL parameterization as listed in Table 10.1 of a book on polymer properties [56]. We also use the model of Bosque and Sales [57] and refer to it as BS02. They probed the limits of a simple additive model without any explicit accounting for the environment of an atom. Introducing an unphysical constant term in their model, they obtained a fit to the experimental polarizabilities of 340 liquids with a mean absolute percent deviation (MAPD) of 2.3%, and found a MAPD of 1.93% on their test set of 86 liquids. The sixth purely empirical model we use is model 2E of Wang *et al.* [58]. We refer to this model as WXHX07. Wang *et al.* [58] extended Bosque and Sales' model by reintroducing the dependence of the effective atomic polarizabilities upon the hybridization state and, not surprisingly, concluded that the MAPD could be reduced by almost a factor of two (1.24%).

Clearly, it is possible to fit additivity models to *ab initio* polarizabilities as well. For example, in a series of papers, Doerksen, Kassimi, and Thakkar computed 2nd-order Møller–Plesset (MP2) polarizabilities for azoles [59], azines [60], oxazoles [61],

azaborinines [62], azaboroles and oxazaboroles [63]. In each of their papers they fit atom- and bond-additive models to all the polarizabilities computed in that and previous papers in the series. In the last paper [63], Doerksen and Thakkar reported fits to MP2 polarizabilities of 104 planar, five- and six-membered heteroaromatic rings. Dykstra and colleagues [64–66] published three parameterizations, denoted here as SD95 [64], SD98 [65] and ZD00 [66], that differ only in the number of atom types for which parameter values were obtained and the data to which they were fitted. The SD95 and ZD00 parameters were fitted to Hartree–Fock (HF) polarizabilities for 30 and 58 small molecules, respectively, at idealized geometries. The SD98 parameters were fitted to MP2 polarizabilities for the same 30 molecules as those used to parameterize SD95. A mixture of experimental and theoretical polarizabilities was used by van Duijnen and Swart [67] to re-parameterize and extend Thole’s model. Voisin and Cartier [68] parameterized the models of Thole [46] and Miller [50] using MP2 polarizabilities for 20 small molecules [68, 69]. Ewig and coworkers [70] fit an additive model, referred to as EWM02 in this work, in which the effective atomic polarizability depends on its environment via bond increments. They used a training set of HF polarizabilities for 30 carefully chosen organic molecules, and reported scale factors that effectively correct for electron correlation.

Kassimi and Thakkar [71] took an approach akin to one they had used for relating the polarizability of purine to its fragments [72]. They cleave a molecule AB into two suitable fragments A and B, cap both fragments with a hydrogen atom to form the AH and BH molecules, and compute the polarizability of AB as the sum of the polarizabilities of the capped fragments minus twice the polarizability of a capping hydrogen. They called that procedure the “hydrogen elimination” (HE) model. The fragments can be capped with methyl groups instead of hydrogen atoms. In the latter case, their model is called the “methyl elimination” (ME) model. If the partitioning does not lead to fragments that are small enough, then the large fragments can themselves be decomposed into smaller ones in the same manner. Kassimi and Thakkar [71] calculated the fragment polarizabilities at the MP2 level but emphasized that any suitable *ab initio* or experimental method could have been used instead.

Much work has been devoted to the a posteriori decomposition of *ab initio* molecular properties into contributions from constituent atoms, bonds and functional groups. The partitioned quantities need to be transferable from molecule to molecule if they are to be of use in additivity schemes. Bader and others have considered polarizabilities and their additivity [73–77] from the perspective afforded by Bader’s theory of atoms in molecules [78]. Bader and Bayles [79] point out that “transferability of a group and its properties is in general, only apparent, being the result of compensatory transferability wherein the changes in the properties of one group are compensated for by equal but opposite changes in the properties of the adjoining group.” Other polarizability decomposition methods include those of Karlström and colleagues [80–82], Stone and coworkers [83–87] and others [88, 89]. However, the practical implications of polarizability partitioning methods for additivity models have been rather limited so far [90].

12.3

Polarizabilities of the Amino Acids

Only a few experimental studies of the polarizabilities of the 20 naturally occurring amino acids have been reported. The refractive indices of amino acid crystals were measured by Lacourt and Delande [91, 92] for use in microthermal identification. The values of the molar refraction measured by McMeekin and colleagues [93, 94] at a wavelength of $\lambda = 578$ nm in aqueous solution lead, via the Lorentz–Lorenz relationship [1, 2], to the experimental values of the polarizability listed in Table 12.1. Solubility difficulties prevented McMeekin *et al.* from measuring the molar refractions for aspartic acid (Asp), cysteine (Cys) and tyrosine (Tyr). The molar refractions that they list for Asp and Tyr are estimates obtained by subtracting the molar

Table 12.1 Polarizabilities of the amino acids from *ab initio* computations and experiment.

Acid	MP2 ^{a)}	DFT ^{b)}	DFT(ω) ^{c)}	DFT(Z) ^{d)}	EST ^{e)}	Exptl ^{f)}
Ala	55.26	55.98	57.44	58.09	58.83	55.86 $\pm$ 0.7
Arg	120.16	121.71	125.59	124.58	126.91	115.57 $\pm$ 0.2
Asn	78.33	78.98	81.24	82.05	83.66	79.77 $\pm$ 0.7
Asp	72.71	73.43	75.33	76.92	78.1	(76.35)
Cys	75.80	76.76	79.10	80.34	81.72	(74.96)
Gln	90.80	91.85	94.44	94.19	95.73	91.22 $\pm$ 0.6
Glu	85.51	86.66	88.95	89.48	90.62	90.42 $\pm$ 0.4
Gly	43.09	43.57	44.70	47.13	47.78	44.25 $\pm$ 0.6
His	101.01	102.40	105.93	105.87	108.01	102.59 $\pm$ 0.4
Ile	91.24	92.67	95.04	94.88	95.82	95.24 $\pm$ 0.6
Leu	91.98	93.47	95.90	96.00	96.94	94.49 $\pm$ 0.4
Lys	102.37	104.18	107.04	106.96	108.01	101.20 $\pm$ 0.5
Met	102.22	103.81	106.99	106.56	108.15	102.14 $\pm$ 0.1
Phe	122.39	123.64	128.57	127.82	131.5	122.90 $\pm$ 0.3
Pro	74.18	75.19	77.13	76.82	77.75	73.49 $\pm$ 0.4
Ser	59.90	60.69	62.27	63.86	64.65	61.24 $\pm$ 0.4
Thr	71.56	72.56	74.37	75.47	76.28	73.70 $\pm$ 0.4
Trp	155.77	157.49	165.42	161.72	167.93	157.76 $\pm$ 0.5
Tyr	128.26	129.76	135.17	134.43	138.34	(128.60)
Val	79.03	80.19	82.26	82.31	83.22	81.49 $\pm$ 0.4

a) Static ($\omega = 0$) MP2/aug-cc-pVDZ//B97-1/cc-pVDZ polarizability for conformer F1, Reference [102].

b) Static ($\omega = 0$) B97-1/aug-cc-pVDZ//B97-1/cc-pVDZ polarizability for conformer F1, Reference [102].

c) Dynamic B97-1/aug-cc-pVDZ//B97-1/cc-pVDZ polarizability at $\lambda = 578$ nm ($\omega = 0.0788$ au) for conformer F1, Reference [102].

d) Static ($\omega = 0$) B97-1/aug-cc-pVDZ//B97-1/cc-pVDZ polarizability for the zwitterion structure, Reference [102].

e) Estimated dynamic polarizability at $\lambda = 578$ nm ($\omega = 0.0788$ au) for the zwitterion structure. Computed from $EST = MP2(F1) + [DFT(\omega) - DFT] + [DFT(Z) - DFT]$.

f) Dynamic polarizability at $\lambda = 578$ nm extracted from the experimental molar refractions given in References [93, 94]. Values in parentheses are estimates; see text.

refraction of glycyl residue from the molar refractions of glycyl aspartate and glycyl tyrosine, respectively. Their molar refraction for Cys was obtained from effective atomic molar refractions [93, 94]. More recently [95], the refractive indices of alanine (Ala), proline (Pro) and valine (Val) were measured in aqueous solution at three different wavelengths using interferometric methods. Building on earlier work by Orttung and Meyer [96], Khanarian and Moore [97] measured the Kerr effect of amino acids in water. By combining their results with those of McMeekin *et al.* [93, 94], Khanarian and Moore were able to extract the polarizability anisotropies as well.

There have been a few *ab initio* calculations of the polarizabilities of the 20 proteinogenic amino acids. Voisin and Cartier [68] reported a MP2 calculation of the polarizability for glycine (Gly). Tulip and Clark [98] calculated the polarizability tensors of alanine (Ala), leucine (Leu), isoleucine (Ile) and valine (Val) using density functional perturbation theory implemented within the plane wave pseudopotential framework. Swart *et al.* [99] used time-dependent (TD) density functional theory (DFT) to compute molecular polarizabilities of the residues of the 20 amino acids. Hansen *et al.* [100] calculated the frequency-dependent polarizabilities of the 20 amino acids at the HF level. Guthmuller and Simon [101] reported TDDFT computations of the frequency-dependent polarizabilities of tryptophan (Trp), tyrosine (Tyr) and phenylalanine (Phe). Most recently, Millefiori *et al.* [102] reported DFT and MP2 computations of the static polarizabilities of all 20 fundamental amino acids. We focus on their work because it is the most comprehensive and probably the most accurate as well.

Millefiori *et al.* [102] began by optimization of the geometries of the two lowest-energy, neutral conformers, F1 and F2, of each of the 20 amino acids at the B97-1/cc-pVDZ level [103]. They also optimized zwitterionic (Z) structures in aqueous solution at the same level using the conductor-like polarizable continuum model (C-PCM) [104] for the solvent. Then they calculated static polarizabilities at the F1 and F2 geometries using the HF, MP2 and B97-1 methods with an aug-cc-pVDZ basis set. The MAPD between their MP2(F1) and MP2(F2) results is merely 0.4%, and the MAPD between their B97-1(F1) and B97-1(F2) results is only 0.6%. Evidently, conformational effects are not important as far as the static isotropic polarizability is concerned. Hence, only their MP2(F1) and B97-1(F1) results are listed in Table 12.1 as MP2 and DFT, respectively. Table 12.1 and Figure 12.1 show that the B97-1 polarizabilities are consistently larger than their presumably more reliable MP2 counterparts but only by an average of 1.3 and 0.9%, respectively, for the F1 and F2 conformers. Thus B97-1 should be adequate to obtain an estimate of other, smaller effects on the polarizability. Table 12.1 lists, as $\text{DFT}(\omega)$, Millefiori *et al.*'s B97-1/aug-cc-pVDZ polarizabilities calculated for the F1 conformer at a wavelength of $\lambda = 578$ nm to match experiment. Table 12.1 and Figure 12.1 show that the $\text{DFT}(\omega)$ polarizability is consistently larger than its zero-frequency counterpart $\text{DFT}(\text{F1})$ by an average of 3.0% and a maximum of 5.0% for tryptophan (Trp). Table 12.1 lists, as $\text{DFT}(\text{Z})$, Millefiori *et al.*'s B97-1/aug-cc-pVDZ static polarizabilities calculated at the zwitterionic structures. Table 12.1 and Figure 12.1 show that the $\text{DFT}(\text{Z})$ polarizability is consistently larger than $\text{DFT}(\text{F1})$ by an average of 3.6% and a maximum of 8.2% for glycine (Gly). We then define an estimated polarizability for $\lambda = 578$ nm at

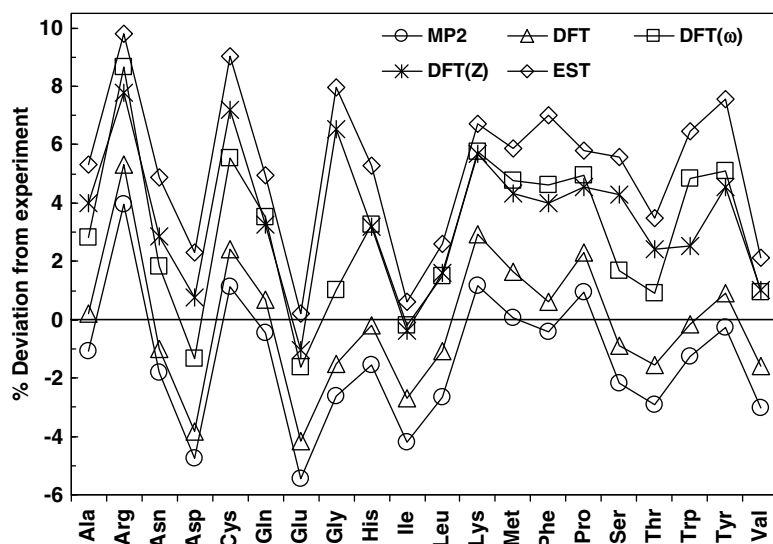


Figure 12.1 Comparison of *ab initio* polarizabilities with the experimental data of McMeekin *et al.* The labels of the various methods are defined in the footnotes to Table 12.1.

the zwitterionic geometry by $EST = MP2(F1) + [DFT(\omega) - DFT] + [DFT(Z) - DFT]$ in which all quantities on the right-hand side are from Millefiori *et al.* [102]. This estimate is also listed in Table 12.1.

We now turn to a comparison of *ab initio* polarizabilities with experiment. Figure 12.1 and Table 12.1 show that, of all the *ab initio* values, the static MP2 polarizabilities are closest to the experimental values of McMeekin *et al.* [93, 94] with a MAPD of 2.1%. The largest differences between MP2 and experiment are 5.4, 4.8 and 4.2% for glutamic acid (Glu), aspartic acid (Asp) and isoleucine (Ile), respectively. Unfortunately, this relatively good agreement must be fortuitous since the MP2 values are for infinite wavelength and the gas-phase whereas the experimental values are for $\lambda = 578$ nm and aqueous solution. Moreover, the best estimate (EST) that accounts for the effects of non-zero frequency and the zwitterionic structure expected in aqueous solution is consistently larger than the experimental values and has a significantly larger MAPD of 5.2%, with a maximum difference of 9.8% for arginine (Arg). As concluded earlier [71, 104], this unsatisfactory situation remains unresolved. We note that de Hemsy *et al.* [95] reported measured values of 56.33, 76.34 and 81.21 for alanine (Ala), proline (Pro) and valine (Val), respectively. Their value for proline is noticeably closer to EST than McMeekin *et al.*'s value is to EST [93, 94].

Next, we consider five additive models of polarizability that are based on *ab initio* calculations. Table 12.2 lists the predictions of the Stout–Dykstra [64] model (SD95), the Zhou–Dykstra [66] model (ZD00), the model of Ewig and coworkers [70] (EWM02), and the hydrogen elimination (HE) and methyl elimination (ME) models of Kassimi and Thakkar [71]. Ewig *et al.* had recommended different scale factors to

Table 12.2 Static polarizabilities of the amino acids from additive models based on *ab initio* calculations.

Acid	HE(19) ^{a)}	HE(17) ^{b)}	ME(15) ^{c)}	ME(13) ^{d)}	EWM02 ^{e)}	ZD00 ^{f)}	SD95 ^{g)}
Ala	55.47	56.02	55.47	56.28	56.29	53.51	56.34
Arg	118.31	118.86	119.26	120.08		117.73	120.97
Asn	79.15	79.55	79.21	80.03	77.82	78.77	79.97
Asp	73.62	74.17	73.51	74.33	72.51	73.52	76.31
Cys	75.70	76.10	76.02	76.84	77.93		
Gln	90.99	91.54	91.57	92.39	90.76	91.07	92.62
Glu	85.29	85.84	85.87	86.69	85.45	85.81	88.96
Gly	43.63	44.03	43.63	43.92	43.35	41.21	43.70
His	100.53	101.08	101.45	102.26			106.30
Ile	92.18	92.58	92.39	93.20	95.11	90.40	94.29
Leu	91.66	92.21	92.39	93.20	95.11	90.40	94.29
Lys	100.58	101.13	101.75	102.57	104.00	99.31	105.37
Met	100.35	100.90	100.94	101.76	104.37		
Phe	121.36	121.91					121.43
Pro	74.70	75.25			76.84	79.25	81.64
Ser	60.07	60.47	60.22	61.04	59.99	57.01	61.39
Thr	72.37	72.77	72.27	73.09	72.93	69.30	74.04
Trp	154.42	154.97					151.83
Tyr	126.91	127.46					126.47
Val	79.82	80.22	80.04	80.85	82.17	78.10	81.64

a) Hydrogen elimination model based on MP2 calculations for 19 fragments [71].

b) Hydrogen elimination model based on MP2 calculations for 17 fragments [71].

c) Methyl elimination model based on MP2 calculations for 15 fragments, Reference [71] and unpublished work. MP2 polarizability used for 4-methyl-1*H*-imidazole is 62.50.

d) Methyl elimination model based on MP2 calculations for 13 fragments, Reference [71] and unpublished work.

e) Calculated from the additive model of Ewig *et al.* [70], with a scale factor of 1.17 optimized in this work for the amino acids.

f) Calculated in Reference [71] from the additive model of Zhou and Dykstra [66].

g) Calculated in Reference [71] from the additive model of Stout and Dykstra [64].

be applied to their model depending on the functional group that characterized a molecule. However, the amino acids have more than one functional group and none of the scale factors is directly applicable. We used a scale factor of 1.17 because it seems to work well for amino acids. Table 12.2 lists two versions of the HE and ME models [71] that differ in the number of molecular fragments used to construct the amino acids; for example, HE(19) was based on 19 fragments. As seen in Table 12.2, there was enough data to apply the HE, ME, EWM02, ZD00 and SD95 models to 20, 16, 15, 14 and 18 amino acids, respectively.

Now we compare the above models with experiment. Figure 12.2 and Table 12.2 show that the ME and HE models are closest to the experimental values of McMeekin *et al.*, with a MAPD ranging between 1.5 and 1.9%. Fortunately, the HE and ME models agree better with experiment than any of the fully *ab initio* calculations of

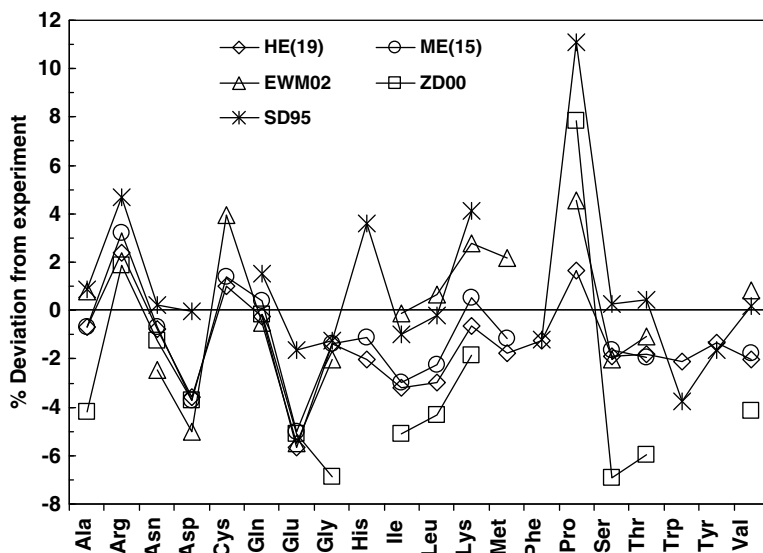


Figure 12.2 Comparison of polarizabilities predicted by additive models based on *ab initio* data with the experimental data of McMeekin *et al.* The labels of the various models are defined in the footnotes to Table 12.2.

Millefiori *et al.* [102]. The SD95 and EWM02 models are not far behind with a MAPD of 2.1 and 2.3%, respectively. The ZD00 model is significantly different with a MAPD of 4.2%.

Since all these models are based on *ab initio* static polarizabilities, it is perhaps more relevant to compare them with the MP2 static polarizabilities of Millefiori *et al.* [102]. The HE and ME models are in the best agreement with the MP2 values with a MAPD of 0.8, 0.9, 1.1 and 1.5% for ME(15), HE(19), HE(17) and ME(13), respectively. Both the HE and ME methods [71] mimic the fully *ab initio* methods as well as could be expected. Keeping in mind that ME was applied to only 16 molecules and that the fragments needed for ME are larger than those needed for HE, we think that HE is to be preferred over ME. The MAPD between MP2 and the EWM02, ZD00 and SD95 models is 1.8, 2.4 and 3.1%, respectively. The HE and ME models perform significantly better at least in part because the fragments are tailored to the problem under consideration.

Next, we consider six additive models based on fits to experimental data. Table 12.3 lists the polarizabilities predicted by the WXHX07, BS02, NCJS93, M90, KJM90 and G69 models. We could not find Goedhart's parameters [56] for aromatic nitrogens, primary amide groups and imine groups; perhaps they do not exist. In any case, we could apply G69 to only 15 amino acids. Since the experimental data used in the fits are usually measured at a finite wavelength, agreement with the experimental polarizabilities of McMeekin *et al.* [93, 94] is a fair test. Figure 12.3 provides a visual comparison. All six models show significant

Table 12.3 Polarizabilities of the amino acids from additive models based on experimental data.

Acid	WXHX07 ^{a)}	BS02 ^{b)}	NCJS93 ^{c)}	M90 ^{d)}	KJM90 ^{e)}	G69 ^{f)}
Ala	57.33	55.54	57.32	56.80	56.37	55.72
Arg	120.24	115.40	119.38	114.45	119.76	
Asn	80.12	77.81	79.69	78.11	79.30	
Asp	74.71	73.42	75.10	75.02	73.63	72.35
Cys	79.69	75.72	79.70	77.39	76.61	76.40
Gln	92.61	90.29	91.79	90.28	91.68	
Glu	87.19	85.91	87.20	87.14	86.01	84.79
Gly	44.84	43.05	45.22	44.60	43.99	43.38
His	103.07	102.57	103.41	100.32	103.27	
Ile	94.78	92.99	93.62	93.71	93.52	92.61
Leu	94.78	92.99	93.62	93.71	93.52	93.11
Lys	104.50	101.22	103.71	103.29	102.64	102.02
Met	104.66	100.68	103.90	101.99	101.38	101.78
Phe	123.48	121.27	123.96	121.47	121.56	120.90
Pro	75.51	78.21	76.27	76.60	75.91	75.87
Ser	61.63	59.39	62.23	61.54	60.67	59.56
Thr	74.12	71.87	74.33	73.70	73.05	71.65
Trp	150.74	149.88	153.02	154.30	157.75	
Tyr	127.79	125.11	130.69	129.05	129.84	125.69
Val	82.29	80.51	81.52	81.37	81.14	80.36

a) Calculated from Model 2E of Wang *et al.* [58].

b) Calculated from the model of Bosque and Sales [57].

c) Calculated from the CDEAP model of No *et al.* using averaged net atomic charges [51].

d) Calculated from the ahc model of Miller, Reference [47] with parameters from Reference [49].

e) Calculated from the ahp model of Kang and Jhon [48], with parameters from Miller [49].

f) Calculated from the unpublished model of Goedhart, as reported in Table 10.1 of Reference [56]. Constitutional corrections were included.

discrepancies for glutamic acid (Glu) and proline (Pro). They all predict a polarizability for proline that is closer to the measurement of de Hemsey and coworkers [95] than to the value of McMeekin *et al.* The MAPDs with respect to McMeekin *et al.*'s experimental values are 1.5% for both M90 and KJM90, 2.0% for WXHX07 and NCJS93, 2.2% for BS02 and 2.3% for G69. Miller's parameterizations [49] of M90 and KJM90 seem to be the most accurate.

12.4

Concluding Remarks

Figure 12.4 is a concise summary of how well the existing *ab initio* calculations and additive models do for the polarizabilities of the amino acids. Figure 12.4 and the results in Section 12.3 show quite clearly that both empirical and *ab initio* additive models are very useful as reasonably accurate predictors of polarizabilities for the amino acids.

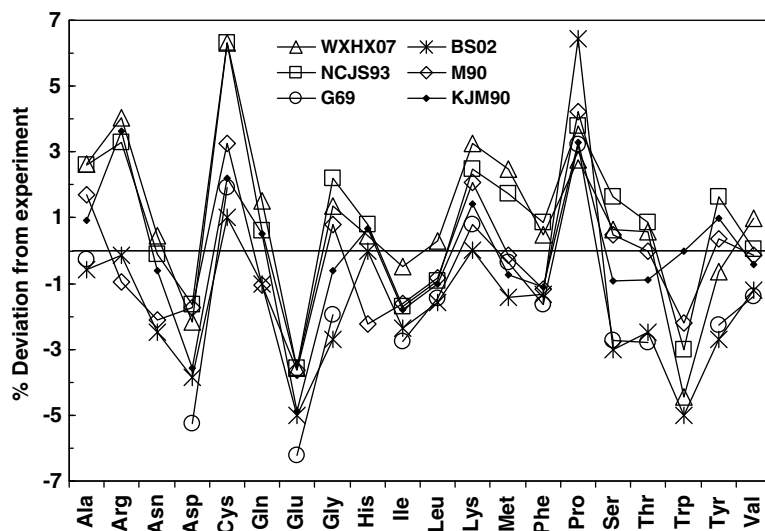


Figure 12.3 Comparison of polarizabilities predicted by additive models based on experimental data with the experimental data of McMeekin *et al.* The labels of the various models are defined in the footnotes to Table 12.3.

However, the current state of fully *ab initio* calculations of polarizabilities for the 20 fundamental amino acids is unsatisfactory. More work needs to be done to understand why the static polarizabilities are in such good agreement with the experimental values whereas the best estimates including the effects of a non-zero

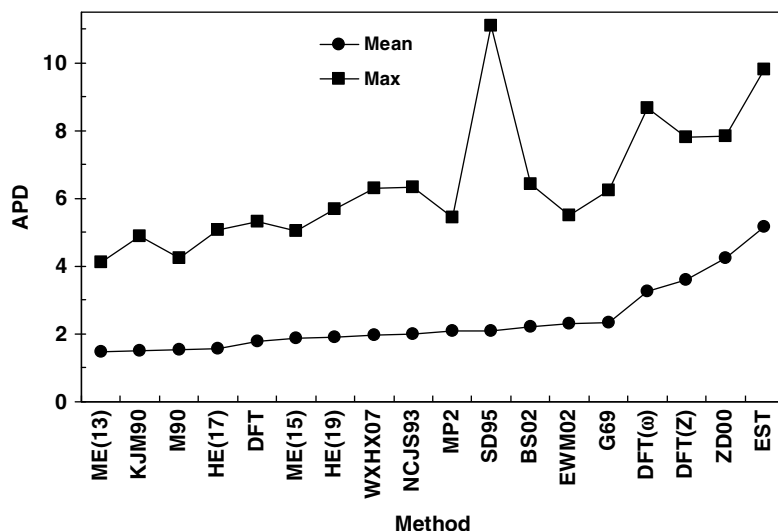


Figure 12.4 Absolute percent deviation (APD) of polarizabilities with respect to the experimental values of McMeekin *et al.*

frequency and the zwitterionic structure expected in aqueous solution are not in agreement with experiment. The anisotropy of the polarizability tensors for the amino acids is a challenge for the future.

There are so many additive models of polarizabilities in the literature that we could mention only a selected subset within the space available. We tried to present a representative selection of the models. No disrespect is meant to the authors of the models that have not been mentioned.

Acknowledgments

This chapter is dedicated to the memory of David M. Bishop who contributed so much to the theoretical study of polarizabilities. AJT enjoyed many discussions with him over the years. The Natural Sciences and Engineering Research Council of Canada supported this work.

References

- Hirschfelder, J.O., Curtiss, C.F., and Bird, R.B. (1954) *Molecular Theory of Gases and Liquids*, John Wiley & Sons, Inc., New York.
- Bonin, K.D. and Kresin, V.V. (1997) *Electric-Dipole Polarizabilities of Atoms, Molecules and Clusters*, World Scientific, Singapore.
- Thakkar, A.J. (2001) Intermolecular interactions, in *Encyclopedia of Chemical Physics and Physical Chemistry* (Vol. I. Fundamentals) (eds J. Moore and N. Spencer), Institute of Physics Publishing, Bristol.
- Miller, T.M. and Bederson, B. (1977) *Adv. At. Mol. Phys.*, **13**, 1–55.
- Miller, T.M. and Bederson, B. (1988) *Adv. At. Mol. Phys.*, **25**, 37–60.
- Gould, H. and Miller, T.M. (2005) *Adv. At. Mol. Phys.*, **51**, 343–361.
- Dykstra, C.E. (1988) *Ab Initio Calculation of the Structures and Properties of Molecules*, Elsevier, Amsterdam.
- Maroulis, G. (ed.) (2006) *Atoms, Molecules and Clusters in Electric Fields: Theoretical Approaches to the Calculation of Electric Polarizability*, Imperial College Press, Oxford, UK.
- Kopp, H. (1839) *Poggendorfs Ann. Phys. Chem.*, **123**, 133–153.
- Kopp, H. (1842) *Ann. Chem. Pharm.*, **41**, 79–89.
- Kopp, H. (1842) *Ann. Chem. Pharm.*, **41**, 169–189.
- Kopp, H. (1855) *Ann. Chem. Pharm.*, **96**, 1–36.
- Kopp, H. (1855) *Ann. Chem. Pharm.*, **96**, 153–185.
- Thorpe, T.E. (1893) *J. Chem. Soc. Trans.*, **63**, 775–815.
- Gladstone, J.H. and Dale, T.P. (1863) *Philos. Trans. R. Soc. London*, **153**, 317–343.
- Brühl, J.W. (1880) *Justus Liebigs Ann. Chem.*, **200**, 139–231.
- Brühl, J.W. (1880) *Justus Liebigs Ann. Chem.*, **203**, 1–63.
- Brühl, J.W. (1880) *Justus Liebigs Ann. Chem.*, **203**, 255–285.
- Brühl, J.W. (1880) *Justus Liebigs Ann. Chem.*, **203**, 363–368.
- Eisenlohr, F. (1910) *Z. Phys. Chem. (Leipzig)*, **75**, 585–607.
- Silberstein, L. (1917) *Philos. Mag.*, **33**, 92–128.
- Eisenlohr, F. (1912) *Z. Phys. Chem. (Leipzig)*, **79**, 129–146.
- Vogel, A.I. (1948) *J. Chem. Soc.*, 1833–1855.
- von Steiger, A.L. (1921) *Ber. Dtsch. Chem. Ges.*, **54**, 1381–1393.
- Smyth, C.P. (1925) *Philos. Mag.*, **50**, 361–375.
- Smyth, C.P. (1925) *Philos. Mag.*, **50**, 715.

- 27 Denbigh, K.G. (1940) *Trans. Faraday Soc.*, **36**, 936–947.
- 28 Vickery, B.C. and Denbigh, K.G. (1949) *Trans. Faraday Soc.*, **45**, 61–81.
- 29 Vogel, A.I., Cresswell, W.T., Jeffery, G.J., and Leicester, J. (1950) *Chem. Ind.*, p. 358.
- 30 Vogel, A.I., Cresswell, W.T., Jeffery, G.H., and Leicester, J. (1952) *J. Chem. Soc.*, 514–549.
- 31 Shockley, W. (1946) *Phys. Rev.*, **70**, 105.
- 32 Roberts, S. (1949) *Phys. Rev.*, **76**, 1215–1220.
- 33 Tessman, J.R., Kahn, A.H., and Shockley, W. (1953) *Phys. Rev.*, **92**, 890–895.
- 34 LeFèvre, R.J.W. (1965) *Adv. Phys. Org. Chem.*, **3**, 1–90.
- 35 Applequist, J., Carl, J.R., and Fung, K.K. (1972) *J. Am. Chem. Soc.*, **94**, 2952–2960.
- 36 Born, M. (1915) *Phys. Z.*, **16**, 251–258.
- 37 Oseen, C.W. (1915) *Ann. Phys.*, **48**, 1–56.
- 38 Gray, F. (1916) *Phys. Rev.*, **7**, 472–488.
- 39 Silberstein, L. (1917) *Philos. Mag.*, **33**, 521–533.
- 40 Rowell, R.L. and Stein, R.S. (1967) *J. Chem. Phys.*, **47**, 2985–2989.
- 41 Mortensen, E.M. (1968) *J. Chem. Phys.*, **49**, 3732–3733.
- 42 DeVoe, H. (1965) *J. Chem. Phys.*, **43**, 3199–3208.
- 43 Birge, R.R. (1980) *J. Chem. Phys.*, **72**, 5312–5319.
- 44 Olson, M.L. and Sundberg, K.R. (1978) *J. Chem. Phys.*, **69**, 5400–5404.
- 45 Applequist, J. (1993) *J. Phys. Chem.*, **97**, 6016–6023.
- 46 Thole, B.T. (1981) *Chem. Phys.*, **59**, 341–350.
- 47 Miller, K.J. and Savchik, J.A. (1979) *J. Am. Chem. Soc.*, **101**, 7206–7213.
- 48 Kang, Y.K. and Jhon, M.S. (1982) *Theor. Chim. Acta*, **61**, 41–48.
- 49 Miller, K.J. (1990) *J. Am. Chem. Soc.*, **112**, 8533–8542.
- 50 Miller, K.J. (1990) *J. Am. Chem. Soc.*, **112**, 8543–8551.
- 51 No, K.T., Cho, K.H., Jhon, M.S., and Scheraga, H.A. (1993) *J. Am. Chem. Soc.*, **115**, 2005–2014.
- 52 No, K.T., Grant, J.A., and Scheraga, H.A. (1990) *J. Phys. Chem.*, **94**, 4732–4739.
- 53 No, K.T., Grant, J.A., Jhon, M.S., and Scheraga, H.A. (1990) *J. Phys. Chem.*, **94**, 4740–4746.
- 54 Park, J.M., No, K.T., Jhon, M.S., and Scheraga, H.A. (1993) *J. Comput. Chem.*, **14**, 1482–1490.
- 55 Park, J.M., Kwon, O.Y., No, K.T., Jhon, M.S., and Scheraga, H. (1995) *J. Comput. Chem.*, **16**, 1011–1026.
- 56 van Krevelen, D.W. and Hoftyzer, P.J. (1976) *Properties of Polymers: Their Estimation and Correlation with Chemical Structure*, 2nd edn, Elsevier, Amsterdam.
- 57 Bosque, R. and Sales, J. (2002) *J. Chem. Inf. Comput. Sci.*, **42**, 1154–1163.
- 58 Wang, J.M., Xie, X.Q., Hou, T.J., and Xu, X.J. (2007) *J. Phys. Chem. A*, **111**, 4443–4448.
- 59 Kassimi, N.E.-B., Doerksen, R.J., and Thakkar, A.J. (1995) *J. Phys. Chem.*, **99**, 12790–12796.
- 60 Doerksen, R.J. and Thakkar, A.J. (1996) *Int. J. Quantum Chem.*, **60**, 1633–1642.
- 61 Kassimi, N.E.-B., Doerksen, R.J., and Thakkar, A.J. (1996) *J. Phys. Chem.*, **100**, 8752–8757.
- 62 Doerksen, R.J. and Thakkar, A.J. (1998) *J. Phys. Chem. A*, **102**, 4679–4686.
- 63 Doerksen, R.J. and Thakkar, A.J. (1999) *J. Phys. Chem. A*, **103**, 2141–2151.
- 64 Stout, J.M. and Dykstra, C.E. (1995) *J. Am. Chem. Soc.*, **117**, 5127–5132.
- 65 Stout, J.M. and Dykstra, C.E. (1998) *J. Phys. Chem. A*, **102**, 1576–1582.
- 66 Zhou, T. and Dykstra, C.E. (2000) *J. Phys. Chem. A*, **104**, 2204–2210.
- 67 van Duijnen, P.T. and Swart, M. (1998) *J. Phys. Chem. A*, **102**, 2399–2407.
- 68 Voisin, C. and Cartier, A. (1993) *J. Mol. Struct. (THEOCHEM)*, **105**, 35–45.
- 69 Voisin, C., Cartier, A., and Rivail, J.L. (1992) *J. Phys. Chem.*, **96**, 7966–7971.
- 70 Ewig, C.S., Waldman, M., and Maple, J.R. (2002) *J. Phys. Chem. A*, **106**, 326–334.
- 71 Kassimi, N.E.-B. and Thakkar, A.J. (2009) *Chem. Phys. Lett.*, **472**, 232–236.
- 72 Kassimi, N.E.-B. and Thakkar, A.J. (1996) *J. Mol. Struct. (THEOCHEM)*, **366**, 185–193.
- 73 Bader, R.F.W. (1989) *J. Chem. Phys.*, **91**, 6989–7001.
- 74 Laidig, K.E. and Bader, R.F.W. (1990) *J. Chem. Phys.*, **93**, 7213–7224.
- 75 Bader, R.F.W., Keith, T.A., Gough, K.M., and Laidig, K.E. (1992) *Mol. Phys.*, **75**, 1167–1189.

- 76 Stone, A.J., Hattig, C., Jansen, G., and Angyan, J.G. (1996) *Mol. Phys.*, **89**, 595–605.
- 77 Arturo, S.G. and Knox, D.E. (2006) *J. Mol. Struct. (THEOCHEM)*, **770**, 31–44.
- 78 Bader, R.F.W. (1990) *Atoms in Molecules: A Quantum Theory*, Oxford University Press, Oxford.
- 79 Bader, R.F.W. and Bayles, D. (2000) *J. Phys. Chem. A*, **104**, 5579–5589.
- 80 Karlström, G. (1982) *Theor. Chim. Acta*, **60**, 535–541.
- 81 Gagliardi, L., Lindh, R., and Karlström, G. (2004) *J. Chem. Phys.*, **121**, 4494–4500.
- 82 Söderhjelm, P., Krogh, J.W., Karlström, G., Ryde, U., and Lindh, R. (2007) *J. Comput. Chem.*, **28**, 1083–1090.
- 83 Stone, A.J. (1985) *Mol. Phys.*, **56**, 1065–1082.
- 84 Lesueur, C.R. and Stone, A.J. (1993) *Mol. Phys.*, **78**, 1267–1291.
- 85 Lesueur, C.R. and Stone, A.J. (1994) *Mol. Phys.*, **83**, 293–307.
- 86 Williams, G.J. and Stone, A.J. (2004) *Mol. Phys.*, **102**, 985–991.
- 87 Misquitta, A.J. and Stone, A.J. (2006) *J. Chem. Phys.*, **124**, 024111.
- 88 Ferraro, M.B., Caputo, M.C., and Lazzarotti, P. (1998) *J. Chem. Phys.*, **109**, 2987–2993.
- 89 Lillestolen, T.C. and Wheatley, R.J. (2007) *J. Phys. Chem. A*, **111**, 11141–11146.
- 90 Rick, S.W. and Stuart, S.J. (2002) *Rev. Comput. Chem.*, **18**, 89–146.
- 91 Lacourt, A. and Delande, N. (1962) *Mikrochim. Acta*, **50**, 48–54.
- 92 Lacourt, A. and Delande, N. (1964) *Mikrochim. Acta*, **52**, 547–560.
- 93 McMeekin, T.L., Groves, M.L., and Wilensky, M. (1962) *Biochem. Biophys. Res. Commun.*, **7**, 151–156.
- 94 McMeekin, T.L., Groves, M.L., and Hipp, N.J. (1964) Refractive indices of amino acids, proteins, and related substances, in *Amino Acids and Serum Proteins*, vol. 44, Advances in Chemistry, American Chemical Society, Washington, D.C. 54–66.
- 95 de Hemsy, M.E.B., de Molina, M.A.A., Miñano, A.S.M., and Lobo, P.W. (1976) *Anal. Asoc. Quím. Argentina*, **64**, 105–114.
- 96 Orttung, W.H. and Meyers, J.A. (1963) *J. Phys. Chem.*, **67**, 1911–1915.
- 97 Khanarian, G. and Moore, W.J. (1980) *Aust. J. Chem.*, **33**, 1727–1741.
- 98 Tulip, P.R. and Clark, S.J. (2004) *J. Chem. Phys.*, **121**, 5201–5210.
- 99 Swart, M., Snijders, J.G., and van Duijnen, P.Th. (2004) *J. Comput. Methods Sci. Eng.*, **4**, 419–425.
- 100 Hansen, T., Jensen, L., Astrand, P.O., and Mikkelsen, K.V. (2005) *J. Chem. Theory Comput.*, **1**, 626–633.
- 101 Guthmuller, J. and Simon, D. (2006) *J. Phys. Chem. A*, **110**, 9967–9973.
- 102 Millefiori, S., Alparone, A., Millefiori, A., and Vanella, A. (2008) *Biophys. Chem.*, **132**, 139–147.
- 103 Cramer, C.J. (2004) *Essentials of Computational Chemistry: Theories and Models*, 2nd edn, John Wiley & Sons, Inc., Hoboken.
- 104 Cossi, M., Reggi, N., Scalmani, G., and Barone, V. (2003) *J. Comput. Chem.*, **24**, 669–681.

13

Methods in Biocomputational Chemistry: A Lesson from the Amino Acids

Hugo J. Bohórquez, Constanza Cárdenas, Chérif F. Matta, Russell J. Boyd,
and Manuel E. Patarroyo

13.1

Introduction

Computer-aided drug design (CADD) requires accurate and fast methods to identify and characterize molecules with potential therapeutic use. While quantum mechanics (QM) provides the best available theoretical framework to predict molecular properties, it is computationally expensive for biologically-relevant molecules, molecules that are usually composed of hundreds of atoms such as proteins and nucleic acids. This practical limitation dictates the use of approximate methods that are fast enough to screen large sets of biochemical compounds.¹⁾ Ideally, these methods are designed to identify molecules with a specific biological activity *in silico*. Predictive methods employed in drug design, such as statistical analysis (SA) and molecular mechanics (MM), address different levels of detail of the molecular problem. Statistical methods, based mainly on database records, are designed to provide averaged molecular properties such as secondary-structure propensities or the hydrophobic character of a polypeptide chain. Molecular mechanics (MM) methods provide information about specific functional groups and their interactions in terms of Newtonian (classical) mechanics through force fields parameterized for a given class of biomolecules. This parameterization is based on the results of quantum mechanical computations. Hence, QM plays an *indirect* but crucial role in approximate MM biocomputational methods. In more recent years, and when a specific reactive center is known, one can combine QM and MM in a single calculation in what has become known as QM/MM (and its variants). (See Chapters 2–4, and the literature cited therein.)

Here we explain three strategies developed for studying peptides that include statistical analysis over quantum mechanical data to characterize amino acids'

1) See for example the methods reviewed in the first four chapters of this book: Quantum crystallography (Chapter 1), ONIOM and QM/MM (Chapters 2 and 3), and the continuum methods of solvating large molecules (Chapter 4).

functional similarities and activity in proteins. These strategies were developed in stages, each one focusing on a different aspect of the problem.

The first stage addresses the question: which theoretical variables describe the conformational trends of the amino acids best? This question underlies the structure–activity relationship (SAR) paradigm, according to which similar structures yield similar bioactivity [1]. Hence, from this standpoint, it would be advantageous to select the parameters that optimally describe similarity in the construction of quantitative structure–activity relationship (QSAR) models. A principal result from this research is that electrostatic variables sufficiently discriminate structural and chemical trends [2]. This work is summarized in Section 13.2.

The second stage of the study examines a smaller set of capped amino acid (AA) models $[(\text{HC}=\text{O})-\text{AA}-\text{NH}_2]$, in two conformations, that is, 40 molecules in total. The quantum theory of atoms in molecules (QTAIM) was used to analyze the resulting electron densities. The variables studied are the electronic energy and the multipole moments (polarizations) of the amino acids side chains. This set of variables defines a ten-dimensional space. The similarities in this 40-molecule/ten-variable system are determined in two ways. The first method is graphical, based on a multidimensional projection known as the Andrews plot [3]. The second is an unbiased pairing method (neighbor joining) that determines similarities on the basis of the distance between the vectors representing each side chain. Remarkably, this procedure can replicate the standard biochemical classification of the genetically-encoded amino acids, providing a quantum theoretical classification of amino acids [4], the first to our knowledge. Section 13.3 provides details about this work and its future extensions.

The third stage of this research illustrates the practical application of the previously mentioned findings through a method that incorporates the electrostatic variables for the study of peptide–host interactions [5]. Section 13.4 illustrates the advantages of using a Mulliken multipole-based approach to the study of MHC–antigenic peptide complexes. Comments about the strengths and future directions of this approach conclude this chapter.

13.2

Conformers, Rotamers and Physicochemical Variables

The number of possible molecules formed from a given set of atoms is determined by the combinatorial number of allowed stable bonding interactions between these atoms. If we count amino-acid based penta-peptides, for example, the number of possible molecular structures is $\sim 20^5$, that is, 3.2×10^6 molecules. This number is based on the 20 genetically-encoded amino acids only, a subset of the about 300 amino acids found in living systems, excluding unnatural amino acids. Not surprisingly, the idea of drug design appears to be a hopeless quest. How can we effectively reduce such diversity to a manageable set that, eventually, will display the desired drug properties?

A first step consists of selecting a set of variables that can be obtained consistently for every molecule. Each molecule is then represented by a vector whose components are the selected variables. Each variable should be well-defined and, at least in principle, also be a measurable property. Each molecule is represented by its respective set of properties in the multidimensional space, that is, by a vector $V_A \subset R^N$. The representation of every molecule in this multidimensional vector space, R^N , enables one to define a Euclidian distance, d_{AB} , between two molecules A and B. Ideally, two molecules separated by the shortest distance in this vector space are also the most chemically similar among the set. This hypothesis is based on the realization that similar molecules must exhibit similar molecular physicochemical properties. In this approach, molecular design implies the identification of similarities in this vector space.

The biochemical behavior of a protein is encoded in its primary structure, that is, the amino acid sequence that determines its functionality via the secondary and tertiary structures. In the study of the genetically encoded amino acids is important to determine which variables account for their idiosyncratic biochemical features.

Within the context of protein-based drug design, the following question is addressed: what theoretical variables better represent the highly specific yet overlapping biochemical functions displayed by each of the genetically-encoded amino acids? To answer this question, two models have been built that mimic the electronic environment of an amino acid residue inside a peptide chain. The models differ in the capping groups for the N- and the C-terminuses (Figure 13.1). The non-zwitterionic amino acid models studied are $H(C=O)|AA|NH_2$ and $Ala|AA|Ala$. The second model allows the determination of the effect of a neighboring amino acid on the properties of the central amino acid.

Each side chain has preferred side chain torsion angles χ_1 [6–8]. Three of these side chain conformers or rotamers were selected: gauche(+) = -66.7° , gauche(–) = 64.1° and trans = 183.68° . The main chain conformers were set at five α -helical and five β -sheet conformations (defined by ψ and ϕ angles as shown in

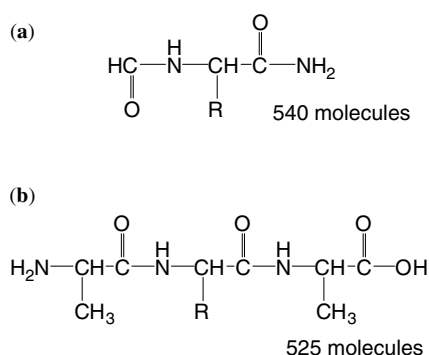


Figure 13.1 Capped amino acid residues used in this study. Each molecule was represented by the variables listed in Table 13.1. The total number of molecules studied using each model is indicated.

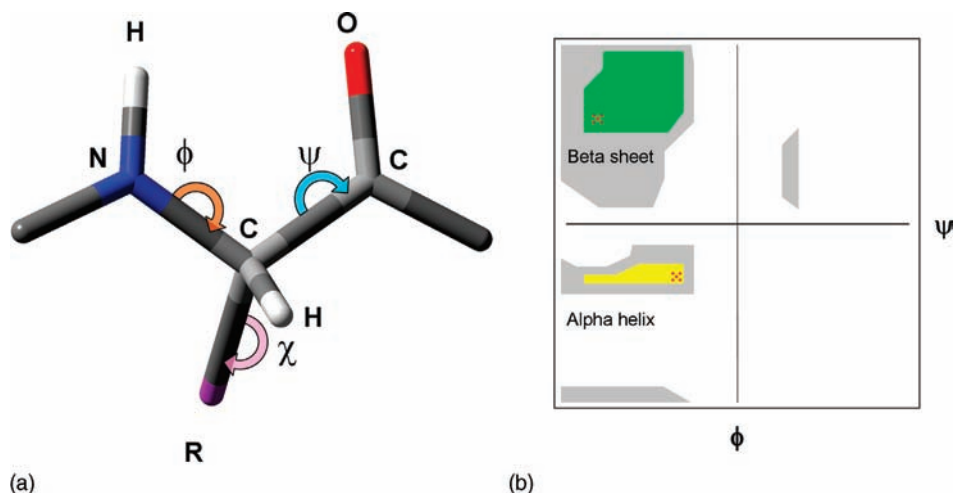


Figure 13.2 (a) Stick model of an amino acid residue with the standard dihedral angles defining the main chain (ϕ , ψ) and the side chain (χ) conformations; (b) Ramachandran plot with the studied α -helical and β -sheet conformation regions highlighted (red dots).

Figure 13.2). The five backbone torsion angles corresponding to an α -helical conformation are $\phi = -65^\circ + b4^\circ$ and $\psi = -39^\circ + b4^\circ$, and those corresponding to the β -strand conformations are $\phi = -130^\circ + b5^\circ$ and $\psi = 120^\circ + b5^\circ$, with b (0, ± 1). These conformations are indicated by the red dots in Figure 13.2b.

A total of 40 theoretical variables were considered: (i) 19 graph descriptors, that is, connectivity descriptors of the molecular structure [9, 10] and (ii) 21 physicochemical variables obtained from quantum mechanical calculations. Table 13.1 lists the variables representing the amino acids. A total of 1065 molecules representing the 20 amino acids in different conformations and capping models were studied. The graph-theory indices were calculated with Codessa [11, 12], and the QM computations were performed at the HF/6-31G(d) level, with polarization functions on heavy atoms capable of forming hydrogen bonds (N, O and S).

A hierarchical cluster analysis was carried out with the NTSYS program [13], with the unweighted pair group method with arithmetic mean (UPGMA) method. Figure 13.3 depicts a schematic representation of the steps followed in this strategy. Principal components analysis (PCA) [14] was used to determine the principal variables that specify the similarity between the amino acids. An important result is that the amino acids are separated into statistically disjoint groups, and those groups are segregated mainly by their electrostatic properties alone. Figure 13.4 shows the classification obtained from the PCA. The group of amino acids containing π electrons is clearly identifiable, which includes the aromatic amino acids, such as Phe, Tyr and Trp as well as His and Arg. Two amino acids, Gly and Pro, are clearly outliers in this classification, which reflects their particular biochemical behavior: The first has the smallest side chain (a hydrogen atom) and the second is an imino acid, that is, its side chain is cycled over the backbone. The analysis was able to

Table 13.1 Properties selected for representation of the amino acids.

Graph theory indices	Quantum variables
Wiener index	Moment of inertia
Randic indices of order 0–3	Molecular weight of the amino acid residue
Kier and Hall connectivity indices of order 0–3	Electronic spatial extent
Kier and Hall shape indices of order 1 through 3	Nuclear repulsion energy
Kier flexibility index	Total energy
Shadow indices	Highest occupied molecular orbital energy, HOMO
	Lowest unoccupied molecular orbital energy, LUMO
	Mulliken partial charges
	Sum of the Mulliken partial charges for the side chain atoms
	Total dipole moment
	Electric potential
	Sum of the electric potential for the side chain atoms
	Quadrupole norm

automatically distinguish between the two large groups differing only in the conformation of the backbone, namely, the α - and β -conformers of all of the amino acids separate into two large statistically-distinguishable groups, as can be seen in Figure 13.2.

In general, side chain rotamers are grouped closely to each other according to their respective amino acid. Rotamers of isoelectronic side chains, such as the Asn-Asp or Gln-Glu isoelectronic pairs, are also located closely in the 40-dimensional vector space, which indicates that the method generates valid results according to intrinsic (total) molecular properties, but misses details on the specific smaller functional groups.

The groups obtained for each amino acid conformer are conserved across the two capping groups, which indicates that the selected variables capture the intrinsic nature of amino acid properties. These groups can be reproduced by eight variables only (five quantum, three from graph theory), as indicated by the PCA analysis.

In conclusion, the classification of the set of amino acids studied is driven principally by electrostatic properties. Clearly, from the clustering shown in Figure 13.4 the main groups are those side chains containing electrons with π symmetry, and the two backbone conformations, α and β , whose multipole moments are oriented in different directions, which means, therefore, that they have different electrostatic interactions.

The results reviewed here show that the structural features of amino acids are sufficiently accounted for by electrostatic variables alone. Some quantum QSARs come to a similar conclusion. For example, Brinck *et al.* have studied approx. 100 QM-based variables to predict the water–octanol partition coefficient ($P_{o/w}$) from the molecular wavefunctions. These authors concluded that three electrostatic variables – surface area, the surface electrostatic potential and the spatial minima of the

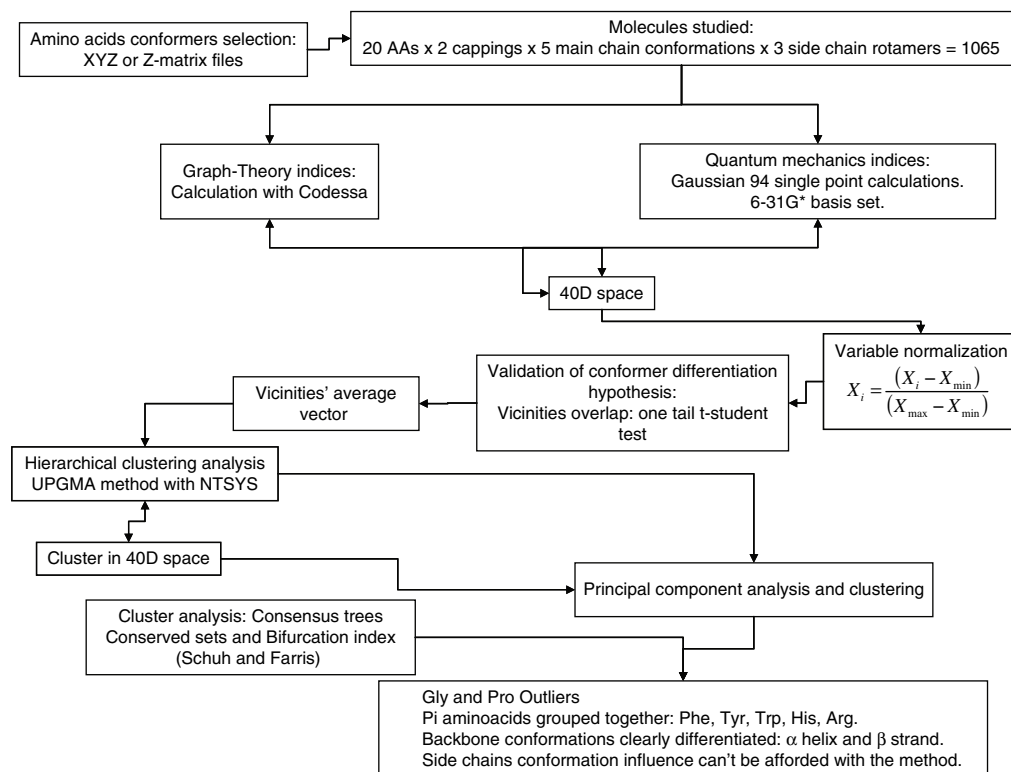


Figure 13.3 Schematic diagram of the steps followed to determine the theoretical variables responsible for the main structural propensities of the amino acids. The key factor in this

approach is the representation of every molecule by a set of properties in a multidimensional (40D) vector space for performing a PCA and a clustering analysis.

electrostatic potential – can give good correlations with $\log P_{o/w}$ for several molecules with biological and pharmacological interest [15]. This is a remarkable result in the sense that $\log P_{o/w}$ is an experimentally determined biochemical property usually measured at standard conditions yet the thermodynamic factors are not included in the quantum computation. These results as well as those reviewed in this chapter suggest that the electrostatic variables are good descriptors regardless of the thermodynamic conditions, providing support to the validity of the isolated (“gas phase”) QM model.

13.3

QTAIM Side Chain Polarizations and the Theoretical Classification of Amino Acids

Clearly, from the results reviewed earlier, the amino acids can be adequately described in terms of the electrostatic variables. In this section we describe the use of QTAIM for characterizing the genetically-encoded amino acids.

QTAIM partitions the molecular properties into additive atomic contributions and, in doing so, allows the characterization of molecular transferable fragments such as the amino acids' side chains. We omit details about the theory here as they can be found elsewhere [16]. This section will focus on the similarity of the amino acids within the context of this theory. The earlier work of Bader and coworkers on peptides and amino acids has been extended in greater detail by Matta and Bader more recently [17–19]. Chapter 14 of this book provides a review of this latter work.

The local structural properties of each amino acid determine the overall tertiary structure of a protein, and the side chains are responsible for its specific bioactivity. Therefore, it is vital to characterize the physicochemical properties of the amino acid side chains to understand and predict the bioactivity of peptides and proteins. The amino acid model studied is shown in Figure 13.1a, which was initially studied by Bader [20–22]. Two main backbone angles and a single rotamer per amino acid were studied, giving a total of 40 molecules, including a total of 888 atoms at the HF/6-31G (d) level with polarization functions on N, O and S. The atomic properties were computed with the AIMPAC suite of routines from Bader's group [23]. To compare tensor and vector properties (which are origin- and orientation-dependent) all the amino acids were properly aligned by the common atoms of the backbone and the first atom in the side chain. The origin of the coordinates was placed at the α -carbon atom.

Each amino acid was represented by the three first terms in the multipole moment expansion of the side chain charge density [side chain charges (monopoles), and side chain dipolar and quadrupolar polarizations] and the side chain total electronic energy. These electronic multipole moments (polarizations) should not be confused with the total multipole moments of the amino acids. Multipole moments provide a basis for a general procedure to systematically extract the symmetries of a continuous distribution, such as the charge density, and hence they characterize its shape. They depend on the origins and the relative orientations of the coordinate system and therefore the molecules were pre-aligned as described above.

The energy of the side chains measures their size (Figure 13.5). Figure 13.5 shows that the side chain energy magnitude can be linearly fitted to the side chain mass with a correlation $R^2 = 0.95$, for all the non-sulfur side chains. The electronic energy of the side chains involving only elements located in the first two rows of the periodic table exhibit a linear correlation with mass; the same correlation does not apply for side chains involving a third row atom, such as sulfur.

It is desirable to visualize similarities between the molecules under study *before* performing any further statistical survey, but any multi-dimensional molecular representation always entails a graphical challenge. Andrews plots (APs) are a useful tool for addressing this task. As illustrated in Figure 13.6, each molecule can be represented by a single strand, which is easily obtainable from the following formula:

$$g(r) = \frac{1}{\sqrt{2}} \left\{ \begin{aligned} &E + Mx[\sin(t) + \cos(t)] + My[\sin(2t) + \cos(2t)] + Mz[\sin(3t) + \cos(4t)] + \\ &+ Q_{xx}[\sin(5t) + \cos(5t)] + Q_{xy}[\sin(6t) + \cos(6t)] + Q_{xz}[\sin(7t) + \cos(7t)] \\ &+ Q_{yz}[\sin(8t) + \cos(8t)] + Q_{zz}[\sin(9t) + \cos(9t)] \end{aligned} \right\} \quad (13.1)$$

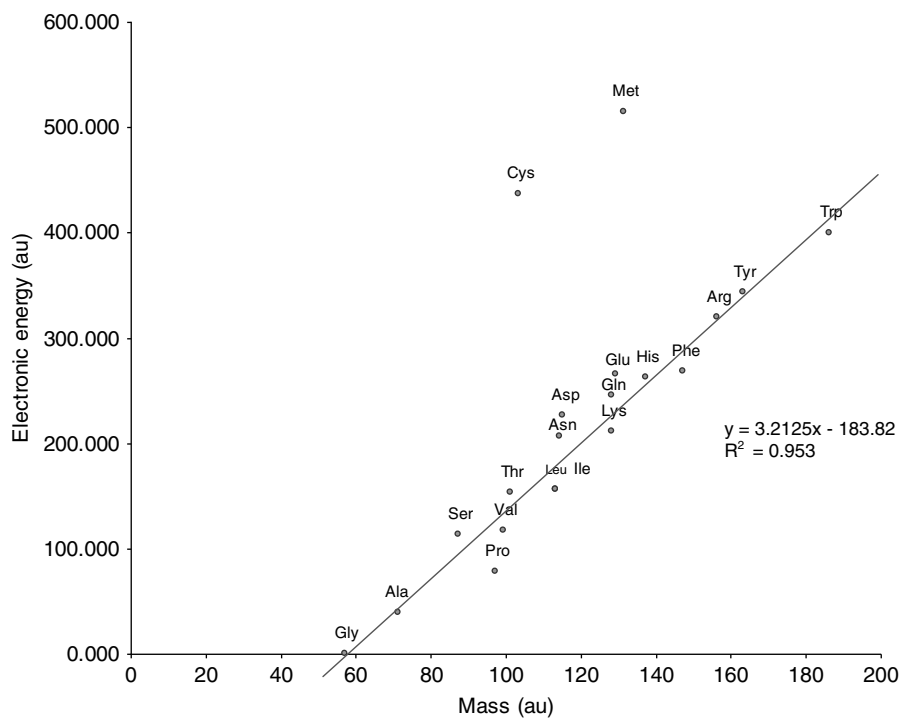


Figure 13.5 Energy magnitude versus mass, as provided by QTAIM, for the genetically-encoded amino acid side chains. The linear fit excludes the sulfur-containing side chains, Cys and Met.

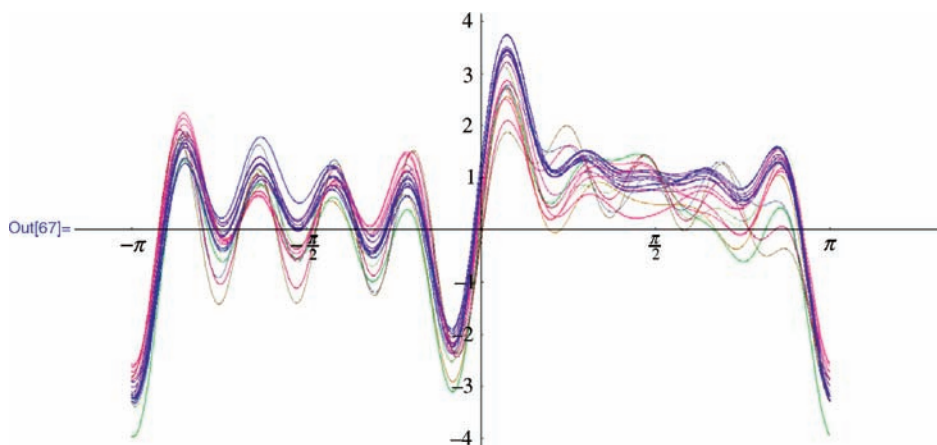


Figure 13.6 Andrews plots for the ten QTAIM variables on 40 amino acid side chains. Molecular similarities appear as similar colors and shapes.

where E is the side chain energy, M_i and Q_{ij} are the dipole and quadrupole polarization components, respectively, and $t \in [-\pi, \pi]$. The values used were standardized as explained in detail in Reference [4].

Each strand represents a side chain as a smooth function, with coefficients equal to the corresponding physical properties. We also added a color code to each strand by assigning each component of the color code to the (standardized) magnitudes of the energy, dipole and quadrupole moment. The final color is a combination of three basic tones: red, green and blue. Each tone is defined as a number within the interval $[0, 1]$. For the present case, we choose $RGB = [1 - M, Q, E]$, where M , Q and E are the normalized magnitudes of the dipolar polarization, quadrupolar polarization and energy, respectively (i.e., each of these variables lies within the interval $[0, 1]$). Therefore, we can visually identify similar shapes and colors that correspond to similar molecules. Figure 13.7 shows the APs of the 40 side chains studied. The distinctive shape in blue groups the aliphatic side chains (Gly, Ala, Pro, Val, Ile, Leu), while the group (Asn, Gln, Asp, Glu) exhibit a similar red color. This simple analysis reveals the existence of underlying similarities within the set of amino acids.

The graphical analysis shows the existence of similarities between the side chains, but to quantitatively determine these similarities a systematic classification procedure is required. Consequently, we used a multivariate classification of the side chains in the 10D vector space that is based on the distance between elements in this vector space. The neighbor joining method applied over a twofold distance measure provides the amino acids classification shown in Figure 13.8. Clearly, the main biochemical features coincide with several of the groups obtained.

This theoretical classification of the amino acids, the first quantum theoretical classification we are aware of, provides a rich variety of clearly identifiable biochemical groups on the sole basis of transferable properties provided by QTAIM. In contrast, experimentally-based classifications tend to emphasize certain molecular features and downplay others, which explains why the classification resulting from their associated matrices coincides with the biochemical classification only for major groups such as “aliphatic AAs” or “charged AA,” while several amino acids appear as outliers [24–26], as recently reported by Esteve and Falceto [27].

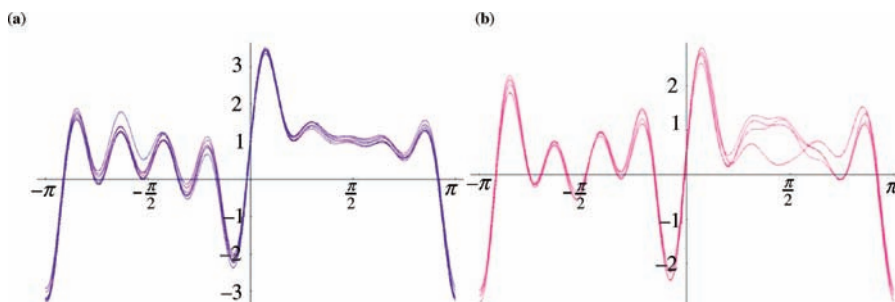


Figure 13.7 Similar side chains as revealed by their corresponding Andrews plots (color and shape). (a) Gly, Ala, Val, Ile and Leu; (b) Asn, Gln, Asp and Glu; the later exhibits a different pattern than the others at $t = \pi/2$.

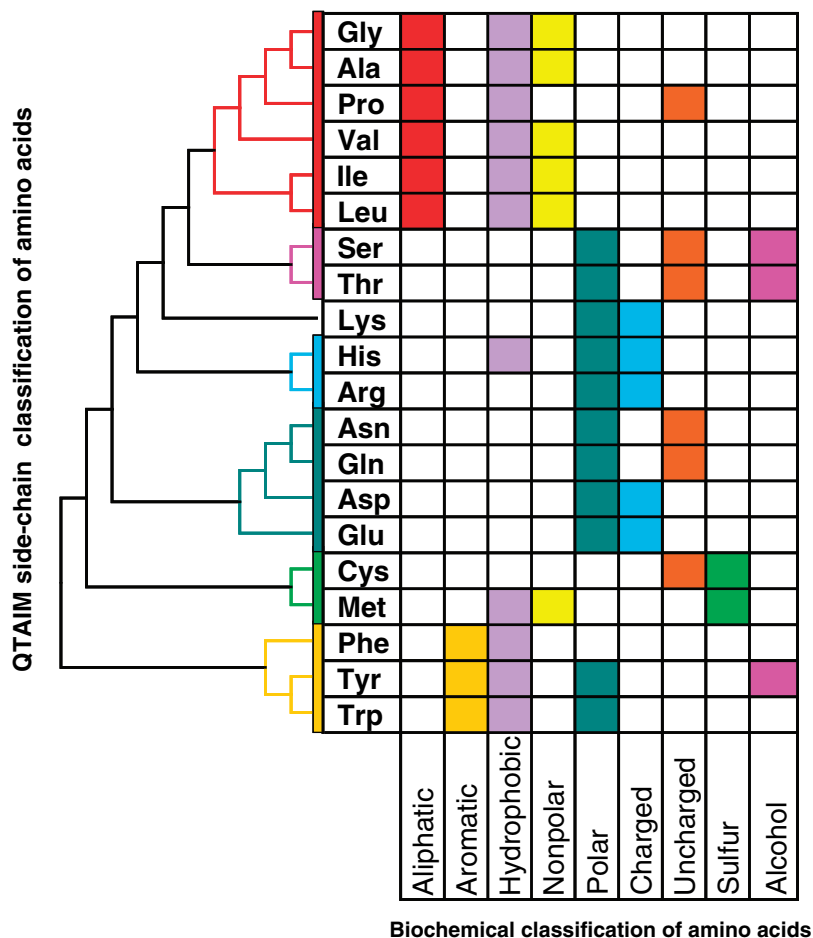


Figure 13.8 Quantum theoretical classification of genetically-encoded amino acids. This classification was obtained after applying a clustering procedure to the side chain properties: energy, dipolar polarization and quadrupolar polarization as provided by QTAIM computations at HF/6-31G(d) level of theory. The table highlights the typical physicochemical properties of the side chains. The main clusters were colored according to these properties.

We attribute the successful classification of the amino acids *in silico* to the quality of the atomic and group properties provided by the quantum theory of atoms in molecules. We have shown above how one can use QTAIM group properties in conjunction with clustering analysis to recover a well-known biochemical classification of a set of functionally-related molecules (the amino acids). Amino acid classification based on the electrostatic moments is superior to those obtained by scoring matrices widely used in protein biostatistics. One key advantage of the theoretically-based classification over experimentally-based ones is the homogeneity

of the quality of the input dataset. Experimentally-based amino acid properties that serve as a basis for the replacement matrices, for example, involve various data sources with different precisions, compromising the outcome of the analysis. As an extension of this work, we are developing a theoretical amino acid replacement matrix for bioinformatics that will potentially overcome several of the drawbacks faced by the empirical ones.

The methodology outlined here can be replicated for any other set of molecules, and it emerges as an alternative to QSAR methods in the sense that it provides unbiased quantitative similarities among the studied set, indicating potential replacements among them.

13.4

Quantum Mechanical Studies of Peptide–Host Interactions

In the previous section we have shown that calculated QM electrostatic properties provide a biochemical classification of the amino acids consistent with their known chemical and physical properties. The main hypothesis for the study of peptide–host interactions is that these interactions are, to a large extent, dominated by these electrostatic properties of the AA residues constituting the interacting peptides. This research program is motivated by the development of a synthetic anti-malarial vaccine at the Fundación Instituto de Inmunología de Colombia (FIDIC) [28, 29].

A key step for developing a specific immune response against a pathogen is the formation of a stable complex between the major histocompatibility complex (MHC) molecule and antigenic peptides, capable of bringing information to the T-cell receptor (TCR) molecules necessary to trigger an immune response against the pathogen. We omit the details of this process, to focus on the QM/MM hybrid approach used for the peptide–host interaction studies for design of a synthetic peptide-based anti-malarial vaccine. The MHC–peptides (MHC-P) interaction is a prototypical ligand–receptor interaction, and hence the approach outlined here can be used to study other similar biochemical complexes. The extended peptide (~9 amino acids) forms a noncovalent complex with the host MHC protein at the peptide binding region (PBR) through certain spots that act as anchoring sites, known as *pockets*. Figure 13.9 shows a MHC class II PBR [30], with the pockets in color, as obtained from the Protein Data Bank (PDB).

According to our hypothesis, the MHC-P interaction can be described by the quantum-based electrostatic potential, which in terms of the multipole expansion has the form:

$$V = \frac{1}{4\pi\epsilon_0} \left[\sum_k \frac{q_k}{r} + \sum_k p_k \delta_k \frac{1}{r^2} + \frac{1}{2} \sum_{ij} \frac{1}{3} Q_{ij} \delta_i \delta_j \frac{1}{r^3} + \dots \right] \quad (13.2)$$

where the index k runs over all the host atoms involved in the interaction. Therefore, a partitioning scheme is necessary to provide atomic contributions for each multipole moment that appears in this expansion. Unfortunately, the number of atoms involved

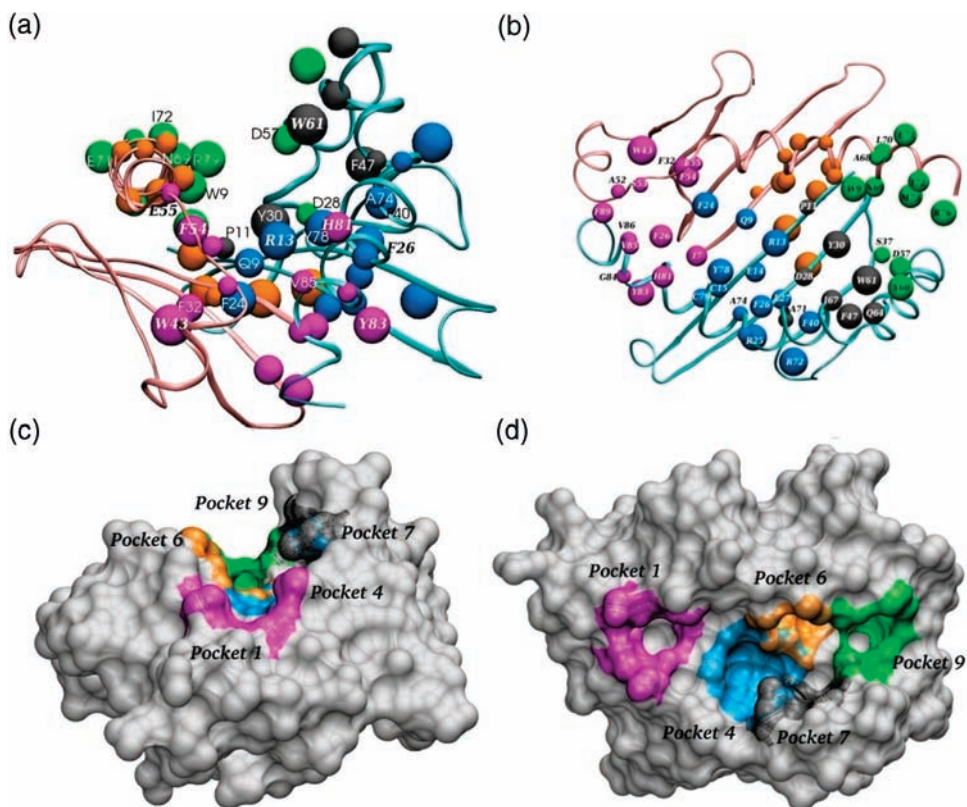


Figure 13.9 Peptide binding region (PBR) of the major histocompatibility complex class II (MHCII-P); LA-DR β 1*1501 molecule with the α -chain as a pink ribbon and the β -chain as a light blue ribbon: (a) frontal view and (b) top view. Pocket amino acids are represented as spheres with different sizes and colors: pocket 1 (magenta), pocket 4 (dark blue), pocket 7 (gray) and pocket 9 (green); molecular surface showing (c) a frontal view of the PBR and (d) the top view, showing the relative depth of the different pockets. P1 and P9 are deeper whereas pockets 4, 6 and 7 are more superficial, lying towards the walls of the groove. (Graphic reprinted from the Reference [30] under the Creative Commons Attribution License (CCAL).)

in a MHC-P interaction exceeds the practical application of QTAIM, which would be the ideal partition scheme. Instead of QTAIM, point-charge multipoles derived from the Mulliken population were obtained from standard quantum mechanical calculations (as provided by programs such as Gaussian). Accordingly, the dipolar and quadrupolar moments and their respective norms are:

$$p_k = \sum_{k=1}^N q_k r_k \quad (13.3)$$

$$d_k = \sqrt{p_k \cdot p_k} \quad (13.4)$$

$$Q_{x_i x_j} = \sum_{k=1}^N q_k (3x_i x_j - r_k^2 \delta_{ij}) \quad (13.5)$$

$$C_k = \sqrt{\sum_{i=1, j=1}^3 Q_{i,j}^2} \quad (13.6)$$

with q_k and r_k the charge and position of the k -th atom, respectively.

Each amino acid residue involved in the complex formation was systematically replaced by each of the remaining 19 genetically-encoded amino acids to quantitatively determine its relevance in the MHC-P complex stability. Figure 13.10 shows a detailed account of the steps followed in this approach. The changes observed after each replacement were estimated by examining three aspects:

- 1) Multipole moments; to estimate the effect that each specific amino acid exerts in the pockets, the Mulliken-derived electrostatic multipoles are evaluated over the pockets (Figure 13.11). The isolated complexing peptide is used as the reference system for evaluating the changes in the multipoles.

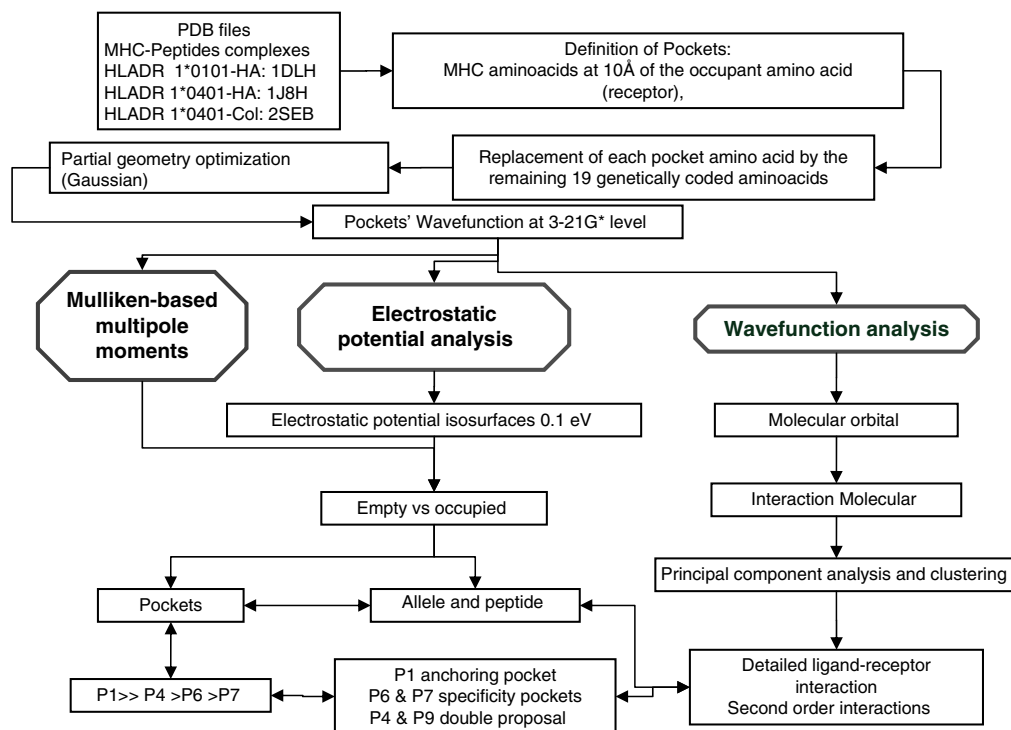


Figure 13.10 Diagram of the quantum study of MHC-P complexes. Each specific analysis is detailed in Figures 13.11–13.13.

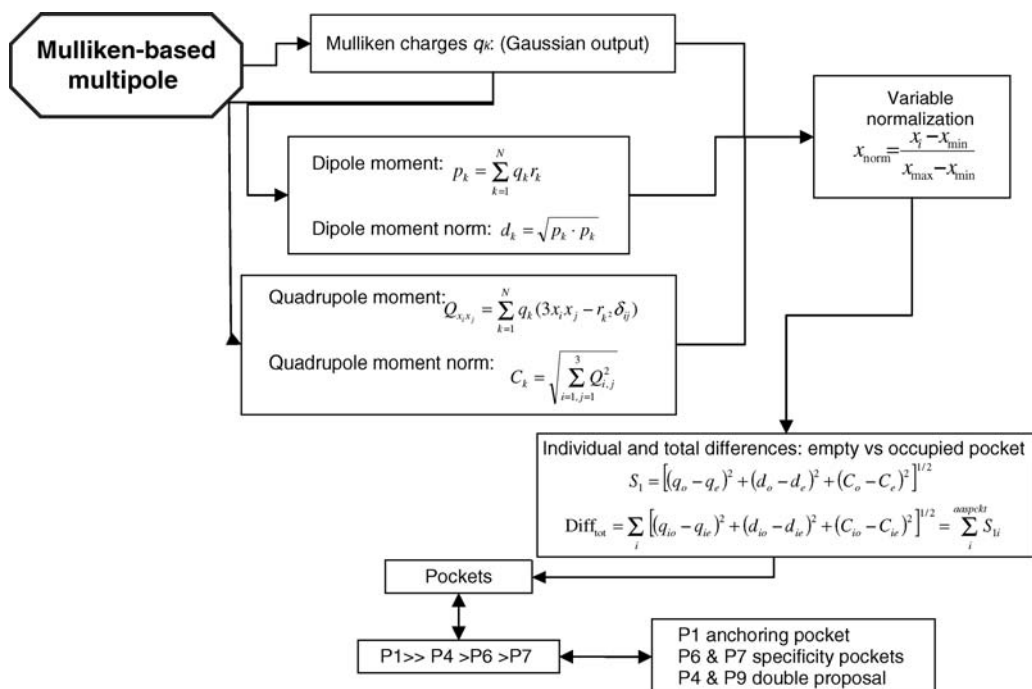


Figure 13.11 Diagram of the quantum study of MHC-P complexes using the Mulliken-based multipole method.

- 2) Electrostatic potential as projected over a molecular surface (Figure 13.12); a traditional study of the QM potential projected over an electron density surface guides the analysis for the atoms directly involved in the complex.
- 3) Identification of those orbitals contributing directly to the complex formation; the orbital expansion coefficients are classified according to pocket and peptide contributions, by a statistical analysis, as schematically explained in Figure 13.13.

While a graphic study of the electrostatic field reveals some details of the peptide–host complex formation, only the multipole and wavefunction analysis provide a hierarchy of relevance among the pocket sites, which is highly correlated with that observed experimentally [31–33]. For example, in a MHCII-peptide complex study, the prevalence for aromatic amino acids in pocket #1 was unambiguously determined (see Table 2 in Reference [5]). Such specific prevalence for aromatic side chains plays a significant role in the complex stability as this pocket works as an anchoring site for the guest peptide. In this way, the traditional direct visual study of the electrostatic potential of the MHC-P complexes provides merely a complementary analysis that verifies the quantitative classification given

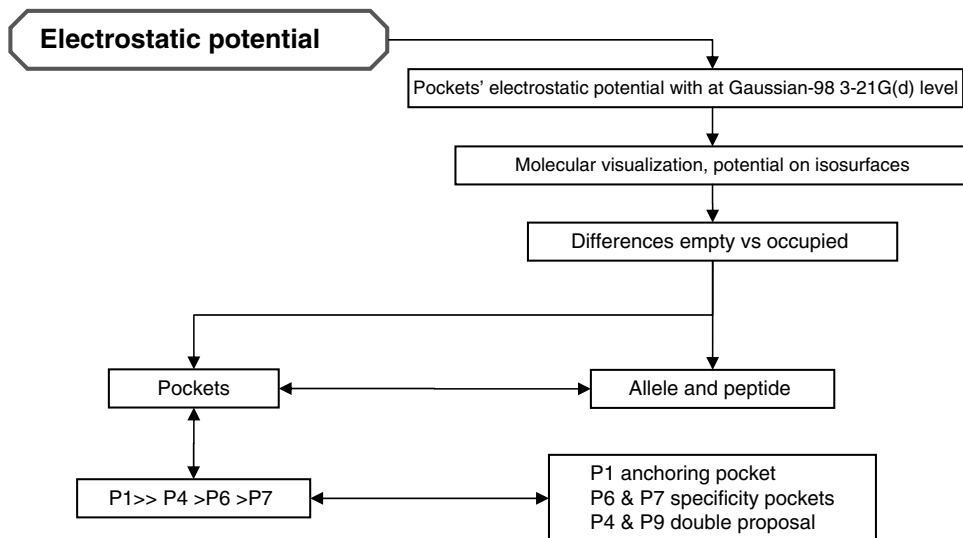


Figure 13.12 Diagram of the quantum study of MHC-P complexes using the electrostatic potential.

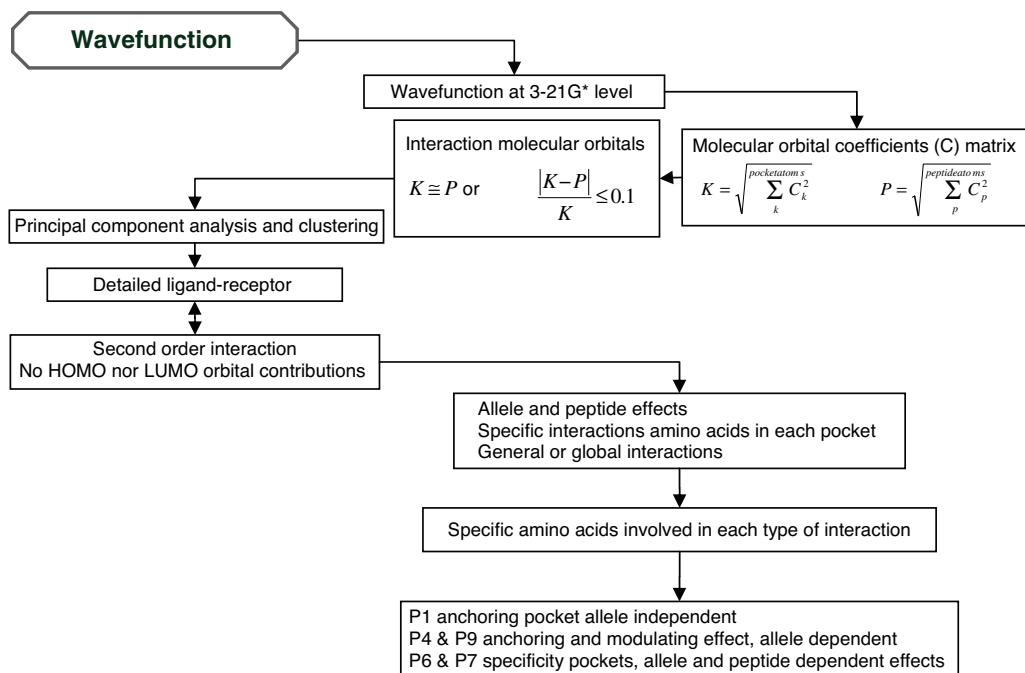


Figure 13.13 Diagram of the quantum study of MHC-P complexes using a wavefunction analysis.

by the other two methods. The success obtained so far in the description of the essential amino acids that are responsible for the complex stability and their respective synonymous replacements validates the overall proposal reviewed in this section.

13.5

Conclusions

This chapter is focused on approaches to extract biochemical information from the wavefunctions of biomolecules, once such wavefunctions are available from QM calculations. The work reviewed in this chapter supports the idea that much of the biochemical information carried by the amino acids is encoded in an electrostatic language. Initially, a principal components analysis (PCA) over a set of amino acid conformers allows one to identify those *ab initio* variables that best describe the features of the amino acids. The electrostatic multipole moments sufficiently account for the characteristic features of the side chains and their interactions to account for known amino acid similarities [2, 4] and their interactions in peptide–host complexes [5, 30–33].

Several concepts related to the quantification of molecular similarity and the relation between theoretically-accessible indices and the bioactivity of molecules are reviewed. The methods and strategies discussed are used to study small peptides but can also be applied to other sets of molecules. The chapter reviews and complements the original work published over the past ten years or so, primarily developed at the Fundación Instituto de Inmunología de Colombia (FIDIC), on the quantum mechanics-based molecular design of a synthetic anti-malarial vaccine.

The strengths of the statistical analysis leading to the identification and classification of the biochemical propensities in peptides and proteins are emphasized in this chapter. These studies reveal that the information regarding the relative physicochemical properties of biomolecules is, to a large extent, encoded in the electrostatic properties of these molecules captured in the multipole expansion. Physical variables, particularly electrostatic properties, account for much of the structural and functional similarities of the amino acids, as shown by a PCA survey of over 1065 amino acids models.

With the ever-increasing power of computers, quantum mechanical calculations are more and more accessible and applicable to larger and larger biomolecular systems. One can only anticipate a parallel ever-increasing reliance on calculated descriptors to predict and classify the physical properties of biomolecules and to correlate these properties with their biological functions.

Acknowledgments

Special thanks to Gavin Heverly-Coulson for his comments on the manuscript and Alfonso Leyva for the initiative that made possible this long-lasting and fruitful

project. We acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada (NSERC) and the provision of computing resources by ACEnet, the regional high performance computing consortium for universities in Atlantic Canada. C.M. acknowledges NSERC for a Discovery Grant, Canada Foundation for Innovation (CFI) for a research infrastructure “Leaders Opportunity” fund, and Mount Saint Vincent University for an internal research grant. We also want to express our gratitude to Edgar Daza from the GQT at Universidad Nacional de Colombia, and to Jose L. Villaveces from GQT at Universidad de los Andes. MEP acknowledges financial support from COLCIENCIAS, Universidad Nacional de Colombia and Universidad del Rosario.

References

- Kumar, D.A. (2001) *Mini Rev. Med. Chem.*, **1**, 187.
- Cárdenas, C., Obregón, M., Llanos, E., Machado, E., Bohórquez, H., Villaveces, J., and Patarroyo, M. (2002) *J. Comput. Chem.*, **26**, 631.
- Andrews, D. (1972) *Biometrics*, **28**, 125.
- Bohórquez, H., Obregón, M., Cárdenas, C., Llanos, E., Suárez, C., Villaveces, J.L., and Patarroyo, M.E. (2003) *J. Phys. Chem. A*, **107**, 10090.
- Cárdenas, C., Villaveces, J.L., Bohórquez, H.J., Llanos, E., Suárez, C., Obregón, M., Patarroyo, M.E. (2004) *Biochem. Biophys. Res. Commun.*, **323**, 1265.
- Lee, K.H., Xie, D., Freire, E., and Amzel, L.M. (1994) *Proteins: Struct. Funct. Genet.*, **20**, 68.
- Schrauber, H., Eisenhaber, F., and Argos, P. (1993) *J. Mol. Biol.*, **230**, 592.
- Bosco, K.H. and Agard, D.A. (2008) *BMC Struct. Biol.*, **8**, 41.
- Trijnstic, N. (1992) *Chemical Graph Theory*, CRC Press Inc, Boca Raton Fl.
- García-Domenech, R., Galvez, J., de Julian-Ortiz, J.V., and Pogliani, L. (2008) *Chem. Rev.*, **108**, 1127.
- SemiChem and the University of Florida (1995) CODESSA: Comprehensive Descriptors for Structural and Statistical Analysis.
- Katritzky, A.R., Karelson, M., Maran, U., and Wang, Y. (1999) *Collect. Czech. Chem. Commun.*, **64**, 1551.
- Rohlf, F.J. (1992) *Numerical Taxonomy and Multivariate Analysis System (NTSYS)*, Ver. 1.8, Exeter Publishing, Ltd, Setauket, NY.
- Jolliffe, I. (2005) Principal component analysis, *Encyclopedia of Statistics in Behavioral Science*, John Wiley & Sons, Ltd., Aberdeen, U.K.
- Haeberlin, M. and Brinck, T. (1997) *J. Chem. Soc., Perkin Trans. 2*, 289.
- (a) Bader, R.F.W. (1990) *Atoms in Molecules: A Quantum Theory*, Oxford University Press, Oxford; (b) Matta, C.F. and Boyd, R.J. (eds) (2007) *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*, Wiley-VCH Verlag GmbH, Weinheim.
- Matta, C.F. and Bader, R.F.W. (2000) *Proteins: Struct. Funct. Genet.*, **40**, 310.
- Matta, C.F. and Bader, R.F.W. (2002) *Proteins: Struct. Funct. Genet.*, **48**, 519.
- Matta, C.F. and Bader, R.F.W. (2003) *Proteins: Struct. Funct. Genet.*, **52**, 360.
- Bader, R.F.W., Popelier, P.L.A., and Chang, C. (1992) *J. Mol. Struct. (THEOCHEM.)*, **255**, 145.
- Chang, C. and Bader, R.F.W. (1992) *J. Phys. Chem.*, **96**, 1654.
- Popelier, P.L.A. and Bader, R.F.W. (1994) *J. Phys. Chem.*, **98**, 4473.
- (a) Biegler-König, F.W., Bader, R.F.W., and Tang, T.-H. (1982) *J. Comput. Chem.*, **13**, 317; (b) Bader, R.F.W. <http://www.chemistry.mcmaster.ca/aimpac/>.
- Kawashima, S., Pokarowski, P., Pokarowska, M., Kolinski, A., Katayama, T., and Kanehisa, M. (2008) *Nucl. Acid. Res.*, **36**, D202.
- Huang, J., Kawashima, S., and Kanehisa, M. (2007) *Genome Inform.*, **18**, 152.
- Kidera, A., Konishi, Y., Ooi, T., and Scheraga, H.A. (1985) *J. Protein Chem.*, **4**, 265.

- 27 Esteve, J.G. and Falceto, F. (2005) *Biophys. Chem.*, **115**, 177.
- 28 Patarroyo, M.E., Amador, R., Clavijo, P., Moreno, A., Guzman, F., Romero, P., Tascon, R., Franco, A., Murillo, L.A., Ponton, G., and Trujillo, G. (1988) *Nature*, **332**, 158.
- 29 Patarroyo, M.E., Romero, P., Torres, M.L., Clavijo, P., Moreno, A., Martinez, A., Rodriguez, R., Guzman, F., and Cabezas, E. (1987) *Nature*, **328**, 629.
- 30 Agudelo, W.A., Galindo, J.F., Ortiz, M., Villaveces, J.L., Daza, E.E., and Patarroyo, M.E. (2009) *PLoS ONE*, **4**, e4164.
- 31 Cardenas, C., Ortiz, M., Balbin, A., Villaveces, J.L., and Patarroyo, M.E. (2005) *Biochem. Biophys. Res. Comm.*, **330**, 1162.
- 32 Cardenas, C., Villaveces, J.L., Suarez, C., Ortiz, M., Villaveces, J.L., and Patarroyo, M.E. (2005) *J. Struct. Biol.*, **149**, 38.
- 33 Balbin, A., Cardenas, C., Villaveces, J.L., and Patarroyo, M.E. (2006) *Biochimie*, **88**, 1307.

14

From Atoms in Amino Acids to the Genetic Code and Protein Stability, and Backwards*

Chérif F. Matta

The electron density lies at the center of the “observable” universe.

(Quoted from this chapter)

14.1

Context of the Work

Genetic information is stored and transcribed in nucleic acid language, a language written in the ink of hydrogen bonding specificity and molecular complementarity. During storage and transcription, the relationship between the genetic information and the physicochemical nature of the letters in which it is written, that is, the nucleic acid bases, is similar to the one between the meaning of the words written in this book and the chemical nature of the ink and paper used in its production; in other words, there exist no relationship. This genetic message is pure *information* in the Shannon’s sense [1, 2], that is, a *capacity* to store and transmit instructions, that can be quantified just like information stored in a line of text or a computer hard disk as [3]:

$$H = -K \sum_i p_i \log p_i \quad (14.1)$$

where K determines the dimensions/units of H ¹⁾ and p_i is the probability of occurrence of a particular letter of the alphabet. DNA language, for example, has an alphabet of four letters, namely, adenine (A), guanine (G), thymine (T), and cytosine (C), with T being replaced by uracil (U) in RNA language.

* The construction of the title of this chapter has been inspired by a chapter by Professors Piero Macchi and Angelo Sironi entitled “Interactions Involving Metals: From ‘Chemical Categories’ to QTAIM and Backwards” published in 2007. I have obtained their permission to adopt a similar construction.

1) When K equals the Boltzmann constant k_B , then H is in dimensions of entropy and if $K = \log_2 e$, then H is in bits.

Equation 14.1 applies to a stretch of nucleic acid provided the bases are (i) *independent* whereby the probability of a letter is unaffected by the nature of the previous letters, such a linear sequence of symbols is termed a “zero-order” or “zero-memory” Markov chain, and (ii) *equiprobable*, that is to say that all the bases occur with equal overall frequencies ($p_A = p_G = p_T$ (or $p_U) = p_C = 1/4$). Deviations from independence and equiprobability must be accounted for when calculating the amount of information stored in real nucleic acids, as discussed in detail in Lila L. Gatlin’s important monograph [3].

After the information is transcribed into mRNA and subsequently “read” by the ribosome, this cell organelle translates the message into amino acid language in the form of a polypeptide chain. (See Chapter 16 for the mechanism of the peptide bond formation in the ribosome.) The polypeptide chain might be the end product (the protein) or it might need to bind to other peptides and/or other molecules before it forms the protein.

Thus, in the ribosome, Shannon’s information contained in the linear sequence of nucleic acid bases is converted into a three-dimensional object, the polypeptide, consisting of amino acid residues linked by peptide bonds. Each of the 20 genetically encoded amino acid residues in the polypeptide is characterized by the (highly transferable) total charge density $Q_{\text{total}}(\mathbf{r})$ of its side chain (the R group) [4–11], where the total density is given by:

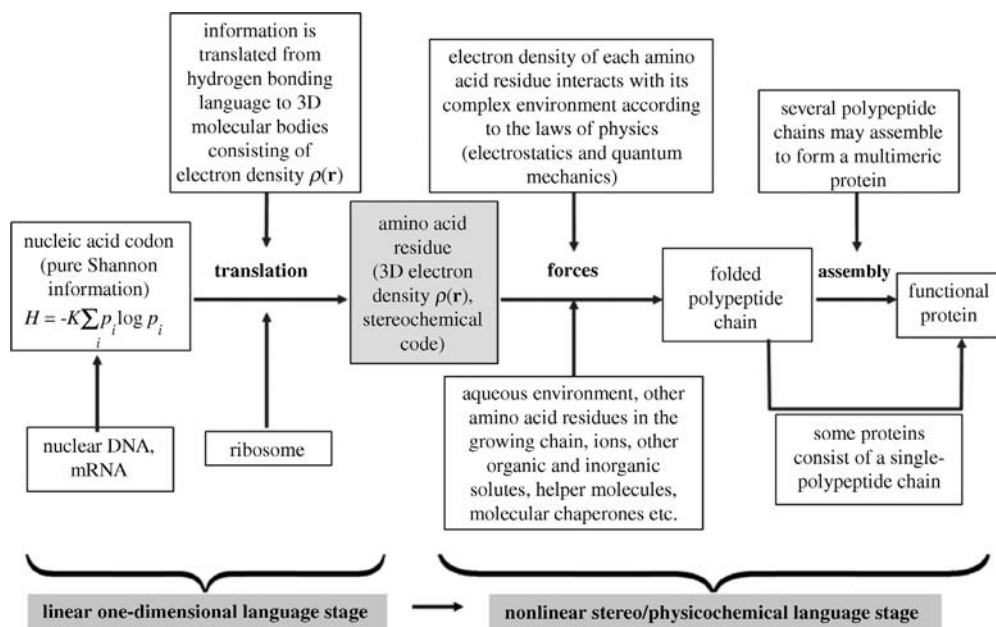
$$Q_{\text{total}}(\mathbf{r}) = -Q(\mathbf{r}) + \sum_{\alpha} Z_{\alpha} \delta(\mathbf{r} - \mathbf{R}_{\alpha}) \quad (14.2)$$

where $Q(\mathbf{r})$ is the electron density and the second term represents the discrete distribution of point-like nuclear charge, Z_{α} being the charge of nucleus α and $\delta(\mathbf{r} - \mathbf{R}_{\alpha})$ a Dirac delta function.

As soon as an amino acid residue in the nascent polypeptide leaves the ribosome it starts interacting with its complex environment, including other amino acid residues in the chain, water molecules, ions and molecules present in the cellular matrix, molecular chaperones that assist the polypeptide to fold in its proper functional native state, and so on. These interactions are rendered *specific* by the properties of the R group of a given amino acid residue, properties that are *completely and solely* determined by the charge density distribution of that residue $Q_{\text{total}}(\mathbf{r})_R$. The charge density of the side chain “stamps” an amino acid with its identity just like the atomic nucleus that stamps an atom with its identity.

The ribosomal translation of the genetic code into a three-dimensional physical field, the total charge density $Q_{\text{total}}(\mathbf{r})_R$, leads to another level of “reading” and “translation”: the charge distribution of the side chain is “read” by its environment with a concomitant “translation” into physical forces that inevitably and uniquely determine a folded polypeptide geometry. Lattman and Rose describe this higher level coding as a *stereochemical code* [12]. Scheme 14.1 summarizes the views described above.

The amount of information encoded in nucleic acid language is linear (additive), similar to a string of zeroes and ones in machine language or a sequence of letters of the alphabet in human languages [3]. In contrast, the stereochemical code [12] introduces another dimension to the genetic language by determining the folding of



Scheme 14.1

the polypeptide/protein through its interaction with its environment. This stereochemical coding dimension is to protein language what is often described by the cliché “read between the lines” to human languages. It is this extra dimension that this chapter is about, which we can translate into the following question:

How strongly does the charge density of a given amino acid determine (correlate with) its physicochemical and biological properties?

In principle, this question has a unique answer: infinitely strong, that is, complete interdependence. *In practice*, however, a comprehensive *ab initio* statistical mechanical theory that connects the quantum properties of the amino acids with their very complex interactions in solution is lacking. As an alternative, we resort to empiricism in the work reviewed in this chapter [6–10], that is, we explore empirical correlations between measured properties of the amino acids and calculated properties of the charge density of their side chains.

Empirical modeling that inevitably introduces simplifications and assumptions cannot match the elegance and power of a first-principle statistical mechanical theory. However, such modeling can (and do) lead to qualitative insight into the interaction of amino acid residues with their environment. Models that yield strong correlations can also serve as a practical quantitative structure–activity (or property) relationship (QSAR/QSPR) tool, an approach of considerable importance in drug and material design [13]. The modeling reviewed in this chapter is based on quantities derived from the underlying electron density, a physical observable. In basing the modeling

on properties derived from a real physical field, one eliminates an important source of uncertainty in the modeling, that is, the quality and nature of its component.

14.2

The Electron Density $\varrho(\mathbf{r})$ as an Indirectly Measurable Dirac Observable

The electron density at any point specified by the position vector $\mathbf{r} = i\mathbf{x} + j\mathbf{y} + k\mathbf{z}$ is defined as the probability density of finding an electron, any electron regardless of spin, in a particular volume element $d\tau = dx dy dz$ surrounding point $\mathbf{r}$ *weighted by the total number of electrons* (N) in the system, that is:

$$\varrho(\mathbf{r}) = N \int d\tau' \Psi^*(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) \Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) \quad (14.3)$$

where $\mathbf{x}_i$ is the set of three spatial coordinates and the spin coordinate of the i th electron, Ψ the many-electron Born–Oppenheimer wavefunction, and the mode of integration denoted by $\int d\tau'$ implies integration over the spatial coordinates of all electrons except one followed by the summation over all spins.

Integrating the electron density (Equation 14.3) over a region of space ω will yield the quantum average of the electron population in that region $N(\omega)$, and if the integration covers all space, then the integral delivers the total number of electrons N in the molecule, since in this case:

$$\int \varrho(\mathbf{r}) d\mathbf{r} = N \underbrace{\int_{\omega=\text{all space}} d\mathbf{r} \int d\tau' \Psi^*(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) \Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)}_1 = N \quad (14.4)$$

provided that Ψ is normalized.

According to the celebrated Hohenberg and Kohn (HK) theorem [14], the electron density $\varrho(\mathbf{r})$ determines the external potential $V[\varrho]$ uniquely, and therefore it determines the charge and positions of the nuclei $\{Z_\alpha \delta(\mathbf{r} - \mathbf{R}_\alpha)\}$ uniquely, that is, it determines the molecular geometry. By determining the external potential, the electron density also determines the total charge density (Equation 14.2). Also, by determining the total number of the electrons $N[\varrho]$, $\varrho(\mathbf{r})$ fixes the Hamiltonian $\hat{H}[\varrho]$ and its eigenfunctions, and therefore completely determines the properties of the ground-state $O[\varrho]$. These relationships are written symbolically:

$$\varrho(\mathbf{r}) \rightarrow \left\{ \begin{array}{c} V[\varrho] \\ N[\varrho] \end{array} \right\} \rightarrow \hat{H}[\varrho] \rightarrow \Psi[\varrho] \rightarrow O[\varrho] \quad (14.5)$$

Even though the functional relationships linking the electron density to most other observables are not generally known, the important fact is that the electron density *does* fix them uniquely. In a sense, if one imagines an n -dimensional Euclidean space $\mathbb{R}^n$, where n is the number of observables, then each molecule would be represented by a vector in this space. The closer two molecules are in this “observable” space, the more they are similar to one another. Because of mapping 14.5, the position



Figure 14.1 Escher's Development II (1939) labeled with the symbol of the electron density, $\rho(\mathbf{r})$. The reptiles represent the ground-state properties that are all uniquely determined by the density and, reciprocally, collectively determine it uniquely. (Adapted from a private

communication with Professor Philip Coppens and used with his permission. The artwork has been reproduced with the permission of M.C. Escher's "Development II" © 2009 The M.C. Escher Company–Holland. All right reserved).

associated with a particular molecule in this space is uniquely determined by its electron density distribution. In a sense, then, the electron density lies at the center of the ground-state "observable" universe, where I use "center" only in the metaphorical sense.

Professor Philip Coppens has once illustrated the relationship of the density and the properties of the ground state using "Development II," a masterpiece of one of his (and one of my) favorite artists, M.C. Escher. The reptiles in Development II represent the various ground-state "observables" emanating from the underlying electron density, and in also converging on it, collectively, uniquely determining it (Figure 14.1).

The electron density $\rho(\mathbf{r})$ is a Dirac *observable*, since it satisfies the following two conditions [15] outlined in Dirac's book [16]:

- (a) It is a *real* (as opposed to complex) dynamical variable that is the expectation value of a linear (and, naturally, Hermitian) operator $\hat{\rho}(\mathbf{r})$.
- (b) The eigenstates of $\hat{\rho}(\mathbf{r})$ form a complete set of coordinate states $|\mathbf{r}\rangle$.

Dirac emphasizes the question: "Can every observable be measured?" and provides the following answer ([16], p. 37):

"The answer theoretically is yes. In practice it may be very awkward, or perhaps even beyond the ingenuity of the experimenter, to devise an apparatus which could

measure some particular observable, but the theory always allows one to imagine that the measurement can be made.”

The electron density, in the light of Dirac’s question, is an observable that can be measured *indirectly*, most commonly through the intermediacy of the structure factors determined in a crystallographic X-ray diffraction experiment.

The electron density in a crystal is built from a repeating unit cell with periodicity in each of three spatial dimensions. As any periodic function, the electron density can be expanded as a Fourier series, in this case in three dimensions, where the expansion coefficients $F(hkl)$ (or structure factors) are the unknowns to be determined from the diffraction experiment. Thus, the electron density is obtained from the X-ray experiment through the *reverse* (discrete) Fourier transform of the diffraction pattern [17]:

$$\rho(xyz) = \frac{1}{V} \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} \underbrace{|F(hkl)|}_{\text{magnitude}} \underbrace{e^{i\alpha(hkl)}}_{\text{phase}} e^{-2\pi i(hx + ky + lz)} \quad (14.6)$$

where the summations are truncated to finite limits (and hence introducing an unavoidable truncation error); V is the volume of the unit cell; x , y , and z are the fractional coordinates of the point at which ρ is specified in the unit cell; h , k , and l are the coordinates of a reflection in reciprocal space (or equivalently, the periodicities of the electron density in the crystal along the x -, y -, and z -axes, respectively); and where the reciprocal space is “sampled” only at those positions satisfying the Bragg condition, hence the discrete nature of the Fourier transform.

The structure factors are characterized by phase and magnitude (Equation 14.6), but since experiments can only measure the intensity of a reflection that is proportional to the square of the structure factor, that is, $I(hkl) \propto |F(hkl)|^2$, and since the structure factors are generally complex, the phase information is lost. This is known as the *phase problem*.

The phase problem has long been considered to be unsolvable before a number of elegant solutions have been discovered. Notably among these solutions is the approach known as the “Direct Methods,” discovered by Jerome Karle²⁾ and Herbert Hauptman for which they were rewarded with the 1985 Nobel Prize in Chemistry. The Direct Methods led to a great enhancement of the widespread use of X-ray crystallography to solve crystal structures (see Ref. [17] for a concise review of modern crystallography).

The experimentally *measurable* quantity in an X-ray diffraction experiment is a set of indexed intensities $I(hkl)$ constituting the *observed diffraction pattern* (these intensities are usually corrected for vibration-induced thermal diffuse scattering, extinction, etc.) In parallel, a *calculated diffraction pattern* is obtained by subjecting a “model density” to a *direct* Fourier transform to obtain calculated structure factors:

2) Dr. Jerome Karle has coauthored Chapters 1 and 16 of this book.

$$F(hkl) = \int_0^1 \int_0^1 \int_0^1 \underbrace{\rho(xyz)}_{\text{model density}} e^{2\pi i(hx + ky + lz)} dx dy dz \quad (14.7)$$

where the integral extends over the volume of the crystallographic unit cell and x , y , and z are direct space fractional coordinates with values ranging from 0 to 1 (note the change in sign of the exponential in Equation 14.7 when compared to Equation 14.6). The magnitude of each calculated structure factor, $|F(hkl)|_{\text{calculated}}$, is then compared with the magnitude of the corresponding experimentally measured structure factor $|F(hkl)|_{\text{observed}}$, and an agreement factor is calculated. The overall agreement between the calculated and observed diffraction patterns is measured by the so-called residual factor or R_f , defined:

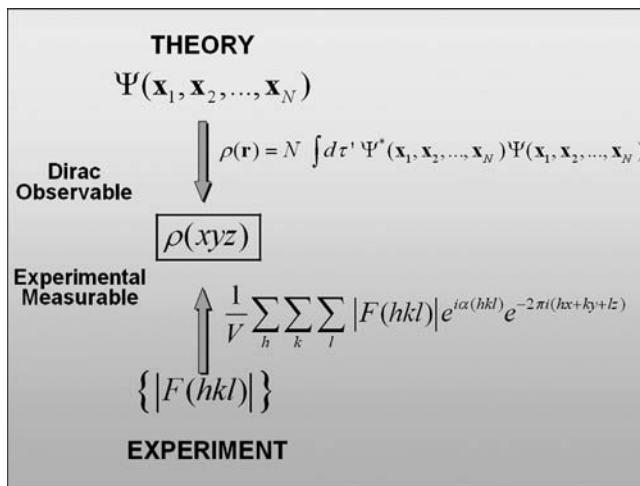
$$R_f = \frac{\sum ||F_{\text{observed}}| - |F_{\text{calculated}}||}{\sum |F_{\text{observed}}|} \quad (14.8)$$

The model density is modified iteratively until the discrepancy between the calculated and observed diffraction patterns (measured by R_f) is minimized and that further Fourier recycling does not decrease its value. When this point is reached, the structure is considered to have been “solved.”

The modeling of the density in the unit cell is often performed using overlapping spherical atomic densities obtained from quantum mechanical calculations on isolated atoms, an approach known as the independent atom model (IAM). Crystallographic experience has demonstrated that the independent atom model is sufficiently accurate for routine crystal structure determinations of molecular geometry. On the other hand, the independent atom model is insufficiently flexible to capture the fine details of chemical bonding when a detailed analysis of the topology and topography of the electron density is the primary goal of the diffraction experiment. In this case, a more sophisticated *nonspherical* modeling of the atomic densities is required to provide the necessary flexibility [18–21]. Data collection for subsequent nonspherical refinement is often performed at very low temperatures to reduce the thermal smearing of the diffraction pattern. This type of accurate measurement of the electron density followed by aspherical modeling has become routine nowadays [19, 20] even in the case of very large biological molecules (see, for example, Refs [22–42]). A study by Luger *et al.* provides one of the numerous examples of the agreement between electron densities obtained from experiment [34] and the corresponding densities obtained from theory [43], in this case, the densities of morphine and related opioids.

This section demonstrates that the electron density lies at the intersection of theory (Equation 14.3) and experiment (Equation 14.6): it is a Dirac quantum mechanical *observable* that is *indirectly measurable* through comparison with a calculated model of the density (Equation 14.8). These interrelationships are summarized in Scheme 14.2.

The electron densities analyzed in the remainder of this chapter were obtained from theory, but the same analysis could have been based entirely on experimental



Scheme 14.2

densities. This independence of the origin of the density is a strength of the theory used in this analysis, namely, the quantum theory of atoms in molecules (QTAIM) [44–46]. In the next section, we present a brief reminder of a few basic concepts of this theory.

14.3

Brief Review of Some Basic Concepts of the Quantum Theory of Atoms in Molecules

This section is meant to give the reader who is unfamiliar with the quantum theory of atoms in molecules (or QTAIM, and in the older literature AIM) a quick overview of some of its basic concepts. It is well beyond the scope of this chapter to provide a thorough and/or mathematically rigorous review of this theory. It is impossible, therefore, to convey here its full beauty, elegance, and interpretative and predictive power. For these, the reader may refer to the original works by Bader and co-workers [44, 47–51].

QTAIM is a “model free” theory of chemistry in the sense that the components of its construction are Dirac observables. The primary observable upon which the theory is based is the electron density (Equation 14.3), the physical field responsible for the space filling manifestation of matter at the level of chemistry and biology.

The electron density has a three-dimensional topography characterized by marked local maxima at the positions of the nuclei. Figure 14.2 is a relief map displaying the topography of the calculated electron density in the plane of a guanine–cytosine Watson–Crick (GC–WC) DNA dimer. Remarkably, the *chemical structure* of the dimer emerges already from this representation of the density in the plane of the dimer. An examination of the relief map in this figure reveals that there exists a ridge of density with a saddle point connecting any two nuclei belonging to a pair of bonded atoms in

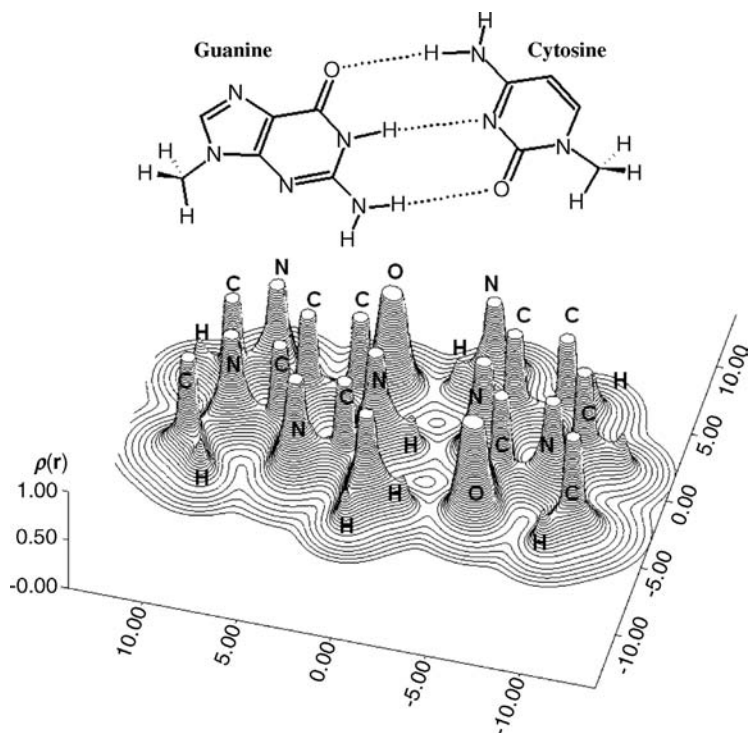


Figure 14.2 Relief map of the electron density in the molecular plane of a guanine–cytosine Watson–Crick base pair along with the chemical structure (the position where each base is connected to the deoxyribose sugar of the DNA backbone has been substituted by a methyl group). The value of $\rho(r)$ is truncated to 1.0

atomic unit (au). The x - and y -axes are labeled in atomic units of length (1 au of length = 1 bohr = a_0). Each contiguous line indicates a constant value of the electron density. The projection of these isodensity lines on the molecular plane constitutes a “contour plot” (see Figure 14.3). (Reproduced from Ref. 10).

the chemical structure. These saddle points are termed bond critical points (BCPs). The gradient path originating at the BCPs and terminating at the nuclei is a line of locally maximum density termed the *bond path* [52, 53].

Figure 14.3 represents the projection of the isocontour lines on the molecular plane of the GC–WC pair (a contour plot of the density). In the figure, the lines linking the nuclei are the bond paths. Figure 14.4 shows the gradient vector field corresponding to the electron density. The gradient vector field provides a natural partitioning of the electron density into nonoverlapping regions each enclosing one and only one nucleus, a partitioning highlighted by coloring each mononuclear region with a given color for every element (red for oxygen, blue for nitrogen, yellow for carbon, and violet for hydrogen). The gradient lines in an atomic region converge on one nucleus enclosed within that region, as can be seen from the figure. It is said that the nuclei are “attractors” because they “attract” the gradient vector field lines. The intersections of the interatomic surfaces that partition space into nonoverlapping

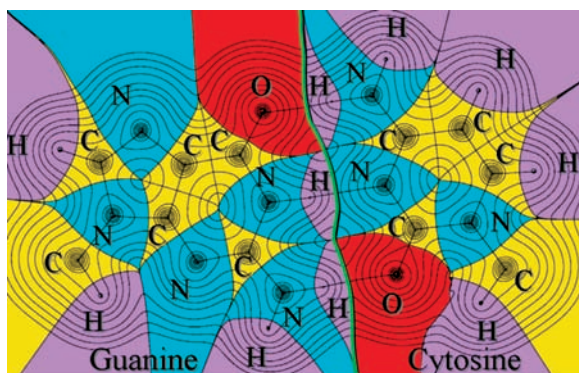


Figure 14.3 Contour map of $\rho(\mathbf{r})$ corresponding to the relief map of the electron density in the molecular plane of a guanine-cytosine Watson-Crick dimer in Figure 14.2. The outermost contour has the isodensity value of 0.001 au followed by 2×10^n , 4×10^n , and 8×10^n au with n starting at -3 and increasing in steps of unity. The lines connecting the nuclei are the bond paths, and the lines delimiting each atom are the intersection of the interatomic zero-flux surfaces with the plane of the figure. The intersection of a bond path with an

associated interatomic surface occurs at the bond critical point, BCP, where $\nabla\rho(\mathbf{r}) = 0$. A green line has been added manually to highlight the zero-flux surface separating the two monomers. This intermonomer G/C zero-flux surface consists of the union of three interatomic surfaces arising due to the three hydrogen-bonded interaction in this Watson-Crick dimer. Crosses not linked by bond paths are the projections of the nuclear positions of atoms out of the plane of the figure. (Adapted from Ref. [10]).

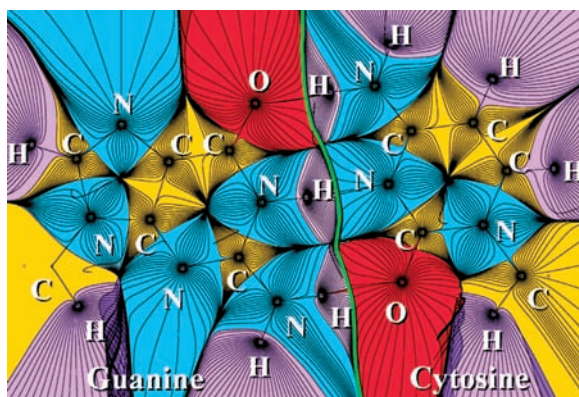


Figure 14.4 Displays of the trajectories of the gradient field of the density ($\nabla\rho(\mathbf{r})$) in the molecular plane of a guanine-cytosine Watson-Crick dimer corresponding to Figure 14.3. All the paths in the neighborhood of a given nucleus terminate at that nucleus and define the atomic basin. Of course, gradient lines never cross, the few lines that do, especially at the bottom of the plot in the basins of the hydrogen atoms and the oxygen atom, are

projections of out-of-plane gradient vector field lines that are within the plane thickness tolerance. A green line has been added to highlight the G/C intermonomer zero-flux surface that consists of the union of three interatomic surfaces arising from the three hydrogen-bonded interactions. Crosses not linked by bond paths are the projections of the nuclear positions of atoms out of the plane of the figure. (Adapted from Ref. [10]).

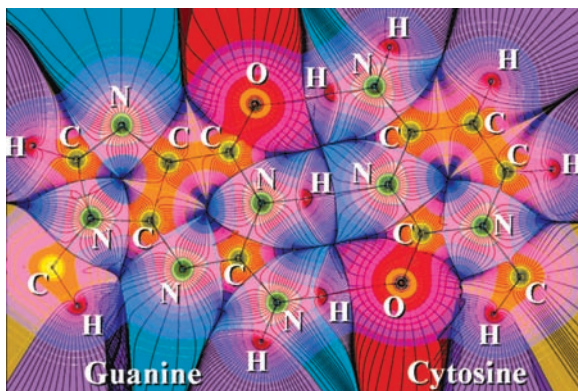


Figure 14.5 A superimposition of the gradient vector field map of a guanine-cytosine Watson-Crick base pair (Figure 14.4) and the corresponding electron density contour map (Figure 14.3) showing the natural partitioning of

the density into separate atomic basins, each containing a single nucleus. *Note the significant departure of atoms in molecules from spherical symmetry.* (Cover graphic art of this book. Adapted from Ref. [10]).

mononuclear regions appear as the lines bounding the atoms in Figures 14.3–14.5. An interatomic surface is never crossed by the gradient vectors of the electron density. Restated mathematically, an interatomic surface satisfies, *locally*, the following condition:

$$\nabla \rho(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r}) = 0, \quad \text{for all } \mathbf{r} \text{ belonging to the surface } S(\Omega) \quad (14.9)$$

where $\mathbf{r}$ is the position vector and $\mathbf{n}(\mathbf{r})$ the unit vector normal to the surface $S(\Omega)$. An interatomic surface is said to be one of the zero flux in the gradient vector field of the density. The surface bounding an atom or a group of atoms in a molecule or a crystal is always one of the zero flux and so is the surface delimiting monomers in a weakly bonded dimer such as the GC-WC base pair. The union of the interatomic surfaces bounding a given atom defines the *shape* of this atom in the molecule.

In Figures 14.3 and 14.4, the intersection of the zero-flux intermonomer surface delimiting the two monomers (G and C) with the plane of the figure is highlighted in green. This intermonomer “hydrogen-bonded surface” is the union of three interatomic surfaces:

$$\begin{aligned} \text{guanine|cytosine} = & (\text{guanine-}\boxed{\text{O}\cdots\text{H}}\text{-N-cytosine}) \cup (\text{guanine-}\boxed{\text{N}\cdots\text{H}}\text{-N-cytosine}) \\ & \cup (\text{guanine-N-}\boxed{\text{H}\cdots\text{O}}\text{-cytosine}) \end{aligned} \quad (14.10)$$

and similarly for the adenine-thymine WC base pair:

$$\begin{aligned} \text{adenine|thymine} = & (\text{adenine-}\boxed{\text{N}\cdots\text{H}}\text{-O-thymine}) \cup (\text{adenine-}\boxed{\text{N}\cdots\text{H}}\text{-N-thymine}) \\ & \cup (\text{adenine-C-}\boxed{\text{H}\cdots\text{O}}\text{-thymine, weak}) \end{aligned} \quad (14.11)$$

where the vertical bar denotes the zero-flux surface. The guanine|cytosine and adenine|thymine(or uracil) hydrogen bonding zero-flux surfaces are the “ink” in

which the genetic code is written, stored, transcribed, and through the intermediacy of which it is read and translated by the ribosome.

Figure 14.5 (this book's cover theme) is a superposition of the plots in Figures 14.3 and 14.4 showing all the elements describing the topography and topology of the density together, namely, the gradient vector field lines (which are always perpendicular to the isodensity contour lines), the isodensity contour lines, the zero-flux surfaces, and the bond paths. In Figures 14.3–14.5, the intersections of the zero-flux surfaces partitioning the GC-WC dimer into separate atomic basins appear as lines that are crossed each by a single corresponding bond paths once, the point of intersection being the BCP. The atomic basin and its associated nucleus constitute a bounded atom in a molecule, an open quantum subsystem [44] with well-defined properties such as energy, charge, dipole (and higher electric multipoles), and so on. In some systems, there exists basins that are associated with attractors other than nuclei, the so-called nonnuclear attractors (NNA) [54–56] and these behave in every respect as open quantum subsystems and hence termed “pseudo-atoms” (because they lack the nucleus). An example of a NNA of relevance to quantum biochemistry is the solvated electron [57] generated, for example, by the interaction of ionizing radiation with water.

Figure 14.6 traces the calculated bond paths and displays the positions of the BCP and ring critical points (RCP), the collection of which is known as the molecular graph. What should be added here is that *every bond path is mirrored by a “shadow path” called the “virial path,”* a line of maximally negative (maximally attractive) potential energy density, that is, of maximal local stability, again in real three-dimensional space [58]. There exists a one-to-one homeomorphic correspondence between bond paths (or molecular graphs) and virial paths (or virial graphs), where molecular and virial graphs denote the collection of all bond or virial paths defining the chemical structure of interest. The homeomorphism of the molecular and virial graphs associates a “chemical bonding structure” with a corresponding “energetic stabilization structure”.

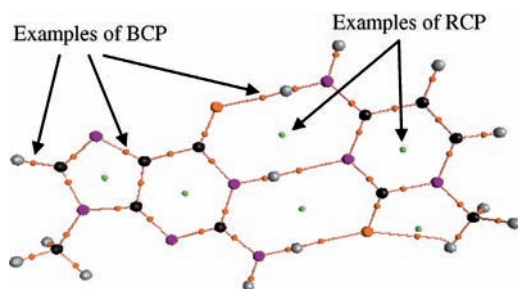


Figure 14.6 The molecular graph of the guanine–cytosine base pair. The set of bond paths displayed in this figure recover the usual chemical structure, with every bonded atom linked by a unique bond path, the small dots

along each bond path is a bond critical point (BCP). The unconnected dots enclosed by rings are the ring critical points (RCP). (Adapted from Ref. [10]).

While the term “bond” or the phrase “there exists a bond between A and B” permeate all of chemistry, these are neither *uniquely* definable in terms of physics nor do they satisfy Dirac’s conditions for quantum observables. Bader has recently demonstrated that the distinction between a “chemical bond” and a “bond path” is not only a question of semantics and grammar but, primarily, one of physics [59]. The concept of the “bond path” is unambiguously defined in terms of the observable and measurable electron density and its associated virial field (potential energy density).

It fascinates the author of this chapter that the chemical bonding structure inferred on the basis of chemical and spectroscopic knowledge emerges naturally and completely as a set of bond paths, including those of weak bonding interactions, from the topology of a real three-dimensional “observable” field, the electron density. It is equally remarkable that whenever a bond path links two nuclei in the electron density field, a path of local energetic stability in real space linking the same pair of nuclei accompany it as its shadow in the potential energy density field.

We now further allude briefly to the virial (potential energy density) field.³⁾ The molecular virial theorem that specifies the relationship between the potential and kinetic energies has been generalized to the following *local* form [60]:

$$\left(\frac{\hbar^2}{4m}\right)\nabla^2 Q(\mathbf{r}) = 2G(\mathbf{r}) + \mathcal{V}(\mathbf{r}) \quad (14.14)$$

where $G(\mathbf{r})$ is the gradient kinetic energy density and $\mathcal{V}(\mathbf{r})$ is the *virial field* that is “ N times the potential energy density of one electron at $\mathbf{r}$ as determined by its average interaction with all the other particles in the system” [44]. The virial field is everywhere negative and integrates to the total potential energy of the molecule. As mentioned above, this field is homeomorphic with the electron density [58], that is, has an identical topology.

The *local virial theorem* (Equation 14.14) is an exact local relationship between potential and kinetic energy densities on one hand and the Laplacian of the electron density on the other, which applies at an arbitrary spatial position $\mathbf{r}$, a remarkable equation. Bader has postulated [61] and proved [44, 62–64] that the integrated form of this theorem constitutes an *atomic virial theorem* that can be used to define the energy of an atom in a molecule $E(\Omega)$, that is, the contribution of this atom to the total molecular energy. To appreciate the nontriviality of $E(\Omega)$, it is perhaps sufficient to remember that such an atomic energy has to include, for example, a contribution from the nuclear–nuclear repulsion energy.⁴⁾

As mentioned above in this section, atoms in molecules are true quantum mechanical open systems [50]. Equation 14.9 has been shown to embody the constraint necessary for the generalization of Schwinger’s principle of stationary action [65] to a quantum subsystem, the physics of a closed isolated system being a special limiting case. The general equation of motion for an observable $\hat{O}$ is expressed [50]:

3) See Chapter 10 of this book by Professors Richard F.W. Bader and Fernando Cortés-Guzmán on the transferability of the virial field in DNA base pairing.

4) For a discussion of the meaning of atomic energies within the frameworks of *ab initio* theory and density functional theory, see Section 15.3 and also the appendices of Ref. [111].

$$N \int_{\Omega} d\mathbf{r} \int d\mathbf{r}' \frac{\partial [\Psi^* \hat{O}(\mathbf{r}) \Psi + \text{cc}]}{\partial t} = \left[\frac{i}{\hbar} \langle \Psi | [\hat{H}, \hat{O}(\mathbf{r})] \Psi \rangle_{\Omega} + \text{cc} \right] - \oint d\mathbf{S}(\mathbf{r}_s; \Omega) \cdot [\mathbf{J}_O(\mathbf{r}_s) + \text{cc}] \quad (14.15)$$

where cc is the acronym of “complex conjugate” and the second term is a surface integral of the net flux in the current density ($\mathbf{J}_O$) of property O through the surface bounding the system $\mathbf{S}$. Equation 14.15 applies to any spatial region satisfying the boundary condition expressed in Equation 14.9, namely, a *proper open system*, that is, a system bounded by a surface of zero flux in $\nabla \mathbf{q}(\mathbf{r})$. As the surface term vanishes at infinity for a closed isolated system, one recovers the usual quantum mechanical theorems.

The disjoint partitioning a molecule (or a crystal) into atomic basins on the basis of Equation 14.9 entails the partitioning of any molecular property that can be expressed as a space-filling density (whether a scalar, vector, or tensor density field) into atomic contributions through its integration over this bounded volume. Examples of such properties include atomic charges (the zeroth-order atomic electrostatic multipole) and higher order electrostatic multipoles, the different contributions to the atomic energies, and even response properties such as the polarizability [66].

Atomic volumes that represent the “steric bulk” of an atom in a molecule can also be defined within this theory as the volume bounded by a (or the union of several) zero-flux surface(s) in the interior of the molecule. If the atom has an outer exposed surface that extends to infinity, experience has shown that the value of $q = 0.001$ au corresponds to experimental van der Waals molecular sizes in the gas phase. The 0.001 au isodensity envelope is thus selected as the outer bounding surface of the molecule. This choice is further justified because it most often contains more than 99% of the electron population of the molecule. The $q = 0.002$ au envelope corresponds to the van der Waals sizes in solution (Figure 14.7).

For a system at equilibrium, the expectation value of an operator averaged over all space can be written as a sum of the expectation values of this operator averaged over each individual atom in the molecule or the crystal, that is:

$$\langle \hat{O} \rangle_{\text{molecule}} = \sum_i^{\text{all atoms in the molecule}} \left(N \int_{\Omega_i} \left\{ \int \frac{1}{2} [\Psi^* \hat{O} \Psi + (\hat{O} \Psi)^* \Psi] d\mathbf{r}' \right\} d\mathbf{r} \right) \quad (14.16a)$$

$$= \sum_i^{\text{all atoms in the molecule}} \left(\int_{\Omega_i} q_O d\mathbf{r} \right) \quad (14.16b)$$

$$= \sum_i^{\text{all atoms in the molecule}} \langle \hat{O}(\Omega_i) \rangle \quad (14.16c)$$

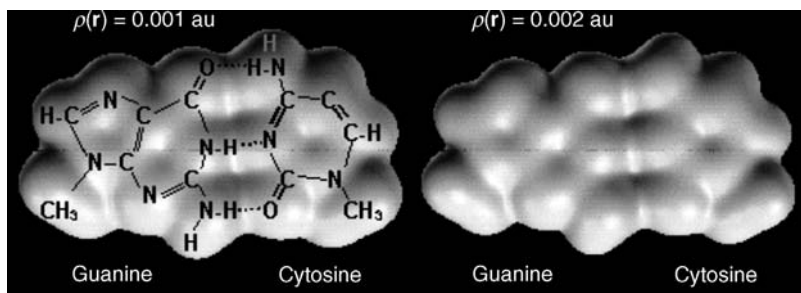


Figure 14.7 Displays of two isodensity $\rho(\mathbf{r})$ envelopes of a guanine–cytosine triply hydrogen-bonded DNA base pair. The $\rho = 0.001$ au envelope corresponds to the van der Waal's molecular size in the gas phase while the $\rho = 0.002$ au is sometimes a better measure of

molecular size in the condensed phase. One can see a ridge of density along the intersection of the zero-flux surface separating the two dimers, the hydrogen bonding surface (the ink by which genetic information is written and transcribed). (Adapted from Ref. [10]).

in which $\hat{O}$ is a linear Hermitian operator corresponding to an observable, $\langle \hat{O} \rangle_{\text{molecule}}$ is its molecular expectation value, and $\langle \hat{O}(\Omega_i) \rangle$ is its corresponding atomic expectation value. The mode of integration is the same as in Equations 14.3 and 14.4. The last equality (14.14c) is the mathematical expression of additivity of atomic properties. Thus, the molecular value of any property O that can be expressed in terms of a real space density “dressed density $\rho_O(\mathbf{r})$ ” can be written as a sum of atomic contributions obtained by averaging the appropriate operator over the volume of the atom. An equivalent statement is that since the atomic properties are defined in complete analogy to the molecular case, the theorems of quantum mechanics that apply to the molecule as a whole also apply to each of its constituent atoms, the virial theorem being an important example.

Modeling electrostatic forces occupies a central role in the modeling of molecular recognition. Multipole moment expansions can be used to express the molecular electrostatic potential in terms of QTAIM atomic moments to an accuracy that depends on the number of terms included in the expansion. The first term, the monopole (or atomic charge) can be obtained by subtracting the atomic population of atom Ω , $N(\Omega)$, from the nuclear charge (Z_Ω) [15, 67]. In its turn, the atomic population, which is the average number of electrons in the atomic basin, is obtained by letting $\hat{O} = \hat{1}$ (the unit operator) in Equation 14.16. In general, by inserting the appropriate operator in Equation 14.16, one obtains the corresponding atomic expectation value of that operator.⁵⁾ As another example, the atomic dipole moment is obtained from:

$$\mu(\Omega) = -e \int_{\Omega} r_{\Omega} \rho(\mathbf{r}) d\mathbf{r} \quad (14.17)$$

where the origin is placed at the position of the nucleus of atom Ω .

5) See Chapter 11 for explicit formulas of some important atomic properties defined with QTAIM (Equations 11.1–11.8).

Popelier *et al.* has shown on several occasions that the molecular electrostatic potential is reproduced with high accuracy from QTAIM multipoles [68–71]. The reader is referred to Bader’s book [44], Popelier’s introduction to QTAIM [45], or the introductory chapter of a recently edited book on QTAIM [72] for a more comprehensive discussion of atomic properties and of QTAIM in general.

To conclude this section, it is important to emphasize that an atomic property is a quantum mechanical expectation value averaged over an open quantum subsystem in analogy to the quantum mechanical expectation value averaged over the closed total system. In this sense, if a Dirac observable is expressible as a real space density, then it can be averaged on equal footing at the molecular and atomic levels. These facts are irreconcilable with claims that atomic charges (for example) are not uniquely defined [67]. The additivity of atomic expectation values is inextricably associated with the exhaustive partitioning of molecular space into nonoverlapping regions, and the uniqueness of these expectation values reflects the uniqueness of the form of the atom in its immediate environment. Figures 14.3–14.5 reveal the considerable deformation of atoms in molecules from spherical symmetry, in contrast with overlapping atoms that do not have a form since they “interpenetrate.” This conclusion, that an atom must have a form, has been reached through reasoning by a philosopher of the stature of Bertrand Russell, the winner of the 1950 Nobel Prize in Literature, who argues in his *A History of Western Philosophy* [73] (p. 165):

“... it is in virtue of the form that the matter is some one definite thing, and this is the substance of the thing. What Aristotle means seems to be plain common sense: a ‘thing’ must be bounded, and the boundary constitutes its form. ... We should not naturally say that it is the form that confers substantiality, but that is because the atomic hypothesis is ingrained in our imagination. Each atom, however, if it is a ‘thing’, is so in virtue of its being delimited from other atoms, and so having, in some sense, a ‘form’.”

14.4

Computational Approach and Level of Theory

The geometries of the 20 genetically encoded amino acids were optimized without constraints [7–9] at the restricted Hartree–Fock (RHF)/6-31+G(*d*) level. This level of theory has been shown to be suitable by Head-Gordon *et al.* [74]. An extensive comparison of the optimized geometries of the amino acids and the corresponding X-ray crystallographic geometries⁶⁾ has provided more support for the suitability of this level of theory [8]. Furthermore, a recent detailed comparison of Hartree–Fock and DFT(B3LYP) geometries on one hand and the corresponding geometries optimized at the MP2 level of theory on the other has shown that HF slightly outperforms DFT

6) An extensive tabulation of accurate X-ray and neutron diffraction determinations of the amino acids can be found in the Appendix at the end of this chapter.

(B3LYP) in reproducing the MP2 geometries of a test set of 30 small molecules representing fragments present in the amino acids [75].

Single point calculations of wavefunctions using 6-311++G(*d*, *p*) basis set have been performed using the optimized geometries described above, a level of theory denoted by RHF/6-311++G(*d*, *p*)/RHF/6-31+G(*d*). Further computational details can be found in the original references [7–9].

The α -amino and α -carboxylic groups of each amino acid were modeled in their neutral form to avoid charge separation, since our primary goal is to model the side chains in proteins where they are attached to α -carbon atoms that are attached to peptide bonds without formal charge separation. The side chains, however, were modeled in their most prevalent ionization state at physiological pH.

14.5

Empirical Correlations of QTAIM Atomic Properties of Amino Acid Side Chains with Experiment

14.5.1

Partial Molar Volumes

The partial molar volume of a solute [$\text{L}^3 \text{mol}^{-1}$] in a two-component system is defined:

$$V^0 = \left(\frac{\partial V}{\partial n_{\text{solute}}} \right)_{T, P, n_{\text{solvent}}} \quad (14.18)$$

where *n* is the number of moles and *V* is the volume of the solution at a given temperature (keeping thermal contributions constant), a given pressure *P*, and a given number of moles of solvent.⁷⁾ V^0 is equivalent to the first pressure derivative of the chemical potential of the solute [76].

What is determined directly in volumetric experiments is not V^0 but rather the “apparent” partial molar volume, V_{app}^0 , a quantity that equals V^0 only at infinite dilution since when the concentration is expressed in molality (*m*), we have [77]:

$$V^0 = V_{\text{app}}^0 + m \frac{\partial V_{\text{app}}^0}{\partial m} \quad (14.19)$$

Thus, experimental partial molar volumes are obtained through extrapolation to infinite dilution.

Partial molar volumes at infinite dilution satisfy *group additivity* (see, for example, Refs [77–79] and the literature cited therein). This additivity parallels that of atomic properties expressed in Equation 14.16c.

At infinite dilution, the effect of solute–solute interaction is eliminated and at a given temperature, V^0 is primarily the result of two contributions: (i) a *positive contribution* due to the volume occupied by the electron density of the solute that

7) In this chapter, the solvent is water unless specified otherwise.

excludes the density of the solvent from the space it occupies, the so-called “steric bulk” of the solute ($V_{\text{intrinsic}}^0$), and (ii) a generally *negative contribution* especially in the case of a polar or an ionic solute due to the pull these species exert on a polar solvent like water causing a contraction of the volume of the solution. The negative contribution is termed *electrostriction* ($V_{\text{electrostriction}}^0$). Figure 14.8 presents a simplified cartoon that depicts these two contributions to the partial molar volume.

From these considerations, the partial molar volume of an amino acid may be expressed [78]:

$$V^0 = V_{\text{intrinsic}}^0 + V_{\text{electrostriction}}^0 \quad (14.20a)$$

and, correspondingly, an atomic or a group contribution to the partial molar volume of an amino acid can be written:

$$V^0(\Omega) = V_{\text{intrinsic}}^0(\Omega) + V_{\text{electrostriction}}^0(\Omega) \quad (14.20b)$$

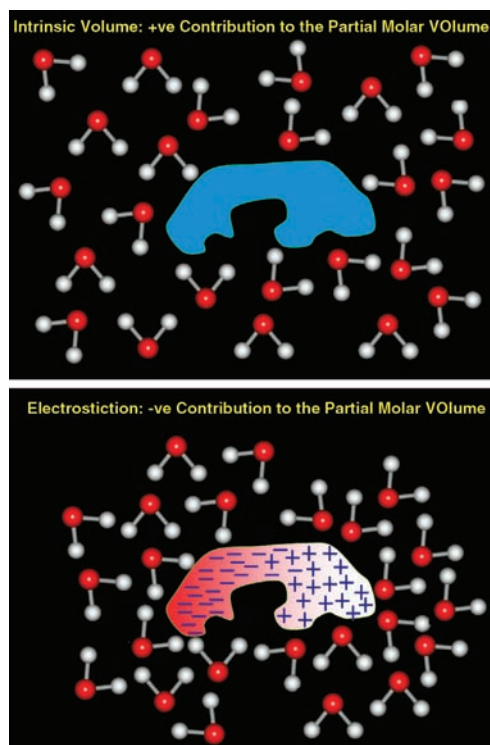


Figure 14.8 Two principal contributions to the partial molar volume. (Top) The intrinsic volume of the solute due to the creation of a cavity in the solvent to accommodate the electron density of the solute. (Bottom) The

negative electrostriction contribution of a polar solute caused by its attraction to the water molecules (note the different orientations of the water molecules in response to the local polarity of the solute molecule).

In the framework of this simple model, the steric (positive) contribution is modeled by the van der Waals volume, that is, the volume occupied by the molecule, group, or atom within the zero-flux boundaries and up to $q = 0.001$ au envelope. The electrostriction (negative) contribution is determined by the local electrostatic field generated by the charge distribution of the amino acid side chain. Both positively and negatively charged atoms in the side chain attract the nearby water molecules, albeit with opposite orientations of the water dipole. Thus, in the following modeling, we take the sum of the unsigned atomic charges as the descriptor that correlates with electrostriction, what is called the *charge separation index* (CSI). The most compelling justification of this simple modeling is that “it works,” and very satisfactorily.

The charge separation index of an amino acid side chain is defined as [80]:

$$\text{CSI} = \sum_{\Omega} |q(\Omega)| \quad (14.21)$$

where $q(\Omega)$ is the charge of atom or group Ω .

The CSI provides an overall measure of the polarity of the side chain with “atomic resolution” (a molecular dipole moment, for example, measures the polarity of the entire molecule but with a coarser “molecular resolution” of the larger the side chain). In view of the small size of the water molecule (a little larger than the size of an oxygen atom), an average local measure of polarity such as the CSI may be more suited than a global measure such as the side chain dipole moment.

Figure 14.9 displays the atomic charges on two neutral side chains: one belonging to a nonpolar amino acid (methionine) and the other to a highly polar one (histidine).

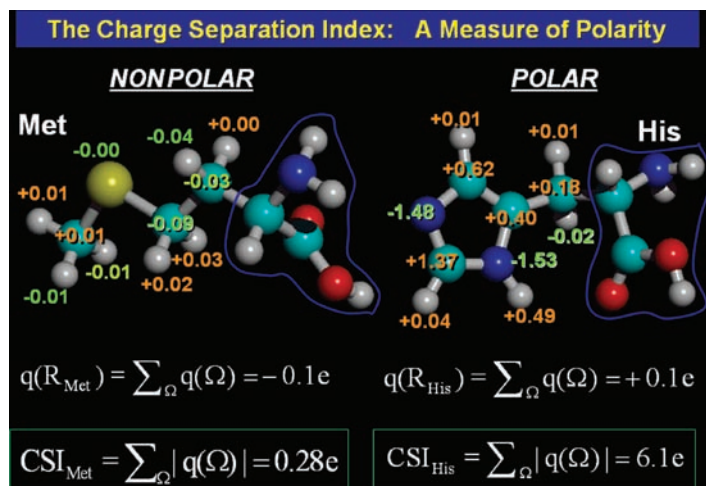


Figure 14.9 Two examples showing how the CSI captures the local polarity of two neutral amino acid side chains: methionine (example of a nonpolar side chain) and histidine (example of a highly polar side chain). While the sum of the

QTAIM atomic charges reveals an overall charge close to zero in both cases, the CSI of the side chain of histidine is more than 20 times that of methionine.

While the sums of the atomic charges on each side chain depart little from electrical neutrality as expected, the sums of the *magnitudes* of these charges, the CSI, clearly distinguish between the polar and the nonpolar side chains (CSI = 6.1 au for the side chain of histidine and only 0.28 au for the side chain of methionine).

The experimental partial molar volumes of the genetically encoded amino acids were obtained from Ref. [77]. These experimental partial molar volumes include contributions from the groups attached to the α -carbon, namely, the $-\text{NH}_2$ and $-\text{COOH}$ groups in their zwitter-ionic forms in addition to the α -hydrogen atom.

The calculated contributions from the groups attached to the α -carbons of the 20 amino acids have been shown to be highly transferable, changing very little from an amino acid to another [9]. Table 14.1 lists the group charges and volumes averaged over the 20 amino acids. A comparison of the magnitudes of the average values of these quantities with the associated standard deviations demonstrates that the spread of the individual values around the mean is very small.

Because the calculated volumes and charges of the α -carbon and its substituents (other than R) are highly transferable, we assume that their total contribution to the partial molar volume is approximately constant across the series of the 20 amino acids. Although we did not carry out calculations on the zwitter-ionic species, we can expect a similar transferability of the properties of the zwitter-ionic form of these groups. On the basis of these observations, the focus of the remainder of this chapter will be exclusively on the side chains R “extracted” from their amino acid.

The modeling of the partial molar volume of the 20 genetically encoded amino acids according to the above assumptions results in the following linear regression equation [9]:

$$V_{\text{AA}}^0 = 37.250 + 0.098 \times V(\text{vdW})_R - 0.884 \times \text{CSI}_R \quad (14.22)$$

$$[r^2 = 0.978, s = 3.887, n = 20]$$

where AA is the acronym of “amino acid”; the subscript R refers to the side chain; $V(\text{vdW})$ is the van der Waal’s volume already defined; CSI is the charge separation index defined in Equation 14.21; r is the linear correlation coefficient that measures the “strength” of the linear tendency of the correlation; s is the estimated standard error, that is, for the normally distributed data, 68 and 95% of the data points lie within $\pm s$ and $\pm 2s$ of the regression line, respectively; and n is the number of amino acids (or amino acid side chains) included in the regression. Naturally, the constants in this equation have the dimensions and units so that the overall equation is “dimensionally homogenous” [81].

Table 14.1 Highly transferable group properties of the α -carbon and its substituents other than the side chain $\text{C}\alpha\text{H}(\text{NH}_2)\text{COOH}$ (data in atomic units).

	COOH		$\text{C}_\alpha\text{H}_\alpha$		NH_2		Total	
	$q(\Omega)$	$\text{vol}(\Omega)$	$q(\Omega)$	$\text{vol}(\Omega)$	$q(\Omega)$	$\text{vol}(\Omega)$	$q(\Omega)$	$\text{vol}(\Omega)$
Average	−0.235	303.8	0.576	87.9	−0.413	173.4	−0.073	565.2
Standard deviation	0.026	3.1	0.020	2.1	0.022	4.5	0.055	6.0

Data from Ref. [9].

In Equation 14.22, the first term accounts for the steric bulk of the side chain and appears with a positive constant, as expected. The second (electrostriction) term is negative as anticipated, which implies that the more polar a side chain is, the more it pulls the surrounding water molecules causing greater contraction. The residual volume represented by the constant ($37.250 \text{ cm}^3 \text{ mol}^{-1}$) is the contribution from the α -carbon and its substituents other than R . The average total volume of this group in its neutral non-zwitterionic form is 565.2 au (Table 14.1). This volume is equivalent to $50.438 \text{ cm}^3 \text{ mol}^{-1}$ which is larger than the residual volume of $37.250 \text{ cm}^3 \text{ mol}^{-1}$. This discrepancy is expected, first since the volume in the Table is strictly the van der Waals contribution without the electrostriction contribution. Furthermore, the experimental values are obtained for amino acids in their zwitter-ionic forms which exhibit significantly more negative electrostriction contributions to partial molar volumes than their non-zwitter counterparts described in Table 14.1.

What is remarkable is the strength of the linear correlation using only these two parameters to fit a dataset of 20 points, this simple model being sufficient to account for 97.8% of the variance in V_{AA}^0 . Table 14.2 compares experimental and calculated partial molar volumes of the 20 genetically encoded amino acids along with the values of $V(\text{vdW})_R$ and CSI_R . The strength of the correlation between calculated and experimental V_{AA}^0 can also be appreciated from Figure 14.10.

Table 14.2 Experimental and calculated partial molar volumes of the free amino acids V_{AA}^0 at infinite dilution in water (25°C) and the terms used in the modeling.

AA	$V_{AA}^0(\text{Experimental})$	$V_{AA}^0(\text{Calculated})$	$V(\text{vdW})_R$	CSI_R
Gly	43.3	41.9	47.25	0.009
Ala	60.5	57.8	212.22	0.221
Ser	60.6	62.0	277.93	2.707
Cys	73.4	76.3	406.30	0.741
Asp ⁽⁻⁾	73.8	79.0	473.10	5.083
Thr	76.9	77.3	435.30	2.805
Asn	78.0	80.5	493.10	5.487
Pro	82.8	81.4	461.70	1.070
Glu ⁽⁻⁾	85.9	93.6	625.23	5.350
Val	90.8	87.3	519.70	0.737
Gln	93.9	95.2	644.87	5.665
His	98.8	96.2	666.83	6.982
Met	105.4	106.9	715.79	0.275
Ile	105.8	101.9	670.83	1.010
Leu	107.8	102.2	674.48	1.019
Lys ⁽⁺⁾	108.5	107.8	760.30	4.175
Phe	121.5	122.5	879.93	0.698
Tyr	123.6	127.3	947.65	2.826
Arg ⁽⁺⁾	127.3	119.1	926.27	9.679
Trp	143.9	146.3	1146.63	3.238

Experimental values were obtained from [77]. V_{AA}^0 are in $\text{cm}^3 \text{ mol}^{-1}$; van der Waals volume and CSI are in atomic units. (Reproduced from Ref. [9] with permission of Wiley-Liss).

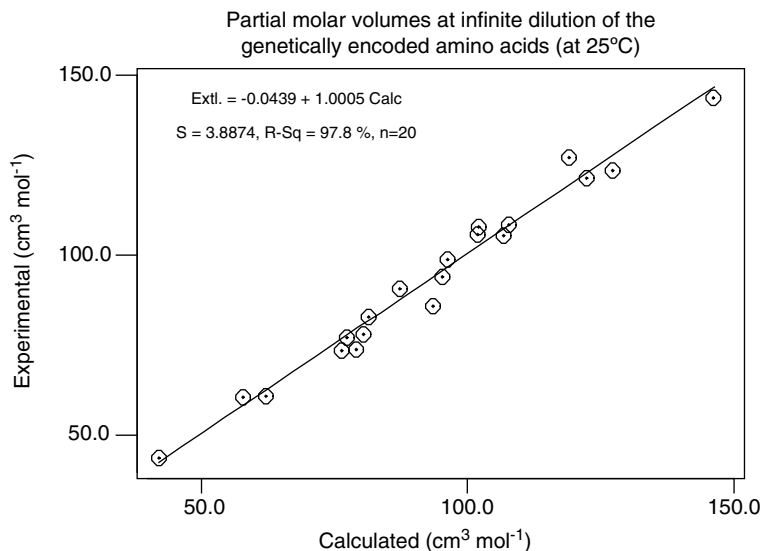


Figure 14.10 Calculated versus experimental partial molar volumes of the genetically encoded amino acids. (Reproduced from Ref. [9] with permission of Wiley-Liss).

The experimentally derived group contributions to the partial molar volume of the amino acids can be linearly correlated to the corresponding calculated group contributions to $V(\text{vdW})$ and CSI , yielding the following regression equation [9]:

$$V_G^0 = -0.925 + 0.127 \times V(\text{vdW})_G - 2.456 \times \text{CSI}_G \quad (14.23)$$

$[r^2 = 0.983, s = 1.122, n = 8]$

where V_G^0 is the experimental group contribution to the partial molar volume, and $V(\text{vdW})_G$ and CSI_G are the sums of the atomic contributions to the van der Waals volume and to the CSI of the group, respectively.

Equation 14.23 accounts for 98.3% of the variance in V_G^0 . If we use $V(\text{vdW})_G$ only as a regressor (and ignore CSI_G), then r^2 drops significantly to only 0.806. This demonstrates the importance of accounting for electrostriction in determining the group contributions to the partial molar volume. The experimental and calculated group contributions to the partial molar volume are listed in Table 14.3 along with CSI_G and $V(\text{vdW})_G$ (Figure 14.11).

Finally, we show that this modeling of the partial molar volume is applicable to molecules of a nature that is significantly different from the amino acids, namely, the nucleic acid bases. The electron densities of the five free nucleic acid bases A, G, C, T, and U, were obtained from density functional theory (DFT) calculations using the B3LYP hybrid functional at the B3LYP/6-311++G(d, p)/B3LYP/6-31+G(d, p) level of theory and subsequently analyzed as described in Section 14.4. The accurate values of the partial molar volumes of these bases in water at two temperatures were obtained from a volumetric study by Lee and Chalikian [82]. These authors report the partial molar volumes of four of the free bases (A, C, T, and U) and those of all five

Table 14.3 Experimental and calculated group contributions to the partial molar volume at infinite dilution in water (V_G^0) and the terms used in the modeling.^{a)}

Group	V_G^0 (Experimental)	V_G^0 (Calculated)	$V(\text{vdW})_G$	CSI_G
–NH–	11.6	10.7	126.83	1.826
–C=O	13.1	12.7	167.78	3.119
–NH ₂	15.4	16.4	175.63	2.013
–CH ₂ –	15.9	17.4	150.49	0.309
–COOH	25.8	25.5	306.26	5.077
–CH ₃	26.5	25.5	213.79	0.277
–CH ₂ OH	28.2	27.7	277.93	2.707
–CONH ₂	28.8	29.6	344.49	5.363

a) Experimental values were obtained from Ref. [77]. Partial molar volumes are in $\text{cm}^3 \text{mol}^{-1}$; van der Waals volume and CSI are in atomic units. (Reproduced from Ref. [9] with permission of Wiley-Liss).

nucleosides, that is, bases attached to the ribose sugar. We compared the reported partial molar volumes of the bases with the corresponding partial molar volumes of the nucleosides to obtain an (indirect) experimental estimate of the contribution of the sugar to the partial molar volume by difference, that is,

$$V_{\text{sugar}}^0 \approx V_{\text{nucleoside}}^0 - V_{\text{base}}^0 \quad (14.24)$$

The contribution of the sugar to the partial volume of the nucleosides of U, C, T, and A that we estimate according to the approximate Equation 14.24 exhibits a

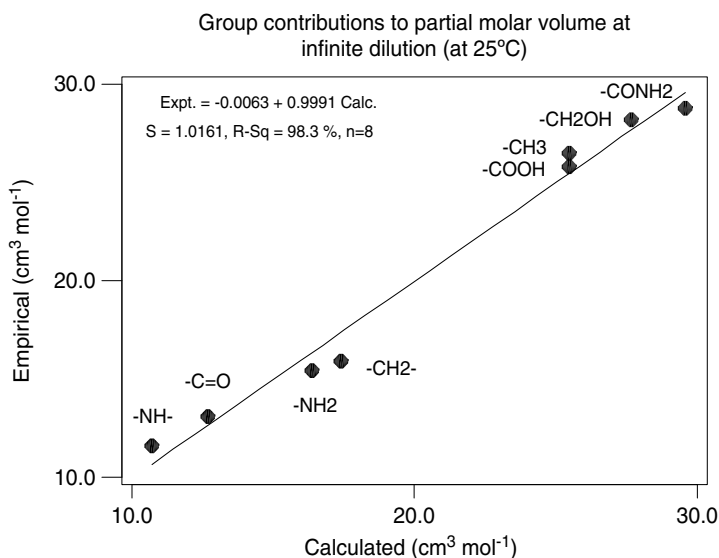
**Figure 14.11** Calculated versus experimental additive group contributions to the partial molar volumes. (Reproduced from Ref. [9] with permission of Wiley-Liss).

Table 14.4 Experimental and calculated partial molar volumes of nucleic acid bases, experimental partial molar volumes of nucleosides, and estimates of the contribution of the ribose sugar to the partial molar volume of the nucleosides.

Molecule	$V^0(18^\circ\text{C})$ (Experimental)	$V^0(18^\circ\text{C})$ (Calculated)	V^0 (55 °C)	V^0 (55 °C) (Calculated)
Experimental ^{a)}				
Free bases				
Uracil	70.2 ± 0.4	69.1	74.6 ± 0.5	74.7
Cytosine	72.4 ± 0.4	74.0	75.8 ± 0.5	79.1
Thymine	86.3 ± 0.4	86.0	91.6 ± 0.5	90.3
Adenine	88.0 ± 0.4	87.5	93.5 ± 0.5	92.2
Guanine ^{b)}	91.6 ± 1.0 ^{b)}	91.8	97.5 ± 1.2 ^{b)}	96.6
Nucleosides				
Uridine	150.7 ± 0.6		154.8 ± 0.7	
Cytidine	152.2 ± 0.6		156.4 ± 0.7	
Thymidine	166.4 ± 0.6		170.4 ± 0.7	
Adenosine	169.2 ± 0.6		175.3 ± 0.7	
Guanosine	172.0 ± 0.6		177.8 ± 0.7	
Differences ^{c)}				
Sugar (U)	80.5 ± 1.0		80.2 ± 1.0	
Sugar (C)	79.8 ± 1.0		80.6 ± 1.0	
Sugar (T)	80.1 ± 1.0		78.8 ± 1.0	
Sugar (A)	81.2 ± 1.0		81.8 ± 1.0	
Sugar (average)	80.4 ± 1.0		80.4 ± 1.0	

a) Experimental data, except for guanine, are obtained from Ref. [82].

b) Estimated from Equation 14.25 (see text).

c) Calculated from Equation 14.24.

remarkable transferability, being practically identical within experimental uncertainties (Table 14.4). The average of this contribution over the available experimental data is $80.4 \pm 1.0 \text{ cm}^3 \text{ mol}^{-1}$, a value that remains constant within the experimental and averaging uncertainties at the two considered temperatures. In view of the high transferability of the group contribution of the sugar to the partial molar volume, we estimate the partial molar volume of guanine (which is not reported in the experimental paper [82]) from the approximation:

$$V_{\text{guanine}}^0 \approx V_{\text{guanosine}}^0 - \langle V_{\text{sugar}}^0 \rangle \quad (14.25)$$

where the last term is the average partial molar volume of the sugar ($80.4 \pm 1.0 \text{ cm}^3 \text{ mol}^{-1}$).

The experimental and estimated values of the partial molar volumes at infinite dilutions of the five nucleic acid free bases in aqueous solutions collected in Table 14.4 were correlated with the two “predictors,” namely, the van der Waals volume $V(\text{vdW})$ and the CSI. The regression equations corresponding to the two experimental temperatures are:

$$V_{\text{base}}^0(18^\circ\text{C}) = 0.09508 \times V(\text{vdW})_{\text{base}} - 1.25304 \times \text{CSI}_{\text{base}} \quad (14.26)$$

$$[r^2 = 0.980, s = 2.527, n = 5(\text{A, T, U, G, C})]$$

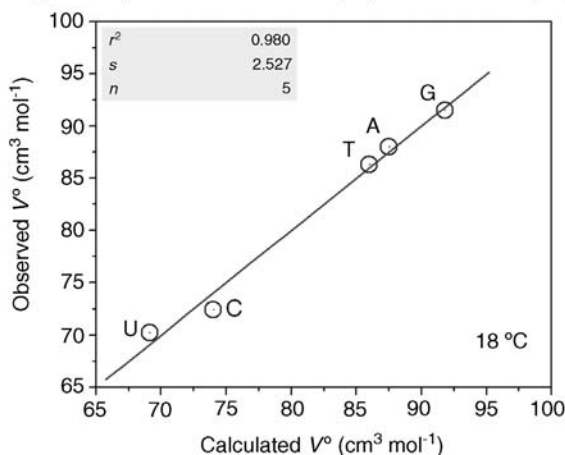
and

$$V_{\text{base}}^0(55^\circ\text{C}) = 0.10050 \times V(\text{vdW})_{\text{base}} - 1.30930 \times \text{CSI}_{\text{base}} \quad (14.27)$$

$$[r^2 = 0.967, s = 4.895, n = 5(\text{A, T, U, G, C})]$$

Figure 14.12 displays the correlation between the experimental and calculated values of the partial molar volumes of the five nucleic acid bases at the two temperatures.

(a) $V^\circ (\text{cm}^3 \text{mol}^{-1}) = 0.0950803 \cdot \text{Vol001 (au)} - 1.253039 \cdot \text{CSI (au)}$



(b) $V^\circ (\text{cm}^3 \text{mol}^{-1}) = 0.1005027 \cdot \text{Vol001 (au)} - 1.309296 \cdot \text{CSI (au)}$

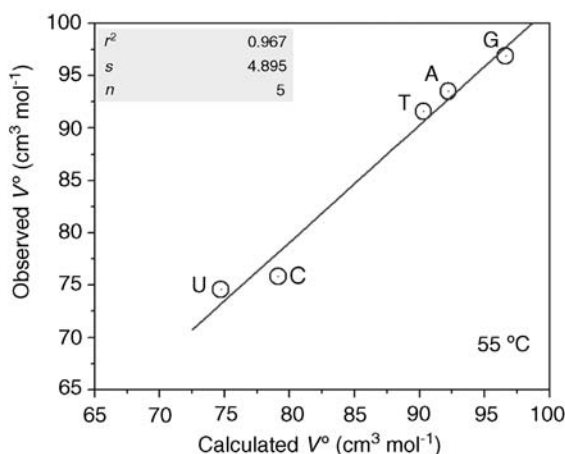


Figure 14.12 Calculated versus experimental partial molar volumes at infinite dilution (in $\text{cm}^3 \text{mol}^{-1}$) of the five nucleic acid bases [adenine (A), guanine (G), cytosine (C), thymine (T), and uracil (U)] in water at (a) 18 °C and (b) 55 °C.

Thermally induced vibrations contribute a (positive) “thermal volume” to the partial molar volume since it increases the size of the cavity created by the solute in the solvent [82]. The increase in the partial molar volumes of the five bases when the temperature is increased from 18 °C (Figure 14.12a) to 55 °C (Figure 14.12b) is very well reproduced by the simple model with a shift of the entire regression line (from U to G) to higher values of the partial molar volumes at the higher temperature. These data are currently being examined in detail in our group and will be published in the future elsewhere. The point stressed here is that the simple model based on $V(\text{vdW})$ and CSI has a wide range of applicability and appears to capture much of the physics determining partial molar volumes.

14.5.2

Free Energy of Transfer from the Gas to the Aqueous Phase

The relative water affinities of the amino acids are of paramount importance in determining the tertiary structure of proteins in solution [83]. An experimental measure of the affinity of a solute to a solvent is provided by the molar free energy of transfer of the solute from the gas phase to the aqueous phase, a quantity termed the molar free energy of hydration.

Wolfenden *et al.* [84] proposed a physicochemical scale to rank the amino acid side chains on the basis of their water affinities. The free energy of transfer from the gas phase to the aqueous phase has been measured at 25 °C for the side chains of the amino acids capped with a hydrogen (i.e., $R\text{-H}$, where H replaces the α -carbon) and the corrected values to pH 7 were termed their “hydration potential” (ΔG_{hydr}) [84].

A single parameter (predictor), the charge separation index defined in Equation 14.21, accounts for 93.6% of the variance in ΔG_{hydr} of the amino acid side chain analogues according to the following linear regression equation:

$$\begin{aligned}\Delta G_{\text{hydr}} &= 1.832 - 2.237 \times \text{CSI}_R \\ [r^2 &= 0.936, s = 1.599, n = 19]\end{aligned}\tag{14.28}$$

The experimental and calculated ΔG_{hydr} are listed along with CSI_R in Table 14.5 (the table is sorted in order of increasing experimental ΔG_{hydr} that is, of decreasing hydrophilicity according to this criterion). The table as well as Figure 14.13 both show the strong agreement between the experimental and calculated values of the hydration potentials. It is noteworthy that a single parameter, the CSI, appears to capture much of the physics of a *free energy*, itself being the sum of an enthalpic term (ΔH_{hydr}) and entropic term ($-T\Delta S_{\text{hydr}}$).

14.5.3

Simulation of Genetic Mutations with Amino Acids Partition Coefficients

Proteins fold in aqueous environment in a manner that tends to minimize the exposure of hydrophobic amino acids to the solvent while maximizing the exposure

Table 14.5 Experimental and calculated free energy of transfer (ΔG_{hydr}) from the gas phase to the aqueous phase of the amino acid side chains capped with a hydrogen atom, along with the CSI values used in the modeling.^{a)}

R-H	ΔG_{hydr} (Experimental)	ΔG_{hydr} (Calculated)	CSI _(R)
Arg ⁽⁺⁾	−19.92	−19.82	9.679
Asp ^(−)	−10.95	−9.54	5.083
His	−10.27	−11.87	6.124
Glu ^(−)	−10.20	−10.84	5.350
Asn	−9.68	−10.44	5.487
Lys ⁽⁺⁾	−9.52	−7.51	4.175
Gln	−9.38	−10.13	5.665
Tyr	−6.11	−4.49	2.826
Trp	−5.88	−5.41	3.238
Ser	−5.06	−4.22	2.707
Thr	−4.88	−4.44	2.805
Met	−1.48	1.22	0.275
Cys	−1.24	0.17	0.741
Phe	−0.76	0.27	0.698
Ala	1.94	1.34	0.221
Val	1.99	0.18	0.737
Ile	2.15	−0.43	1.010
Leu	2.28	−0.45	1.019
Gly	2.39	1.81	0.009

a) ΔG_{hydr} is in kcal mol^{−1} and CSI in atomic units. Experimental values were determined by Wolfenden *et al.* [84]. (Reproduced from Ref. [9] with permission of Wiley-Liss).

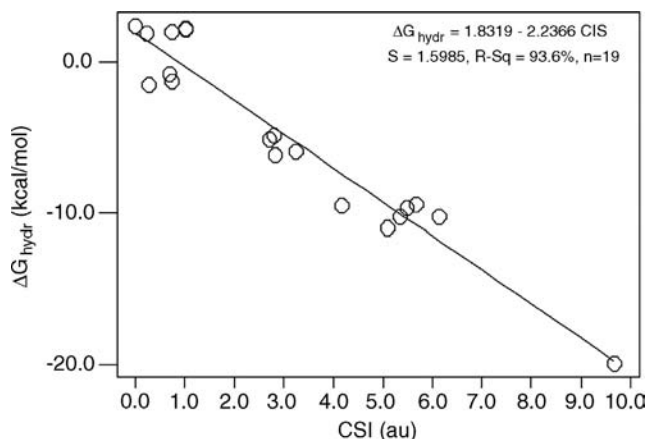


Figure 14.13 Experimental free energy of transfer from the gas phase to the aqueous phase (ΔG_{hydr}) of the genetically encoded amino acids side chains capped with a hydrogen

atom (in kcal mol^{−1}) plotted against the corresponding charge separation index (CSI_R) (in atomic units). (Reproduced from Ref. [9] with permission of Wiley-Liss).

of hydrophilic residues. Denaturation may result in the unfolding of the protein and the exposure of previously buried residues to the aqueous phase [85].

Sharp *et al.* [86] studied the *change* in the protein unfolding energies induced by single point mutations, that is, upon replacement of one amino acid by another at the same site in the polypeptide chain. They found a strong correlation between the change in the protein unfolding energy upon amino acid substitution and the difference in the solvent-to-solvent free energy of transfer of the H-capped side chain of the wild-type amino acid of that of the mutant.

Radzicka and Wolfenden [87] reported $\Delta G_{\text{cyclohexane} \rightarrow \text{water}}$ and $\Delta G_{\text{octanol} \rightarrow \text{water}}$ for 19 and 17 H-capped amino acid side chains, respectively. These are the free energies of transfer of the H-capped amino acid side chains between cyclohexane and water (a model for nonpolar-to-polar mutation) and between octanol and water (a model for polar-to-polar mutation). We have used these experimental values to construct two corresponding difference matrices of all possible $\Delta\Delta G(\text{AA1-AA2}) = \Delta G(\text{AA1}) - \Delta G(\text{AA2})$, where AA1 and AA2 are amino acids 1 and 2, respectively. Each one of these two difference matrices is antisymmetric, that is, the matrix element a_{ij} is equal to the negative of the matrix element a_{ji} . The (upper or lower) triangular part of the cyclohexane–water matrix includes 191 elements and that of the octanol–water matrix includes 153 values. We have similarly constructed three corresponding difference matrices of theoretically calculated quantities, the elements of which are: (1) the differences between all pairs of side chains CSIs (ΔCSI_R), (2) the magnitudes of the difference in the total unsigned net charge of the side chains ($|\Delta q_{(R)}|$), and (3) the differences of their van der Waals volumes [$\Delta V(\text{vdW})_R$].

The experimental $\Delta G_{\text{cyclohexane} \rightarrow \text{water}}$ and $\Delta G_{\text{octanol} \rightarrow \text{water}}$ were then fitted to the following linear regression models using the elements of the theoretically calculated matrices:

$$\begin{aligned} \Delta\Delta G_{\text{cyclohexane} \rightarrow \text{water}} = & -0.308 - \underbrace{(1.750 \times \Delta\text{CSI}_R + 3.910 |\Delta q_R|)}_{\text{Electrostatic term}} \\ & + \underbrace{0.0057 \times \Delta V(\text{vdW})_R}_{\text{Intrinsic volume term}} \quad [r^2 = 0.890, s = 1.873, n = 191] \end{aligned} \quad (14.29)$$

and

$$\begin{aligned} \Delta\Delta G_{\text{octanol} \rightarrow \text{water}} = & 0.272 - \underbrace{(0.192 \times \Delta\text{CSI}_R + 1.260 \times |\Delta q_R|)}_{\text{Electrostatic term}} \\ & + \underbrace{0.0024 \times \Delta V(\text{vdW})_R}_{\text{Intrinsic volume term}} \quad [r^2 = 0.783, s = 0.427, n = 153] \end{aligned} \quad (14.30)$$

In these two equations, $\Delta\Delta G$ is written as a sum of a negative electrostatic term (the change in the side chain CSI and the change in the unsigned charge) and a positive

intrinsic volume term that represents the change in the van der Waals volume of the side chains upon mutating amino acid 1 to amino acid 2.

Figure 14.14 displays plots of the calculated and experimental values of $\Delta G_{\text{cyclohexane} \rightarrow \text{water}}$ and $\Delta G_{\text{octanol} \rightarrow \text{water}}$. The two regression Equations 14.29 and 14.30 have the desirable small ratio of adjustable parameters (three) to data points (191 and 153, respectively). In addition, the elevated values of r^2 indicate a strong linear correlation despite this small parameter/data ratio. The strength of the correlation can also be visually appreciated from Figure 14.14.

The reader may have noticed the smaller numerical values of all the coefficients in Equation 14.30 compared to Equation 14.29, which reflects the much more drastic environmental change felt by the solute (the amino acid H-capped side chain) upon its transfer from cyclohexane to water (nonpolar to polar) in comparison with its transfer from octanol to water (polar to polar). Furthermore, the electrostatic term contributes relatively more to the $\Delta\Delta G$ in partitioning from a nonpolar-to-polar transfer, that is, in Equation 14.29.

14.5.4

Effect of Genetic Mutation on Protein Stability

Oligonucleotide-directed mutagenesis was used by Shortle *et al.* [88] to systematically introduce single-site mutations in an effort to elucidate the effect of amino acid substitutions on the stability of a representative protein, namely, staphylococcal nuclease. Using this technique, these researchers have prepared a series of mutant proteins, each mutant having a single amino acid substitution of a wild-type residue to either glycine or alanine.

Guanidine hydrochloride was then used to reversibly denature the mutant proteins. The equilibrium constants between the folded and denatured protein were determined for the wild-type protein and all of its mutants [88]. The ratio of the equilibrium constant of the wild-type protein to that of a mutant protein can then be used to evaluate the change in the free energy of denaturation upon mutation ($\Delta\Delta G$), which is given by $RT \ln(K_{\text{wild type}}/K_{\text{mutant}})$.

Since a given amino acid residue generally occurs at more than one site in the sequence of staphylococcal nuclease, Shortle *et al.* averaged the values of the change in protein stability upon mutating this residue over several different sites in which it occurs in the sequence. A similar averaging was repeated for all the amino acids considered in their study. Consequently, and as they explain in their paper, the experimental uncertainty they report reflects experimental errors as well as the variance in the data due to the differences in the local environment of a given amino acid residue at different locations in the amino acid sequence of the polypeptide chain [88]. The authors estimate that these “environmental effect” rather than experimental uncertainties are the lead contributor to the experimental uncertainty (depicted as “error bar” in Figure 14.15).

The average change in the stability of staphylococcal nuclease upon a single point mutation can be fitted to the following linear regression equation:

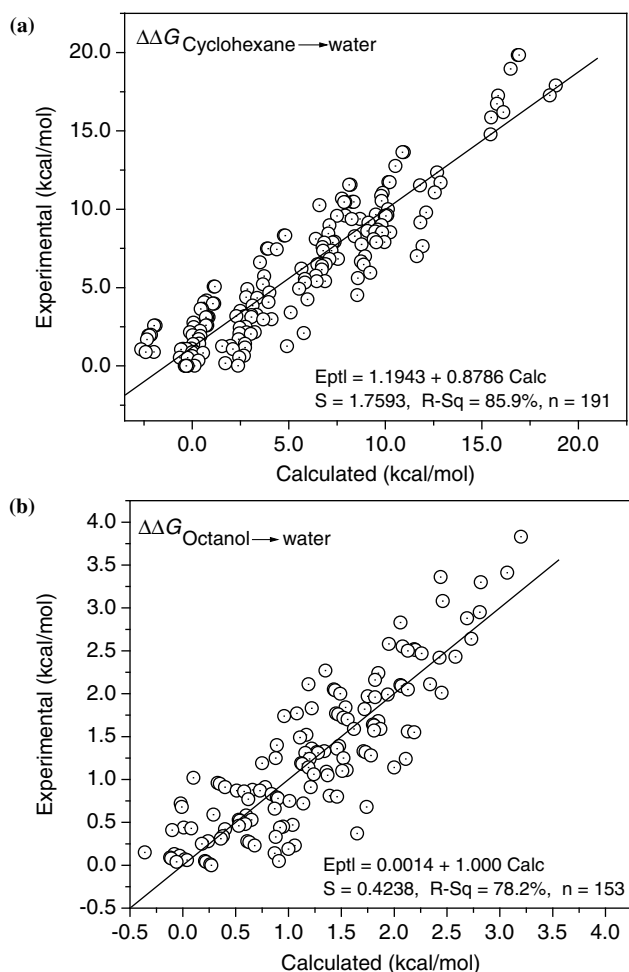


Figure 14.14 Experimental versus calculated. (a) Difference of the free energy of transfer of all pairs of H-capped amino acid side chains from cyclohexane to water. (b) Difference of the free energy of transfer of all pairs of H-capped amino acid side chains from octanol to water. (All data in kcal mol^{-1}). (Adapted from Ref. [9] with permission of Wiley-Liss).

$$\begin{aligned}
 \Delta\Delta G_{\text{mut} \rightarrow \text{wt}} &= \Delta G_{\text{wt}} - \Delta G_{\text{mut}} \\
 &= 0.278 + 1.840 \times \Delta \text{CSI}_R + 0.004 \times \Delta V(\text{vdW})_R \\
 [r^2 &= 0.829, s = 0.4377, n = 10]
 \end{aligned}
 \tag{14.31}$$

where $\Delta \text{CSI}_R = \text{CSI}_{R(\text{wt})} - \text{CSI}_{R(\text{mut})}$, $\Delta V_R = V_{R(\text{wt})} - V_{R(\text{mut})}$, and $\Delta\Delta G_{\text{mut} \rightarrow \text{wt}}$ is the average change in the stability of staphylococcal nuclease upon mutation. Figure 14.15 displays the correlation between the experimental and calculated changes in the stability of staphylococcal nuclease.

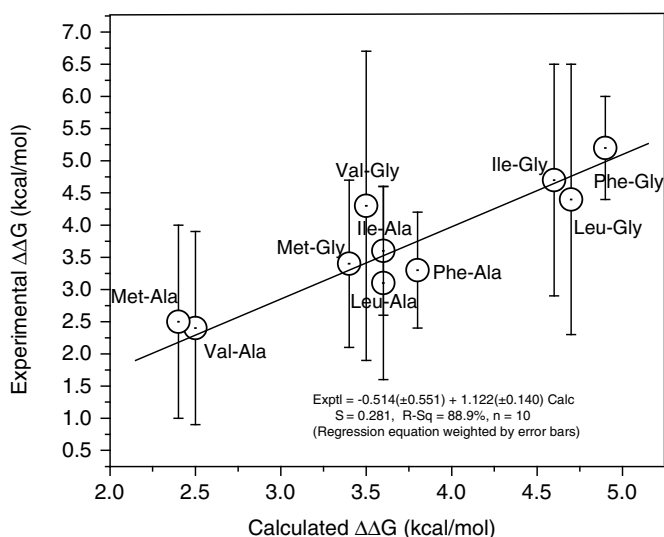


Figure 14.15 Experimental versus calculated (Equation 14.31) change of staphylococcal nuclease stability upon mutations of the type: nonpolar $\rightarrow$ nonpolar. The bars indicate the uncertainties due to variation in the

microenvironment of the amino acid residue in addition to experimental uncertainty as explained in the text. (All data in kcal mol^{-1}). (Reproduced from Ref. [9] with permission of Wiley-Liss).

Equation 14.31 and Figure 14.15 exclude two outliers, namely, Tyr $\rightarrow$ Ala and Tyr $\rightarrow$ Gly, since the $\Delta\Delta G$ values for these two mutations exhibit significant experimental uncertainties of magnitudes comparable to that of the respective $\Delta\Delta G$ values themselves. These uncertainties have been attributed to large differences in the local environment of tyrosine residues within this protein [88].

Guerois *et al.* [89] use a database as a “training set” to optimize a thermodynamical model of ΔG of unfolding. ΔG is modeled as a sum of a van der Waals terms, terms describing the difference in solvation energy for residues on going from the unfolded to the folded state, terms to account for hydrogen bonding with water, an electrostatic term, and entropic terms. The optimization of the weights of each term has been achieved using a training set of 339 mutants in 9 proteins. The resulting model achieved $r = 0.83$ for the database of 1030 mutants [89].

We present here an alternative approach in which the $\Delta\Delta G$ of folding upon mutation is correlated to the corresponding changes in the two descriptors (ΔCSI_R and $\Delta V(\text{vdW})_R$), since these two descriptors have been found to be highly correlated to several other physicochemical properties of the amino acid residues as described above.

Using differential scanning calorimetry (DSC), Loladze, Ermolenko, and Makhatadze [90] reported $\Delta\Delta G$ of unfolding of ubiquitin, but they were also able to measure the enthalpic and entropic terms ($\Delta\Delta H$ and $-T\Delta\Delta S$) as well. We have found a strong statistical correlation between $\Delta\Delta G$ of unfolding of ubiquitin and ΔCSI_R , a single descriptor used in the modeling. Furthermore, the individual entropic

and enthalpic terms are highly correlated to the sum of ΔCSI_R and its square $(\Delta\text{CSI}_R)^2$. Interestingly, the coefficient of the $(\Delta\text{CSI}_R)^2$ terms in the regression equation of the enthalpic component has an almost identical magnitude as (but opposite sign of) the corresponding coefficient in the regression equation of the entropic component. The sum of the enthalpic and entropic components results in an almost complete cancellation of the quadratic term yielding a simple linear equation as a result:

$$\begin{aligned}\Delta\Delta H_{\text{wt} \rightarrow \text{mut}} &= \Delta\Delta H_{\text{mut}} - \Delta\Delta H_{\text{wt}} = -5.6711 + 2.1578\Delta\text{CSI} - 0.2504\Delta\text{CSI}^2 \\ [r^2 &= 0.908, s = 2.5360, n = 12]\end{aligned}\quad (14.32)$$

$$\begin{aligned}-T\Delta\Delta S_{\text{wt} \rightarrow \text{mut}} &= -T(\Delta\Delta S_{\text{mut}} - \Delta\Delta S_{\text{wt}}) = 3.2068 - 1.7252\Delta\text{CSI} + 0.2623\Delta\text{CSI}^2 \\ [r^2 &= 0.867, s = 2.5639, n = 12]\end{aligned}\quad (14.33)$$

$$\begin{aligned}\Delta\Delta G_{\text{wt} \rightarrow \text{mut}} &= \Delta\Delta G_{\text{mut}} - \Delta\Delta G_{\text{wt}} = \Delta\Delta H - T\Delta\Delta S \\ &= -2.3499(-2.4643) + 0.4361(0.4326)\Delta\text{CSI} + 0.0000(0.0190)\Delta\text{CSI}^2 \\ [r^2 &= 0.883, s = 0.5416, n = 12]\end{aligned}\quad (14.34)$$

where quantities related to the wild-type and mutant proteins are subscripted by wt and mut, respectively.

What is labeled as Equation 14.34 is actually two equations: in the parentheses are the constants obtained by summing Equations 14.32 and 14.33 and outside the parentheses are the constants obtained by direct regression of the ΔCSI and the experimental $\Delta\Delta G$ values when a linear model is assumed (i.e., the coefficient of the quadratic term is assumed to be exactly zero from the start).

Figure 14.16a and b displays the relationships between the enthalpic and the entropic components and ΔCSI , and Figure 14.16c exhibits the correlation between the experimental $\Delta\Delta G$ and ΔCSI with an assumed linear model.

14.5.5

From the Genetic Code to the Density and Back

Nirenberg *et al.* [91] remarked that amino acids with similar polarities generally have codons that are similar in their base composition, whether mainly purines or pyrimidines. The genetic code is moderately degenerate at the first position and highly degenerate at the third position. The second position, however, is nondegenerate with the single exception of serine. In other words, the middle letter of the code is always the same in all the synonyms that encode a particular amino acid (except in the case of serine). Later, Alff-Steinberger [92] noted that the substitution of the first position results, in general, in another amino acid with physical properties that are similar to the original amino acid. From these considerations, it has long been recognized that the second position of the codon is the most important in

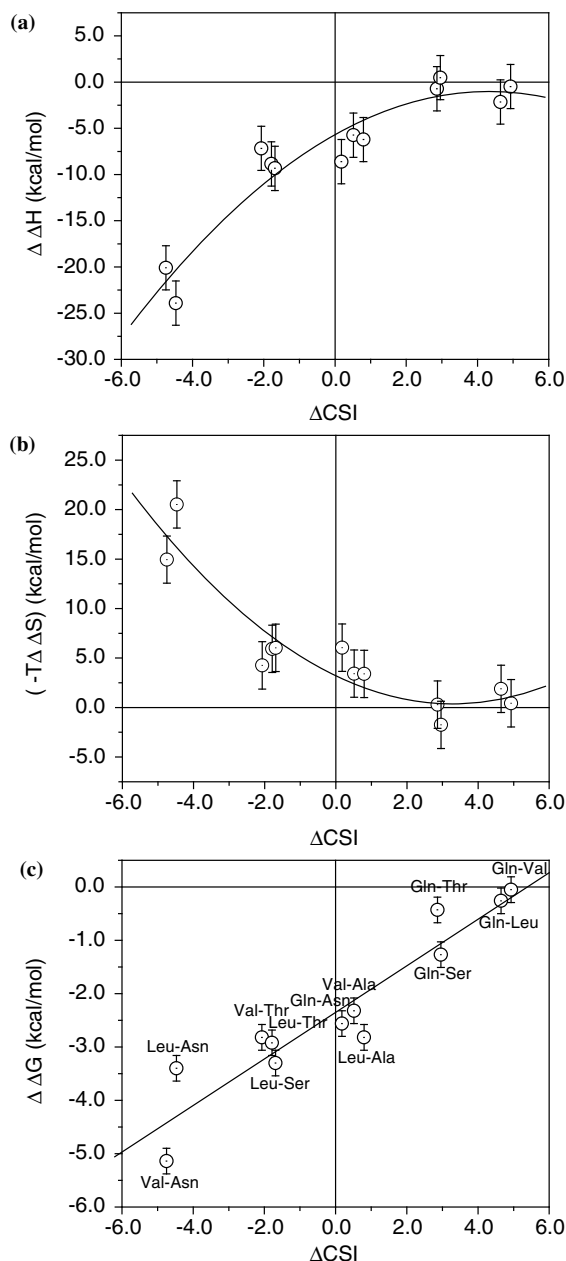


Figure 14.16 (a) Correlation between experimental $\Delta\Delta H$ upon single point mutation and ΔCSI . The fitted line is given by Equation 14.32. (b) Correlation between experimental $-T\Delta\Delta S$ upon single point mutation and ΔCSI . The fitted line is given by Equation 14.33. (c) Correlation between experimental $\Delta\Delta G$ upon single point mutation and ΔCSI . The fitted line is given by

Equation 14.34 taking the values of the constants outside the bracket. The amino acid residue substitutions are indicated by aa1–aa2, where aa1 is the amino acid residue in the wild type and aa2 in the substituent residue in the mutant. (Energies are in kcal mol^{-1} , CSI is in atomic units). (Reproduced from Ref. [9] with permission of Wiley-Liss).

determining the physical properties of the encoded amino acid [92]. Therefore, a mutation at the second place is expected to have more significant consequences on the structure and function of the mutant protein. Further, Alff-Steinberger noticed that similar amino acids tend to exhibit similarities in the chemical class to which the nucleic acid bases belong (purine or pyrimidine) at a given position in the triplet code [92]. The resilience of the genetic code can partly be understood in the light of these observations since mutations that interchange a purine and a pyrimidine are far less common or likely than ones that interchange a base within the same chemical class [92].

Wolfenden *et al.* [84] sorted the 20 genetically encoded amino acids on the basis of increasing water affinity of their side chains (as measured by their respective hydration potentials). A strong bias has emerged from this sorting: in the mRNA version of the genetic code, hydrophilic amino acids have a purine base in the second position (A or G) while hydrophobic amino acids tend to have a pyrimidine base (U or C) in the second position (see Figure 14.17). Since the hydrophilicity/hydrophobicity of the amino acid side chains is highly correlated with the CSI (as are several other physicochemical properties), one can anticipate that this descriptor may provide a direct link between the genetic code and the electron density of the encoded amino acid.

Table 14.6 lists the CSI values of the side chains of 19 genetically encoded amino acids and the second letter in their mRNA codon. Glycine is the only amino acid that has been excluded as it lacks a side chain. The amino acid side chains are listed in order of increasing CSI_R . The table reveals that all amino acid side chains with $0.22 < CSI_R \leq 2.81$ au, listed in the top part of the table, are hydrophobic and all possess a *pyrimidine* in the second position, cysteine being the only exception. The bottom part of the table (side chains with $2.81 \leq CSI_R \leq 9.68$ au) includes polar hydrophilic amino acids. All the amino acids with $CSI_R \geq 2.826$ atomic units possess a *purine* base at the second position, predominantly adenine.

It does not appear coincidental that serine, the only degenerate amino acid in the second position, falls almost on the borderline between the two groups. These findings support the hypothesis that the operation of the genetic code is dominated by the polarity of the amino acid side chain as determined by the second letter of the codon in mRNA.

14.6

Molecular Complementarity⁸⁾

The stereochemical code of Lattman and Rose [12] embodied in the amino acid sequence is brought into life by molecular complementarity, the determinant of protein folding. To fit as a lock and key, two molecules must satisfy two types of complementarity [93]:

⁸⁾ This section is a slightly edited reproduction from Ref. [9] with permission of Wiley-Liss.

PYRIMIDINES				PURINES				
	U		C		A		G	
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
	UUC		UCC		UAC		UGC	
	UUA		UCA		UAA		UGA	
	UUG		UCG		UAG		UGG	
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
	CUC		CCC		CAC		CGC	
	CUA		CCA		CAA		CGA	
	CUG		CCG		CAG		CGG	
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
	AUC		ACC		AAC		AGC	
	AUA		ACA		AAA		AGA	
	AUG		ACG		AAG		AGG	
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
	GUC		GCC		GAC		GGC	
	GUA		GCA		GAA		GGA	
	GUG		GCG		GAG		GGG	

Figure 14.17 The mRNA genetic code. The first letter is to be read from the column on the left, the second letter from the top row, and the third from the rightmost column. The left half of the table has a pyrimidine as the second letter and the encoded amino acids in this half are all nonpolar. The half of the table to the right

includes codons with a purine as the second letter and all the amino acids encoded in this half have polar side chains with the exception of glycine (in which the side chain is a hydrogen atom) and cysteine. Note that the middle letter is always the same for any given amino acid in all its synonyms with the exception of serine, the only amino acid that exhibits degeneracy in the second position of its codon.

- 1) *van der Waals complementarity* which is determined by the size and shape of the atoms or groups that are brought into contact. This type of complementarity is particularly important when nondirectional van der Waals or dispersion forces are dominant, as happens, for example, between two aligned hydrocarbon chains in a phospholipid bilayer. The strength of these interactions increases with the area of contact between the interacting molecules. Consequently, the 0.001 or 0.002 au isodensity surfaces that define a molecule's van der Waals shape [94] and the area of the corresponding surface, the former surface for the gas phase and the latter for a condensed phase, are necessary for predicting the relative orientation and resulting strength of the interaction. It is also well documented that the atomic volumes of QTAIM correlate with additive contributions to the molecular polarizability [95] enabling one to use the atomic volumes to obtain quantitative estimates of the strength of such interactions.
- 2) *Lewis complementarity* that is operative when the mating of the molecules is determined by the pairing of acid–base sites or, equivalently, by the pairing of electrophilic and nucleophilic sites. It is Lewis complementarity that determines the pairing of the bases in DNA Watson–Crick base pairs and the recognition between mRNA and tRNA at the ribosome.

Table 14.6 Charge separation index CSI of the genetically encoded amino acids side chains as a basis for the classification of the genetic code.^{a)}

Amino acid	CSI _R (au)	Second letter of the mRNA codon	Chemical nature of the second letter of the mRNA codon
Ala	0.221	C	Pyrimidine
Met	0.275	U	Pyrimidine
Phe	0.321	U	Pyrimidine
Val	0.737	U	Pyrimidine
Cys	0.741	G	Purine
Ile	1.016	U	Pyrimidine
Leu	1.027	U	Pyrimidine
Pro	1.063	C	Pyrimidine
Ser	2.707	C/G	Pyrimidine/purine
Thr	2.805	C	Pyrimidine
Tyr	2.826	A	Purine
Trp	3.238	G	Purine
Lys ⁽⁺⁾	4.175	A	Purine
Asp ⁽⁻⁾	5.083	A	Purine
Glu ⁽⁻⁾	5.350	A	Purine
Asn(II) ^{b)}	5.487	A	Purine
Asn(I) ^{b)}	5.629	A	Purine
Gln	5.665	A	Purine
His(II) ^{b)}	6.092	A	Purine
His(I) ^{b)}	6.124	A	Purine
His ⁽⁺⁾	7.067	A	Purine
Arg ⁽⁺⁾	9.679	G	Purine

a) Reproduced from Ref. [9] with permission of Wiley-Liss.

b) The roman numerals (I) and (II) refer to different conformations (see Ref. [8]).

The topology of the density, while recovering the concepts of atoms, bonded interactions, and chemical structure, gives no indication of the localized bonded and nonbonded pairs of electrons associated with the Lewis model of structure and reactivity. Lewis complementarity may however be understood and predicted in terms of the topology of the Laplacian of the electron density, the quantity $\nabla^2\varrho(\mathbf{r})$. The Lewis model is concerned with the pairing of electrons, information contained in the *electron pair density* and not in the one-electron density defined by Equation 14.3. Remarkably enough however, the essential information regarding the spatial pairing of electrons is contained in the Laplacian of the electron density [96]. The second derivative of a scalar function such as ϱ determines where this function is locally concentrated (where $\nabla^2\varrho(\mathbf{r}) < 0$) and locally depleted ($\nabla^2\varrho(\mathbf{r}) > 0$). Thus, the negative of the Laplacian, the function $L(\mathbf{r}) = -\nabla^2\varrho(\mathbf{r})$, has maxima and minima that indicate where electronic charge is maximally concentrated and depleted, respectively.

$L(\mathbf{r})$ recovers the shell structure of an isolated atom in terms of a corresponding number of alternating pairs of shells of charge concentration (CC) and charge

depletion. The valence shell of charge concentration loses its uniformity when the atom is bonded to other atoms in a molecule. The valence shell in this case exhibits local maxima, that is, local charge concentrations. The number, relative size, and orientation of these CCs provide a faithful mapping of the localized bonded and nonbonded Lewis pairs assumed in the VSEPR model of molecular geometry [97]. Furthermore, it has been demonstrated that the CCs of $L(\mathbf{r})$ denote the number of electron pairs and their positions relative to a fixed position of a reference pair, as determined by the conditional pair density [98]. Thus, the topology of $L(\mathbf{r})$, including its predicted shell structure, provides a mapping of the essential pairing information from six- to three-dimensional spaces and the mapping of the topology of $L(\mathbf{r})$ onto the Lewis and VSEPR models is grounded in the physics of the pair density.

The integral of $L(\mathbf{r})$ over an atomic basin must vanish as a consequence of the zero-flux surface condition, Equation 14.9, and consequently the creation of regions with $L(\mathbf{r}) > 0$ must be coupled with the creation of others with $L(\mathbf{r}) < 0$. Just as the local maxima in $L(\mathbf{r})$ denote concentrations of electronic charge and hence the presence of sites of basicity or equivalently nucleophilicity, so the corresponding “holes” in $L(\mathbf{r})$ denote regions of local depletions in electronic charge, sites characterized by acidic or electrophilic activity. The complementary matching of the reactive surfaces of two molecules determines their relative orientation and mode of attachment. Numerous examples have been given in Ref. [44] wherein the relative orientation of approaching reactants can be predicted from the alignment of their respective maxima and minima in $L(\mathbf{r})$. The same alignment is responsible for the observed packing of crystals, as stated by Koritsanszky and Coppens [19]: “*Analysis of the Laplacian of the electron density shows that molecules pack in a key-lock arrangement in which regions of charge concentration face electron deficient regions in adjacent molecules in crystals.*”

The *reactive surface* of a molecule is defined by $L(\mathbf{r}) = 0$ envelope, the envelope that separates the shells of charge concentration from those of charge depletion. This surface makes clear the locations of the lumps and holes, the nucleophilic and electrophilic sites, respectively. Figure 14.18a illustrates the matching of the reactive surfaces of the guanine–cytosine Watson–Crick base pair where the “holes” in the shells of charge concentration of the three hydrogen atoms that participate in hydrogen bonding are complemented with the localized CCs on the keto oxygen atoms and on a ring nitrogen atom of cytosine. While the “holes” on the hydrogen atoms are not visible in the reactive surface at the resolution used in this display, the program that determines the critical points in $L(\mathbf{r})$ locates a nonbonded (3, +3) critical point on the N–H axis in the valence shell of charge concentration of each hydrogen, as shown in Figure 14.18b. Such a critical point denotes the presence of a local minimum in the shell of charge concentration and is a characteristic feature of hydrogen bonding [99].

Figure 14.19 displays the reactive surfaces defining the sites of electrophilic and nucleophilic attachment in the three amino acids: Arg⁽⁺⁾, Glu⁽⁻⁾, and His⁽⁺⁾. These sites determine the initial interaction of the amino acid with the enzyme responsible for the acylation of its amino group, the step prior to its esterification to form tRNA. Luger and coworkers have determined experimentally the reactive surfaces of Asn·H₂O, Glu·H₂O, Lys·HCl, Pro·H₂O, Ser, and Val [37, 39, 40]. The dominant

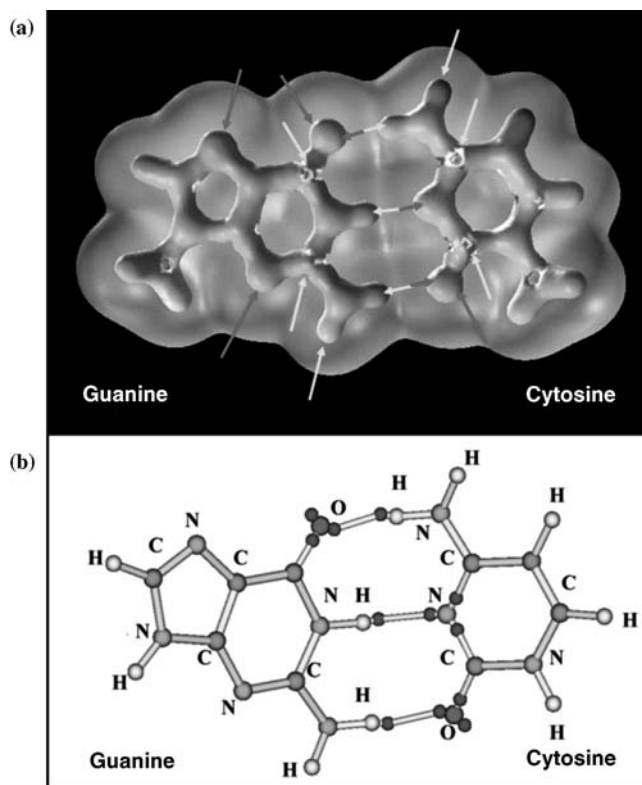


Figure 14.18 (a) A display of the zero-isodensity surface of the Laplacian ($L(r)=0$), the reactive surface, of a guanine–cytosine Watson–Crick base pair encased in a semitransparent $q=0.001$ au isodensity van der Waals envelope with N9 of guanine and N1 of cytosine capped by methyl groups. The light gray arrows denote the locations of “holes,” the sites of nucleophilic attack, and the dark gray arrows indicate the “lumps” of charge concentrations, the sites of electrophilic attack. (b) A molecular graph for the guanine–cytosine base pair with N9 of guanine and N1 of cytosine capped with hydrogen atoms. The location of

the $(3, +3)$ critical points, the centers of charge depletion within the valence shell charge concentrations (VSCCs) of the hydrogens linking the two base pairs, are indicated by small dots. The trio of bonded and nonbonded charge concentrations (CCs) on a pyrimidine nitrogen involved in the hydrogen bonding and each of the keto oxygen atoms are indicated by similar dots. The $(3, +3)$ critical point located on the nonbonded side of a hydrogen atom is linked to a CC on the N or O receptor atom by a bond path defining the hydrogen bond that is noticeably bent in each case. (Adapted from Ref. [9] with permission of Wiley-Liss).

common features to the reactive surface of amino acids include pairs of nonbonded CCs on the oxygen atoms that serve as centers of electrophilic attack; electron-poor regions at the carbon of the carboxyl group (in both neutral and zwitter ions); and a somewhat smaller hole on the α -carbon. The valence shell charge concentrations (VSCCs) of the carboxyl carbon in the side chain of $\text{Glu}^{(-)}$, $\text{C}\delta$, and of the carbon of the guanidino group in $\text{Arg}^{(+)}$, $\text{C}\zeta$, exhibit similar sites open to nucleophilic attack.

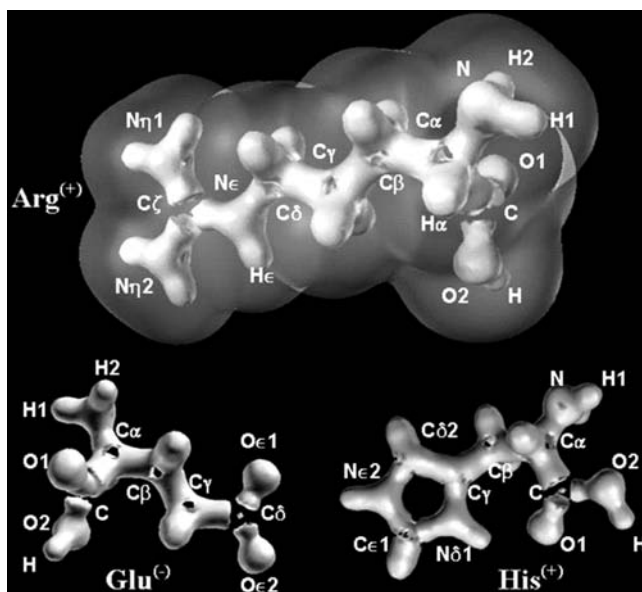


Figure 14.19 Reactive surface maps for the (nonzwitter ionic) amino acids: Arg⁽⁺⁾, Glu⁽⁻⁾, and His⁽⁺⁾. The reactive surface of Arg⁽⁺⁾ is encased in a semitransparent $q = 0.001$ au (van der Waals) envelope. Note the pronounced sites of charge depletion at the carbons of the carboxyl groups and of the amino group. Every saturated carbon exhibits holes in its valence shell charge concentration (VSCC), being most pronounced for the C α atoms. All of the oxygen atoms exhibit a pair of nonbonded charge concentrations (CCs), being particularly evident in the edge-on view of the oxygen atom of the carboxyl O—H group in Glu⁽⁻⁾. Each nitrogen atom of an amino group exhibits a single

nonbonded CC, while the amino group nitrogens have nonbonded CCs located on each side of the plane of the amino group. While the location and number of the nonbonded CCs on the nitrogen and oxygen atoms are as anticipated on the basis of the Lewis model, the Laplacian distribution quantifies the picture by giving the magnitude and hence relative base strength of each CC, together with its precise location. For example, the angle of attack of a nucleophile at the charge depletion of a keto carbon atom from above the plane of the nuclei is determined by the corresponding critical point angle. (Adapted from Ref. [9] with permission of Wiley-Liss).

A display of the experimentally determined Laplacian distribution for a folded protein would offer a clear picture of the operation of the “stereochemical code.” One could in principle map out the reactive surface of an active site in an enzyme by performing a complementary mapping of the substrate’s Laplacian distribution. An attainable goal is to use the Laplacian of the density to follow the complete pathway of the coding and decoding of the genetic information involved in the formation of a polypeptide.

MacDougall and Henze [100, 101] have written and made available a new molecular visualization tool called EVolVis, which is particularly well suited for generating and studying displays of a molecule’s reactive surface defined by the Laplacian of the charge density. They give a number of displays of the reactive surfaces of biological molecules.

14.7

Closing Remarks

This chapter demonstrates that chemical problems in biology can be stated and investigated using a theory based on “observables” such as the electron density. We have shown that one can construct physically reasonable *models* in terms of atomic or group properties that are themselves *model free* and derived directly from the underlying physics. This removes the source of uncertainty in determining the goodness of a model. The *diversity* and *strength* of the statistical correlations reviewed in this chapter adds further evidence to the utility of QTAIM in drug and material design as studies by other groups have repeatedly shown [102–110].

The transferability of the properties of atoms and groups in the genetically encoded amino acids have been discussed in detail elsewhere [7–10]. It is this transferability in conjunction with the additivity of QTAIM properties (Equation 14.16c) that is paralleled by the transferability and additivity of group contributions to properties such as partial molar volume.

We have calculated and tabulated the properties of *every atom* and *every bond* in the 20 amino acids (about 500 atoms and 500 bonds) [7–10], and have used them to construct predictive models for their experimental (genetic and biophysical) properties. These experimental properties include partial molar volumes of the entire side chains as well as group contributions to the partial molar volumes, free energies of hydration, partition coefficients, changes in protein stability upon single point mutations, and the triplet genetic code itself. This work appears to be the first to correlate the change in protein stability upon genetic mutation to the change in the properties of the underlying electron density of the wild type and mutant amino acids. The observations of pioneers such as Alff-Steinberger and Wolfenden have inspired us to search for a direct link between the electron density of the side chains and the genetic code. Such a relationship has been found and it is quite striking, underscoring questions along the lines of:

How did the genetic code evolve to be so strongly correlated to the polarity of the amino acid side chains? Has the code and the encoded ever been in direct physical contact in early evolutionary times?

14.8

Appendix A**X-Ray and Neutron Diffraction Geometries of the Amino Acids in the Literature⁹⁾**

The following table presents literature references to crystallographic determinations of the genetically encoded amino acids. The literature surveyed includes single crystal determinations of the geometries and, in some cases the electron densities, in the free form of the amino acids or as amino acid residues in small peptides.

⁹⁾ Reproduced from Ref. [8] with permission of Wiley-Liss.

Table 14.7 Literature references for crystallographic determinations of the genetically encoded amino acids.*

AA	Authors	Journal	Vol.	Year	Page	Method	Temperature	R factor	Comments
Ala	Gorbitz, C.H.; Dalhus B.	<i>Acta Crystallogr.</i>	C52	1996	1764	X	120 K	0.0854	Zwitter-ionic L-valanyl-L-alanine
Ala	Gatti, C. <i>et al.</i>	<i>J. Mol. Struct. (Theochem)</i>	255	1992	409	Xm, AIM	23 K	0.0203	Zwitter-ionic L-alanine
Ala	Destro, R. <i>et al.</i>	<i>Chem. Phys. Lett.</i>	186	1991	47	Xm, AIM	23 K	0.0203	Zwitter-ionic L-alanine
Ala	Destro, R. <i>et al.</i>	<i>J. Phys. Chem.</i>	92	1988	966	Xm	23 K	0.0203	Zwitter-ionic L-alanine
Ala	Lehmann, M.S. <i>et al.</i>	<i>J. Am. Chem. Soc.</i>	94	1972	2657	N	Room	0.022	Zwitter-ionic L-alanine
Arg	Espinosa, E. <i>et al.</i>	<i>Acta Crystallogr.</i>	B52	1996	519	Xm, AIM	130 K	0.016	Zwitter-ionic L-arginine phosphate monohydrate (LAP)
Arg	Lehmann, M.S. <i>et al.</i>	<i>J. Chem. Soc. Perkin II</i>		1973	133	N	Room	0.034	Zwitter-ionic L-arginine dihydrate
Arg	Karle, I.L.; Karle, J.	<i>Acta Crystallogr.</i>	17	1964	835	X	Room	0.103	Zwitter-ionic L-arginine dihydrate
Asn	Verbist, J.J. <i>et al.</i>	<i>Acta Crystallogr.</i>	B28	1972	3006	N	Room	0.026	Zwitter-ionic L-asparagine monohydrate
Asp	Flaig, R. <i>et al.</i>	<i>J. Am. Chem. Soc.</i>	120	1998	2227	Xm, AIM	20 K	0.0106	Zwitter-ionic D-aspartic acid (nonionized side chain)
Asp	Eggleston, D.S. <i>et al.</i>	<i>Acta Crystallogr.</i>	B37	1981	1428	X	Room	0.040	Zwitter-ionic alpha-L-aspartyl-glycine monohydrate
Cys	Gorbitz, C.H.; Dalhus B.	<i>Acta Crystallogr.</i>	C52	1996	1756	X	120 K	0.0311	Zwitter-ionic L-Cys
Cys	Kerr, K.A.; Ashmore, J.P.	<i>Acta Crystallogr.</i>	B29	1973	2124	X	Room	0.0375	Zwitter-ionic L-cysteine
CysS	Dahaoui, S. <i>et al.</i>	<i>J. Phys. Chem. A</i>	103	1999	6240	Xm, AIM	110 K	0.014	Double-zwitter-ionic L-cysteine (Cys-Cys dimer)
CysS	Jones, D.D. <i>et al.</i>	<i>Acta Crystallogr.</i>	B30	1974	1220	N	Room	0.034	L-Cystine dihydrochloride
Gln	Wagner, A.; Luger, P.	<i>J. Mol. Struct.</i>	595	2001	39	Xm, AIM	130	0.016	Zwitter-ionic L-glutamine

(Continued)

Table (Continued)

AA	Authors	Journal	Vol.	Year	Page	Method	Temperature	R factor	Comments
Gln	Suresh, S. <i>et al.</i>	<i>Acta Crystallogr.</i>	C52	1996	1313	X	Room	0.0406	Zwitter-ionic D-glutamine
Glu	Lehmann, M.S.; Nunes, A.C.	<i>Acta Crystallogr.</i>	B36	1980	1621	X	Room	0.026	Zwitter-ionic and protonated form of L-glutamic acid with nonionized side chain
Glu	Hirayama, N. <i>et al.</i>	<i>Bull. Chem. Soc. Jpn.</i>	53	1980	30	X	Room	0.034	Zwitter-ionic L-glutamic acid with nonionized side chain
Glu	Lehmann, M.S. <i>et al.</i>	<i>J. Cryst. Mol. Struct.</i>	2	1972	225	N	Room	0.026	Zwitter-ionic L-glutamic acid with nonionized side chain
Gly	Destro, R. <i>et al.</i>	<i>J. Phys. Chem. A</i>	104	2000	1047	Xm, AIM	23 K	0.0129	Zwitter-ionic glycine
Gly	Pichon-Pesme V.; Lecomte, C.	<i>Acta Crystallogr.</i>	B54	1998	485	Xm	123 K	0.0251	Zwitter-ionic triglycine
Gly	Legros, J.-P.; Kwick, A.	<i>Acta Crystallogr.</i>	B36	1980	3052	(X-N)m	120 K	$R_F^2 = 0.015$	Zwitter-ionic glycine
Gly	Jonsson, P.-G.; Kwick, A.	<i>Acta Crystallogr.</i>	B28	1972	1827	N	Room	0.030	Zwitter-ionic glycine
His	Coppens, P. <i>et al.</i>	<i>J. Am. Chem. Soc.</i>	121	1999	2585	Xm, AIM	110 K	0.0296	Zwitter-ionic D-histidine with nonionized side chain
His	Kistenmacher, T.J. <i>et al.</i>	<i>Acta Crystallogr.</i>	B28	1972	3352	X	Room	0.029	L-N-acetylhistidine monohydrate (with ionized side chain)
Ile	Lalitha, V. <i>et al.</i>	<i>Int. J. Peptide Protein Res.</i>	24	1984	123	X	Room	0.039	Zwitter-ionic glycyl-glycyl-L-iso-leucine monohydrate
Ile	Torii, K.; Iitaka, Y.	<i>Acta Crystallogr.</i>	B27	1971	2237	X	Room	0.117	Zwitter-ionic L-isoleucine
Leu	Gorbitz, C.H.; Dalhus B.	<i>Acta Crystallogr.</i>	C52	1996	1754	X	120 K	0.0435	Zwitter-ionic L-Leu
Leu	Coll, M. <i>et al.</i>	<i>Acta Crystallogr.</i>	C42	1986	599	X	Room	0.058	Zwitter-ionic L-leucine
Leu	Precigoux, G. <i>et al.</i>	<i>Acta Crystallogr.</i>	C42	1986	721	X	293 K	0.057	N-acetyl-L-prolyl-L-phenylal-nine-L-leucine monohydrate
Leu	Chaney, M.O. <i>et al.</i>	<i>Acta Crystallogr.</i>	B27	1971	544	X	Room	0.098	Zwitter-ionic L-leucine hydroiodide
Lys	Koetzle, T.F. <i>et al.</i>	<i>Acta Crystallogr.</i>	B28	1972	3207	N	Room	0.030	Zwitter-ionic L-lysine monohydrochloride dihydrate (ionized side chain)

Lys	Wright, D.A.; Marsh, R.E.	<i>Acta Crystallogr.</i>	15	1962	54	X	Room	0.057	Zwitter-ionic L-lysine monohydrochloride dihydrate (ionized side chain)
Met	Chen, C.-S.; Parthasarathy, R.	<i>Acta Crystallogr.</i>	B33	1977	3332	X	About 295 K	0.084	Zwitter-ionic N-formyl-L-methionine
Met	Torii, K.; Iitaka, Y.	<i>Acta Crystallogr.</i>	B29	1973	2799	X	Room	0.09	Zwitter-ionic L-methionine and of L-norleucine
Phe	Precigoux, G. <i>et al.</i>	<i>Acta Crystallogr.</i>	C42	1986	721	X	293 K	0.057	N-acetyl-L-prolyl-L-phenylalanine-L-leucine monohydrate
Phe	Al-Karaghoulis, A.R.; Koetzle, T.F.	<i>Acta Crystallogr.</i>	B31	1975	2461	N	Room	0.084	L-Phenylalanine hydrochloride
Pro	Koritsánszky, T. <i>et al.</i>	<i>Science</i>	279	1998	356	Xm, AIM	100 K	0.0208	DL-Proline monohydrate
Pro	Precigoux, G. <i>et al.</i>	<i>Acta Crystallogr.</i>	C42	1986	721	X	293 K	0.057	N-acetyl-L-prolyl-L-phenylalanine-L-leucine monohydrate
Pro	Tanaka, I. <i>et al.</i>	<i>Acta Crystallogr.</i>	B33	1977	116	X	Room	0.051	Benzoyloxycarbonylglycyl L-proline (Z-Gly-Pro)
Ser	Frey, M.N. <i>et al.</i>	<i>Acta Crystallogr.</i>	B29	1973	876	N	Room	0.055, 0.020	Two single-crystal determinations of zwitter-ionic L-serine monohydrate and of DL-serine
Ser	Benedetti, E. <i>et al.</i>	<i>Gaz. Chim. Ital.</i>	103	1973	555	X	Room	0.044	Zwitter-ionic L-serine
Thr	Benabicha, F. <i>et al.</i>	<i>Acta Crystallogr.</i>	B56	2000		Xm, AIM	110 K	0.023	Glycyl-L-threonine dihydrate
Thr	Benabicha, F. <i>et al.</i>	<i>Acta Crystallogr.</i>	B56	2000	155	Xm, AIM	110 K	0.0247	Zwitter-ionic glycyl-L-threonine dihydrate
Thr	Yadava, V.S.; Padmanabhan, V.M.	<i>Acta Crystallogr.</i>	B29	1973	854	X	Room	0.094	Glycyl-L-threonine dihydrate
Trp	Takigawa, T. <i>et al.</i>	<i>Bull. Chem. Soc. Jpn.</i>	39	1966	2369	X	Room	0.088	Zwitter-ionic L-tryptophan hydrochloride and hydrobromide
Trp	Pasternak, R.A.	<i>Acta Crystallogr.</i>	9	1956	341	X	Room	0.155	Glycyl-L-tryptophan dihydrate
Tyr	Dahaoui, S. <i>et al.</i>	<i>Acta Crystallogr.</i>	B55	1999	226	Xm	110 K	0.027	N-acetyl-L-tyrosine ethyl ester monohydrate

(Continued)

Table (Continued)

AA	Authors	Journal	Vol.	Year	Page	Method	Temperature	R factor	Comments
Tyr	Subramanian, E. <i>et al.</i>	<i>Int. J. Peptide Protein Res.</i>	24	1984	55	X	Room	0.059	Zwitter-ionic L-tyrosyl-L-tyrosine dihydrate
Tyr	Frey, M.N. <i>et al.</i>	<i>J. Chem. Phys.</i>	58	1973	2547	N	Room	0.026, 0.041	Zwitter-ionic L-tyrosine and L-tyrosine hydrochloride
Val	Gorbitz, C.H.; Dalhus B.	<i>Acta Crystallogr.</i>	C52	1996	1764	X	120 K	0.0854	Zwitter-ionic L-valanyl-L-alanine
Val	Dalhus B.; Gorbitz, C.H.	<i>Acta Crystallogr.</i>	C52	1996	1759	X	120 K	0.0452	Zwitter-ionic DL-Val
Val	Lalitha, V. <i>et al.</i>	<i>Int. J. Peptide Protein Res.</i>	24	1984	437	X	Room	0.040	Zwitter-ionic glycyl-glycyl-L-valine dihydrate
Val	Koetzle, T.F. <i>et al.</i>	<i>J. Chem. Phys.</i>	60	1974	4690	N	Room	0.031	L-Valine hydrochloride
Val	Torii, K.; Iitaka, Y.	<i>Acta Crystallogr.</i>	B26	1970	1317	X	Room	0.126	Zwitter-ionic L-valine

* The acronyms under "Method" are:

N = Neutron diffraction study.

X = X-ray determination with spherical refinement.

Xm = X-ray determination with multipolar (aspherical refinement).

AIM = The paper reports an atoms in molecules (QTAIM) topological analysis of the experimental density.

Acknowledgments

The author is much indebted to Professor Richard F.W. Bader who was the principal driving force behind this project while I was a graduate student in his group at McMaster University. The author thanks Professor Lou Massa, Professor Anna Gubskaya, and Ms. Alya Arabi for their critical comments on this work; Mr. Hugo Bohorquez for stimulating discussions; and Professor Philip Coppens, Professor Piero Macchi, and Professor Angelo Sironi for their authorizations to adapt intellectually owned material. The author thanks Wiley-Liss, Inc. for granting permissions to reproduce copyrighted material and the Natural Sciences and Engineering Research Council of Canada (NSERC), Canada Foundation for Innovation (CFI), and Mount Saint Vincent University for funding.

References

- 1 Shannon, C. and Weaver, W. (1963) *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, IL.
- 2 Brillouin, L. (2004) *Science and Information Theory*, 2nd edn, Dover Publications, Inc., Mineola, NY.
- 3 Gatlin, L.L. (1972) *Information Theory and the Living System*, Columbia University Press, New York.
- 4 Bohórquez, H.J., Obregón, M., Cárdenas, C., Llanos, E., Suárez, C., Villaveces, J.L., and Patarroyo, M.E. (2003) Electronic energy and multipolar moments characterize amino acid side chains into chemically related groups. *J. Phys. Chem. A*, **107**, 10090–10097.
- 5 Martín, F.J. (2001) Theoretical synthesis of macromolecules from transferable functional groups. Ph.D. thesis. McMaster University, Hamilton.
- 6 Bader, R.F.W., Matta, C.F., and Martín, F.J. (2003) Atoms in medicinal chemistry, in *Medicinal Quantum Chemistry* (eds. Carloni, P. and Alber, F.), Wiley-VCH Verlag GmbH, Weinheim, pp. 201–231.
- 7 Matta, C.F. and Bader, R.F.W. (2000) An atoms-in-molecules study of the genetically-encoded amino acids. I. Effects of conformation and of tautomerization on geometric, atomic, and bond properties. *Proteins: Struct. Funct. Genet.*, **40**, 310–329.
- 8 Matta, C.F. and Bader, R.F.W. (2002) Atoms-in-molecules study of the genetically-encoded amino acids. II. Computational study of molecular geometries. *Proteins: Struct. Funct. Genet.*, **48**, 519–538.
- 9 Matta, C.F. and Bader, R.F.W. (2003) Atoms-in-molecules study of the genetically-encoded amino acids. III. Bond and atomic properties and their correlations with experiment including mutation-induced changes in protein stability and genetic coding. *Proteins: Struct. Funct. Genet.*, **52**, 360–399.
- 10 Matta, C.F. (2002) Applications of the quantum theory of atoms in molecules to chemical and biochemical problems. Ph.D. thesis. McMaster University, Hamilton, Canada.
- 11 Matta, C.F. (2009) The response of the molecular charge density distribution to changes in the external potential and to other perturbations. Habilitation to Direct Research (HDR) Dissertation. Université Henri Poincaré (UHP), Nancy Université – 1: Nancy, Lorraine, France.
- 12 Lattman, E.E. and Rose, G.D. (1993) Protein folding: what's the question? *Proc. Natl. Acad. Sci. USA*, **90**, 439–441.
- 13 Mager, P.P. (1984) *Multidimensional Pharmacochemistry: Design of Safer Drugs*, Academic Press, Inc., London.
- 14 Hohenberg, P. and Kohn, W. (1964) Inhomogeneous electron gas. *Phys. Rev. B*, **136**, 864–871.

- 15 Bader, R.F.W. and Zou, P.F. (1992) An atomic population as the expectation value of a quantum observable. *Chem. Phys. Lett.*, **191**, 54–58.
- 16 Dirac, P.A.M. (1958) *The Principles of Quantum Mechanics*, 3rd edn, Oxford University Press, Oxford.
- 17 Stout, G.H. and Jensen, L.H. (1989) *X-Ray Structure Determination: A Practical Guide*, 2nd edn, John Wiley & Sons, Inc., New York.
- 18 Stewart, R.F. (1976) Electron population analysis with rigid pseudoatoms. *Acta Crystallogr. A*, **32**, 565–574.
- 19 Koritsanszky, T.S. and Coppens, P. (2001) Chemical applications of X-ray charge-density analysis. *Chem. Rev.*, **101**, 1583–1628.
- 20 Coppens, P. (1997) *X-Ray Charge Densities and Chemical Bonding*, Oxford University Press, Inc., New York.
- 21 Hansen, N.K. and Coppens, P. (1978) Testing aspherical atom refinement on small molecules data sets. *Acta Crystallogr. A*, **34**, 909–921.
- 22 Fernández-Serra, M.V., Junquera, J., Jelsch, C., Lecomte, C., and Artacho, E. (2000) Electron density in the peptide bonds of crambin. *Solid State Commun.*, **116**, 395–400.
- 23 Benabicha, F., Pichon-Pesme, V., Jelsch, C., Lecomte, C., and Khmou, A. (2000) Experimental charge density and electrostatic potential of glycyl-L-threonine dihydrate. *Acta Crystallogr. B*, **56**, 155–165.
- 24 Jelsch, C., Teeter, M.M., Lamzin, V., Pichon-Pesme, V., Blessing, R.H., and Lecomte, C. (2000) Accurate protein crystallography at ultra-high resolution: valence electron distribution in crambin. *Proc. Natl. Acad. Sci. USA*, **97**, 3171–3176.
- 25 Housset, D., Benabicha, F., Pichon-Pesme, V., Jelsch, C., Maierhofer, A., David, S., Fontecilla-Camps, J.C., and Lecomte, C. (2000) Towards the charge-density study of proteins: a room-temperature scorpion-toxin structure at 0.96 Å resolution as a first test case. *Acta Crystallogr. D*, **56**, 151–160.
- 26 Dahaoui, S., Pichon-Pesme, V., Howard, J.A.K., and Lecomte, C. (1999) CCD charge density study on crystals with large unit cell parameters: the case of hexagonal L-cystine. *J. Phys. Chem. A*, **103**, 6240–6250.
- 27 Jelsch, C., Pichon-Pesme, V., Lecomte, C., and Aubry, A. (1998) Transferability of multipole charge-density parameters: application to very high resolution oligopeptide and protein structures. *Acta Crystallogr. D*, **54**, 1306–1318.
- 28 Espinosa, E., Lecomte, C., Molins, E., Veintemillas, S., Cousson, A., and Paulus, W. (1996) Electron density study of a new non-linear optical material: L-arginine phosphate monohydrate (LAP). Comparison between X–X and X-(X + N) refinements. *Acta Crystallogr. B*, **52**, 519–534.
- 29 Pichon-Pesme, V., Lecomte, C., and Lachekar, H. (1995) On building a data bank of transferable experimental density parameters: application to polypeptides. *J. Phys. Chem.*, **99**, 6242–6250.
- 30 Wiest, R., Pichon-Pesme, V., Bénard, M., and Lecomte, C. (1994) Electron distributions in peptides and related molecules. Experimental and theoretical study of Leu-enkephalin trihydrate. *J. Phys. Chem.*, **98**, 1351–1362.
- 31 Pichon-Pesme, V., Lecomte, C., Wiest, R., and Bénard, M. (1992) Modeling fragments for the *ab initio* determination of electron density in polypeptides. An experimental and theoretical approach to the electron distribution in Leu-enkephalin trihydrate. *J. Am. Chem. Soc.*, **114**, 2713–2715.
- 32 Leherter, L., Guillot, B., Vercauteren, D.P., Pichon-Pesme, V., Jelsch, C., Lagoutte, A., and Lecomte, C. (2007) Topological analysis of proteins as derived from medium and high-resolution electron density: applications to electrostatic properties, in *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*, (eds. Matta, C.F. and Boyd, R.J.) Wiley-VCH Verlag GmbH, Weinheim, pp. 285–315.
- 33 Luger, P. (2007) Fast electron density methods in the life sciences: a routine application in the future? *Org. Biomol. Chem.*, **5**, 2529–2540.
- 34 Scheins, S., Messerschmidt, M., and Luger, P. (2005) Submolecular

- partitioning of morphine hydrate based on its experimental charge density at 25 K. *Acta Crystallogr. B*, **61**, 443–448.
- 35 Dittrich, B., Koritsánszky, T., Grosche, M., Scherer, W., Flaig, R., Wagner, A., Krane, H.G., Kessler, H., Riemer, C., Schreurs, A.M.M., and Luger, P. (2002) Reproducibility and transferability of topological properties; experimental charge density of the hexapeptide cyclo-(D,L-Pro)₂-(L-Ala)₄ monohydrate. *Acta Crystallogr. B*, **58**, 721–727.
 - 36 Kingsford-Adaboh, R., Dittrich, B., Wagner, A., Messerschmidt, M., Flaig, R., and Luger, P. (2002) Topological analysis of D,L-arginine monohydrate at 100 K. *Z. Kristallogr.*, **217**, 168–173.
 - 37 Flaig, R., Koritsánszky, T., Dittrich, B., Wagner, A., and Luger, P. (2002) Intra- and intermolecular topological properties of amino acids: a comparative study of experimental and theoretical results. *J. Am. Chem. Soc.*, **124**, 3407–3417.
 - 38 Wagner, A. and Luger, P. (2001) Charge density and topological analysis of L-glutamine. *J. Mol. Struct.*, **595**, 39–46.
 - 39 Flaig, R., Koritsánszky, T., Soyka, R., Häming, L., and Luger, P. (2001) Electronic insight into an antithrombotic agent by high-resolution X-ray crystallography. *Angew. Chem., Int. Ed.*, **40**, 355–359.
 - 40 Flaig, R., Koritsánszky, T., Janczak, J., Krane, H.-G., Morgenroth, W., and Luger, P. (1999) Fast experiments for charge-density determination: topological analysis and electrostatic potential of the amino acids L-Asn, D,L-Glu, D,L-Ser, and L-Thr. *Angew. Chem., Int. Ed.*, **38**, 1397–1400.
 - 41 Flaig, R., Koritsánszky, D., Zobel, D., and Luger, P. (1998) Topological analysis of the experimental electron densities of amino acids. 1. D,L-Aspartic acid at 20 K. *J. Am. Chem. Soc.*, **120**, 2227–2238.
 - 42 Luger, P. and Dittrich, B. (2007) Fragment transferability studied theoretically and experimentally with QTAIM: implications for electron density and invariom modeling, in *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*, (eds. Matta, C.F. and Boyd, R.J.) Wiley-VCH Verlag GmbH, Weinheim, pp. 317–341.
 - 43 Matta, C.F. (2001) Theoretical reconstruction of the electron density of large molecules from fragments determined as proper open quantum systems: the properties of the oripavine PEO, enkephalins, and morphine. *J. Phys. Chem. A*, **105**, 11088–11101.
 - 44 Bader, R.F.W. (1990) *Atoms in Molecules: A Quantum Theory*, Oxford University Press, Oxford, UK.
 - 45 Popelier, P.L.A. (2000) *Atoms in Molecules: An Introduction*, Prentice Hall, London.
 - 46 Matta, C.F. and Boyd, R.J. (2007) *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*, (eds. Matta, C.F. and Boyd, R.J.) Wiley-VCH Verlag GmbH, Weinheim.
 - 47 Bader, R.F.W. and Nguyen-Dang, T.T. (1981) Quantum theory of atoms in molecules: Dalton revisited. *Adv. Quantum Chem.*, **14**, 63–124.
 - 48 Bader, R.F.W., Nguyen-Dang, T.T., and Tal, Y. (1981) A topological theory of molecular structure. *Rep. Prog. Phys.*, **44**, 893–948.
 - 49 Bader, R.F.W. (1991) A quantum theory of molecular structure and its applications. *Chem. Rev.*, **91**, 893–928.
 - 50 Bader, R.F.W. (1994) Principle of stationary action and the definition of a proper open system. *Phys. Rev. B*, **49**, 13348–13356.
 - 51 Bader, R.F.W. (1998) *Encyclopedia of Computational Chemistry*, (ed. Schleyer, P. v.-R.) John Wiley & Sons, Ltd, Chichester, UK, pp. 64–86.
 - 52 Bader, R.F.W. (1998) A bond path: a universal indicator of bonded interactions. *J. Phys. Chem. A*, **102**, 7314–7323.
 - 53 Runtz, G.R., Bader, R.F.W., and Messer, R.R. (1977) Definition of bond paths and bond directions in terms of the molecular charge distribution. *Can. J. Chem.*, **55**, 3040–3045.
 - 54 Cao, W.L., Gatti, C., MacDougall, P.J., and Bader, R.F.W. (1987) On the presence of non-nuclear attractors in the charge distributions of Li and Na clusters. *Chem. Phys. Lett.*, **141**, 380–385.

- 55 de Vries, R.Y., Briels, W.J., Feil, D., te Velde, G., and Baerends, E.J. (1996) Charge density study with maximum entropy method on model data of silicon. A search for non-nuclear attractors. *Can. J. Chem.*, **74**, 1054–1058.
- 56 Bader, R.F.W. and Platts, J.A. (1997) Characterization of an *F*-center in an alkali halide cluster. *J. Chem. Phys.*, **107**, 8545–8553.
- 57 Taylor, A., Matta, C.F., and Boyd, R.J. (2007) The hydrated electron as a pseudo-atom in cavity-bound water clusters. *J. Chem. Theor. Comput.*, **3**, 1054–1063.
- 58 Keith, T.A., Bader, R.F.W., and Aray, Y. (1996) Structural homeomorphism between the electron density and the virial field. *Int. J. Quantum Chem.*, **57**, 183–198.
- 59 Bader, R.F.W. (2009) Bond paths are not chemical bond. *J. Phys. Chem. A*, **113**, 10391–10396.
- 60 Bader, R.F.W. (1980) Quantum topology of molecular charge distributions. III. The mechanics of an atom in a molecule. *J. Chem. Phys.*, **73**, 2871–2883.
- 61 Bader, R.F.W. and Beddall, P.M. (1972) Virial field relationship for molecular charge distributions and the spatial partitioning of molecular properties. *J. Chem. Phys.*, **56**, 3320–3328.
- 62 Bader, R.F.W., Beddall, P.M., and Peslak, J., Jr. (1973) Theoretical development of a virial relationship for spatially defined fragments of molecular systems. *J. Chem. Phys.*, **58**, 557–566.
- 63 Srebrenik, S. and Bader, R.F.W. (1975) Towards the development of the quantum mechanics of a subspace. *J. Chem. Phys.*, **63**, 3945–3961.
- 64 Srebrenik, S., Bader, R.F.W., and Nguyen-Dang, T.T. (1978) Subspace quantum mechanics and the variational principle. *J. Chem. Phys.*, **68**, 3667–3679.
- 65 Schwinger, J. (1951) The theory of quantized fields. I. *Phys. Rev.*, **82**, 914–927.
- 66 Keith, T.A. (2007) Atomic response properties, in *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*, (eds. Matta, C.F. and Boyd, R.J.) Wiley-VCH Verlag GmbH, Weinheim.
- 67 Bader, R.F.W. and Matta, C.F. (2004) Atomic charges are measurable quantum expectation values: a rebuttal of criticisms of QTAIM charges. *J. Phys. Chem. A*, **108**, 8385–8394.
- 68 Popelier, P.L.A. (1996) Integration of atoms in molecules: a critical examination. *Mol. Phys.*, **87**, 1196–1187.
- 69 Kosov, D.S. and Popelier, P.L.A. (2000) Atomic partitioning of molecular electrostatic potentials. *J. Phys. Chem. A*, **104**, 7339–7345.
- 70 Kosov, D.S. and Popelier, P.L.A. (2000) Convergence of the multipole expansion for electrostatic potentials of finite topological atoms. *J. Chem. Phys.*, **113**, 3969–3974.
- 71 Popelier, P.L.A., Joubert, L., and Kosov, D.S. (2001) Convergence of the electrostatic interaction based on topological atoms. *J. Phys. Chem. A*, **105**, 8254–8261.
- 72 Matta, C.F. and Boyd, R.J. (2007) *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*, Wiley-VCH Verlag GmbH, Weinheim.
- 73 Russell, B. (1945) *A History of Western Philosophy*, Simon and Schuster, New York.
- 74 Head-Gordon, T., Head-Gordon, M., Frisch, M.J., Brooks, C., III, and Pople, J.A. (1991) Theoretical studies of blocked glycine and alanine peptide analogues. *J. Am. Chem. Soc.*, **113**, 5989–5997.
- 75 Matta, C.F. (2009) How dependent are molecular and atomic properties on the electronic structure method? Comparison of Hartree-Fock, DFT, and MP2 on a biologically-relevant set of molecules. *J. Chem. Theor. Comput.*, in press, DOI: 10.1002/jcc.21417.
- 76 Chalikian, T.V. (2008) On the origin of volumetric data. *J. Phys. Chem. B*, **112**, 911–917.
- 77 Hinz, H.-J. (ed.) (1986) *Thermodynamic Data for Biochemistry and Biotechnology*, Springer-Verlag, Berlin.
- 78 Millero, F.J., Surodo, A.L., and Shin, C. (1978) The apparent molal volumes and adiabatic compressibilities of aqueous amino acids at 25 °C. *J. Phys. Chem.*, **82**, 784–792.

- 79 Lilley, T.H. (1985) Physical properties of amino acid solutions, in *Chemistry and Biochemistry of the Amino Acids*, Chapman & Hall, London, pp. 591–624.
- 80 Collantes, E.R. and Dunn, W.J.I. (1995) Amino acid side chain descriptors for quantitative structure–activity relationship studies of peptide analogues. *J. Med. Chem.*, **38**, 2705–2713.
- 81 Bridgman, P.W. (1931) *Dimensional Analysis*, Yale University Press, New Haven.
- 82 Lee, A. and Chalikian, T.V. (2001) Volumetric characterization of the hydration properties of heterocyclic bases and nucleosides. *Biophys. Chem.*, **92**, 209–227.
- 83 Creighton, T.E. (1983) *Proteins: Structures and Molecular Principles*, W. H. Freeman and Co., New York.
- 84 Wolfenden, R., Andersson, L., Cullis, P.M., and Southgate, C.C.B. (1981) Affinities of amino acid side chains for solvent water. *Biochem.*, **20**, 849–855.
- 85 Fersht, A. (1999) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, W. H. Freeman Co., New York.
- 86 Sharp, K.A., Nicholls, A., Friedman, R., and Honig, B. (1991) Extracting hydrophobic free energies from experimental data: relationship to protein folding and theoretical models. *Biochem.*, **30**, 9686–9697.
- 87 Radzicka, A. and Wolfenden, R. (1988) Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solutions. *Biochem.*, **27**, 1664–1670.
- 88 Shortle, D., Stites, W.E., and Meeker, A.K. (1990) Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease. *Biochem.*, **29**, 8033–8041.
- 89 Guerois, R., Nielsen, J.E., and Serrano, L. (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.*, **320**, 369–387.
- 90 Loladze, V.V., Ermolenko, D.N., and Makhatazde, G.I. (2002) Thermodynamic consequences of burial of polar and non-polar amino acid residues in the protein interior. *J. Mol. Biol.*, **320**, 343–357.
- 91 Nirenberg, M.W., Jones, O.W., Leder, P., Clark, B.F.C., Sly, W.S., and Petska, S. (1963) On the coding of genetic information. *Cold Spring Harbor Symp. Quant. Biol.*, **28**, 549–557.
- 92 Alff-Steinberger, C. (1969) The genetic code and error transmission. *Proc. Natl. Acad. Sci. USA*, **64**, 584–591.
- 93 Bader, R.F.W., Popelier, P.L.A., and Chang, C. (1992) Similarity and complementarity in chemistry. *J. Mol. Struct. (Theochem)*, **255**, 145–171.
- 94 Bader, R.F.W., Carroll, M.T., Cheeseman, J.R., and Chang, C. (1987) Properties of atoms in molecules: atomic volumes. *J. Am. Chem. Soc.*, **109**, 7968–7979.
- 95 Bader, R.F.W., Keith, T.A., Gough, K.M., and Laidig, K.E. (1992) Properties of atoms in molecules: additivity and transferability of group polarizabilities. *Mol. Phys.*, **75**, 1167–1189.
- 96 Bader, R.F.W., MacDougall, P.J., and Lau, C.D.H. (1984) Bonded and nonbonded charge concentrations and their relations to molecular geometry and reactivity. *J. Am. Chem. Soc.*, **106**, 1594–1605.
- 97 Gillespie, R.J. (1972) *Molecular Geometry*, Van Nostrand Reinhold, London.
- 98 Bader, R.W.F. and Heard, G.L. (1999) The mapping of the conditional pair density onto the electron density. *J. Chem. Phys.*, **111**, 8789–8797.
- 99 Carroll, M.T., Chang, C., and Bader, R.F.W. (1988) Prediction of the structures of hydrogen-bonded complexes using the Laplacian of the charge density. *Mol. Phys.*, **63**, 387–405.
- 100 MacDougall, P.J. and Henze, C.E. (2001) Identification of molecular reactive sites with an interactive volume rendering tool. *Theor. Chim. Acc.*, **105**, 345–353.
- 101 MacDougall, P.J. and Henze, C.E. (2007) Fleshing-out pharmacophores with volume rendering of the Laplacian of the charge density and hyperwall visualization technology, in *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*, (eds. Matta, C.F. and Boyd, R.J.) Wiley-VCH Verlag GmbH, Weinheim, pp. 499–514.

- 102 Popelier, P.L.A. (1999) Quantum molecular similarity. 1. BCP space. *J. Phys. Chem. A*, **103**, 2883–2890.
- 103 O'Brien, S.E. and Popelier, P.L.A. (1999) Quantum molecular similarity. Part 2: the relation between properties in BCP space and bond length. *Can. J. Chem.*, **77**, 28–36.
- 104 O'Brien, S.E. and Popelier, P.L.A. (2001) Quantum molecular similarity. 3. QTMS descriptors. *J. Chem. Inf. Comput. Sci.*, **41**, 764–775.
- 105 Song, M., Breneman, C.M., Bi, J., Sukumar, N., Bennett, K.P., Cramer, S., and Tugcu, N. (2002) Prediction of protein retention times in anion-exchange chromatography systems using support vector regression. *J. Chem. Inf. Comput. Sci.*, **42**, 1347–1357.
- 106 Breneman, C.M. and Rhem, M. (1997) QSPR analysis of HPLC column capacity factors for a set of high-energy materials using electronic van der Waals surface property descriptors computed by transferable atom equivalent method. *J. Comput. Chem.*, **18**, 182–197.
- 107 Adam, K.R. (2002) New density functional and atoms in molecules method of computing relative pK_a values in solution. *J. Phys. Chem. A*, **106**, 11963–11972.
- 108 Platts, J.A. (2000) Theoretical prediction of hydrogen bond basicity. *Phys. Chem. Chem. Phys.*, **2**, 3115–3120.
- 109 Platts, J.A. (2000) Theoretical prediction of hydrogen bond donor capacity. *Phys. Chem. Chem. Phys.*, **2**, 973–980.
- 110 Dumitrica, T., Landis, C.M., and Yakobson, B.I. (2002) Curvature-induced polarization in carbon nanoshells. *Chem. Phys. Lett.*, **360**, 182–188.
- 111 Matta, C.F., Arabi, A.A., and Keith, T.A. (2007) Atomic partitioning of the dissociation energy of the PO(H) bond in hydrogen phosphate anion (HPO_4^{2-}): disentangling the effect of Mg^{2+} . *J. Phys. Chem. A*, **111**, 8864–8872.

15

Energy Richness of ATP in Terms of Atomic Energies: A First Step*

Chérif F. Matta and Alya A. Arabi

Their discovery belongs, undoubtedly, to the most brilliant achievement of modern biochemistry [on high energy phosphate bonds].

Albert Szent-Györgyi (Bioenergetics, 1957, p. v)

15.1

Introduction

Adenosine 5'-triphosphate (ATP) is the biological fuel molecule *par excellence* [1–6]. How does this molecule act as an energy currency can be answered, in part, by following the changes in the energies of the atoms composing it as the molecule undergoes one of the reactions to which it is coupled *in vivo*. This atomic level investigation pinpoints the regions of the ATP molecule that are most responsible for its inherent instability (the enthalpic contribution to this instability) in its dominant form at neutral pH. True, this does not provide the full picture for the completion of which one must account for entropic contribution and solvation, and to a lesser extent finite temperature and vibrational corrections, but the atomic partitioning of the electronic part of the enthalpy is a first step, a step that focuses exclusively on the internal electronic structure of the ATP molecule itself in isolation.

Tri- and diphosphorylated molecules (ATP, ADP, GTP, GDP, etc.) all possess at least one “high-energy phosphate bond (P–O)” [1–6]. Free energy necessary to drive otherwise nonspontaneous reactions is made available through coupling of these reactions with the exergonic hydrolysis of ATP ($\Delta G^0 \approx -32 \text{ kJ mol}^{-1}$ at 37 °C):



*Dr. Todd A. Keith has contributed Section 5.3.3 of this chapter. The authors thank him for his important contribution and his useful comments on the remainder of the chapter.

where the superscripts indicate the net electric charge of each species under cellular conditions.

The ubiquitous [7–13] doubly charged magnesium cation, Mg^{2+} , has considerable effects on the electron distribution and on the enthalpy of hydrolysis of ATP. Tracing the atomic origin of these effects is a primary goal of this chapter. The cation has also a significant effect on the shape of the potential energy surface and on the height of the activation energy barrier to hydrolysis, a topic not addressed in this chapter, the atomic roots of these effects being work in progress at the present in our research group.

15.2

How “(De)Localized” is the Enthalpy of Bond Dissociation?

Chemical reactions involve bond making and breaking. Experimentally, bond dissociation energies (BDE) are often estimated from the difference between the heats of formation of the products and those of the reactants. For example, in the reaction:



the BDE can be estimated from [14]:

$$\begin{aligned} \text{BDE}(A-B)^0 &\equiv \Delta H^0(A-B) \\ &= \Delta H_f^0(A) + \Delta H_f^0(B) - \Delta H_f^0(A-B) \end{aligned} \quad (15.3)$$

where $\Delta H_f^0(A)$, $\Delta H_f^0(B)$, and $\Delta H_f^0(A-B)$ are the heats of formation of A , B , and $A-B$ at standard conditions, respectively.

The heat of formation is a *global* property of a molecule. In contrast, a bond dissociation energy (BDE) has primarily a *local* character since it is associated with a particular bond in the molecule. The extent to which an atom is destabilized (or stabilized) in the products of bond dissociation with respect to the reactants is not reflected in a global quantity such as the heat of formation. This issue of the degree of (de)stabilization of every atom upon bond dissociation has been addressed in the literature only recently [15, 16] by comparing atomic energies obtained from the quantum theory of atoms in molecules (QTAIM) [17–19] before and after the reaction (bond dissociation in this case) has taken place. In this framework, an atomic contribution to the (vibrationless, 0 K) electronic BDE is [15, 16]:

$$\Delta E(\Omega) = E(\Omega)_{\text{products}} - E(\Omega)_{\text{reactants}} \quad (15.4)$$

where $\Delta E(\Omega)$ is the change in energy of a particular atom Ω in a molecule, $E(\Omega)_{\text{reactant}}$ is the energy of Ω in the reactant, and $E(\Omega)_{\text{products}}$ is its energy after the dissociation of a given bond. The BDE would then be given by the sum of atomic contributions [15, 16]:

$$\text{BDE}_{\text{electronic}(0\text{K})} = \sum_{\Omega} \Delta E(\Omega) \quad (15.5)$$

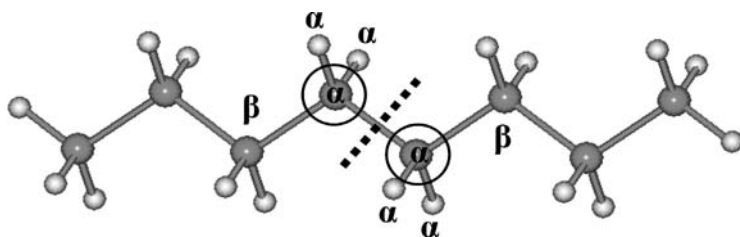


Figure 15.1 *n*-Octane (C_8H_{18}) with the dotted line indicating the bond broken to yield two identical radicals ($C_4H_9\cdot$) along with the labeling scheme.

The application of this approach to simple hydrocarbons shows that contrary to what might be expected, the dissociation of a bond between two carbons does not necessarily result in the destabilization of these two carbons. As an example, breaking the central bond in simple saturated alkanes (Figure 15.1) results in significant destabilization of the α -hydrogens and the β -carbons, but, surprisingly, not the α -carbons between which the bond has been severed (Figure 15.2a). The energetic contributions to the total energy of the α -carbons cancels almost completely resulting in no contribution from these atoms to the overall BDE (with the exception of the first member of the series, ethane, devoid of carbons except the α -carbons, and in which the carbon atoms together contribute approximately 10% of the BDE, the remaining 90% is contributed by the three hydrogen atoms) [15]. The longer the saturated alkane chain, the more the atomic contributions approach an asymptotic limit as is clear from Figure 15.2a. Thus, in the case of alkanes, the enthalpy change accompanying the bond breaking is localized near *but not at* the carbon atoms involved in the bonding.

The situation is different if the bond being broken homolytically is a double bond in alkenes to yield two triplet radicals (Figure 15.2b). In this case, and as can be seen from the figure, at the asymptotic limit, the BDE is contributed in part by two α -carbons (which together contribute $\sim 50\%$ of the BDE), while the rest of the destabilization is contributed primarily by two α -hydrogens (one on each of the triplet radicals), followed by the β -carbons and then the rest of the atoms.

This atomic partitioning of the BDE, if applied to the “energy-rich” P–O chemical bonds of biochemistry, can shed light on the spatial localization of the energy stores in the so-called energy-rich molecules such as ATP, GTP, UTP, and so on. It is important to stress that energy is *not* (never) released, but rather required, to sever a chemical bond, even the so-called energy rich P–O bonds. Energy is released not from these bonds but rather as the net resultant of the formation and the breaking of a number of chemical bonds concurrently.

In order to gain insight into the functioning of ATP (and similar high-energy nucleoside triphosphates), one thus observes the net change in atomic energies resulting from several bond making and breaking events. In other words, instead of a simple reaction of homolysis where an atomic contribution to the BDE is energy of an atom after homolysis minus its energy before homolysis (each species at its respective equilibrium geometry and most stable spin multiplicity), in this case it is

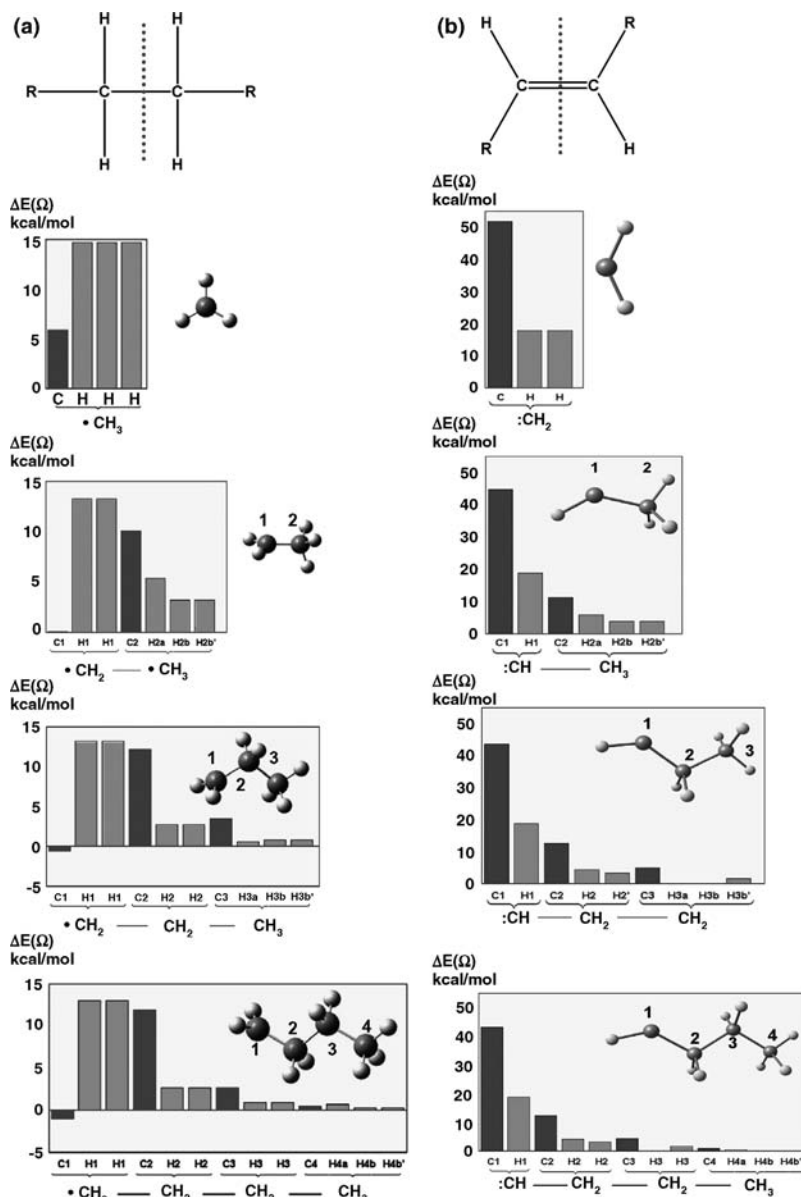


Figure 15.2 Atomic contributions to the bond dissociation enthalpies of the central bond in even-numbered n -alkanes (C_nH_{2n+2}) to two identical doublet $C_nH_{n+1}^\bullet$ radicals (a) and

transalkenes (C_nH_{2n}) to two identical triplet radicals $C_nH_n^\bullet$ (b). (Adapted from Ref. [15] with permission of the American Institute of Physics).

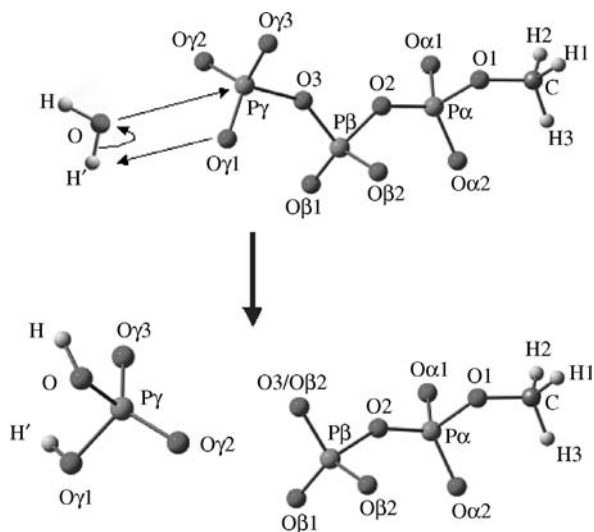


Figure 15.3 Ball-and-stick representation of the optimized geometries of the reactants and products of the hydrolysis of a model of ATP (methyl triphosphate) along with the atomic numbering scheme. (Reproduced from Ref. [20] with permission of the American Chemical Society).

the change of the energy of an atom due to the reaction that involves making and breaking several bonds simultaneously, which is of interest. Before embarking on this program, one must choose (1) the reaction and model molecule and (2) the appropriate level of theory.

In view of the large size of ATP, a compromise must be struck between the size of the model molecule representing it and the level of theory. As a first step in this investigation, we have selected the hydrolysis of ATP using a truncated model of this molecule consisting of its triphosphate tail capped with a methyl (representing the sugar and nucleic acid base). The reaction and atom numbering scheme are depicted in Figure 15.3. The reaction is studied in the absence and presence of complexation with the magnesium cation Mg^{2+} .

15.3

The Choice of a Theoretical Level

15.3.1

The Problem

The choice of an appropriate level of theory for implementing this study is not obvious. On one hand, one has theoretical levels that yield well-defined atomic energies within the framework of QTAIM (and these include *ab initio* methods such as Hartree-Fock, MPn , CI, etc.), and on the other hand, one has to rely on a more affordable level especially if Coulombic correlation is to be accounted for.

Furthermore, the use of a large basis is necessary for an accurate representation of the electron density. A popular choice is the density functional theory (DFT), but if one calculates atomic energies using Kohn–Sham (KS) orbitals using the same numerical procedure as applied to *ab initio* molecular orbitals, then the *meaning* of these atomic energies requires scrutiny [16]. Before delving into the meaning of atomic energies obtained from KS orbitals, we first examine whether, at least *numerically*, one recovers the correct trends of the atomic contributions to the BDE (as, for example, those obtained from MP2 calculations) in a test molecule with similar groups and bonds as in ATP. An inexpensive choice of test molecule is the anion HPO_4^{2-} , free and as complexed with magnesium (MgHPO_4).

15.3.2

Empirical Correlation of Trends in the Atomic Contributions to BDE: Comparison of MP2 and DFT(B3LYP) Results

A partitioning of the homolytic BDE of the P–O bond in HPO_4^{2-} and its magnesium complex MgHPO_4 has been performed at the MP2 (full) level using a 6-311++G(*d*, *p*) basis set and at the DFT(B3LYP) level using the same basis set (where geometry optimizations and calculations of final wavefunctions were all performed at the same respective levels, that is, the calculated species are true minima on their respective potential energy surfaces).

Figure 15.4 displays the atomic contributions to the BDE in the presence and absence of Mg^{2+} at the MP2 and the DFT(B3LYP) levels of theory. In this figure, a negative contribution indicates that the atom is more stable in the bond dissociation products, while a positive contribution means the atom is destabilized by bond dissociation. The effect of complexation with Mg^{2+} on the contributions of various atoms to the BDE can be visualized from Figure 15.4b. It is clear that complexation with Mg^{2+} has a sizable effect both in the magnitude and sign of the atomic contributions to the BDE with a resulting reduction of the BDE; in other words, Mg^{2+} has the net effect of facilitating the dissociation of (weakening) the P–O bond in this model system. Figure 15.4 provides evidence that the numerical trends in atomic contributions to the BDE of the P–O bond are preserved at the two tested levels of theory.

15.3.3

Theory¹⁾

15.3.3.1 QTAIM Atomic Energies from the *ab initio* Methods

For a molecule described by an exact stationary-state Born–Oppenheimer wavefunction Ψ and at an equilibrium geometry, any atom Ω in the molecule satisfies the following virial theorem [17]:

$$2T(\Omega) + V(\Omega) = 2T(\Omega) + V_{\text{ne}}(\Omega) + V_{\text{ee}}(\Omega) + V_{\text{nn}}(\Omega) = 0 \quad (15.6)$$

1) This section is reproduced from Ref. [16] with permission of the American Chemical Society.

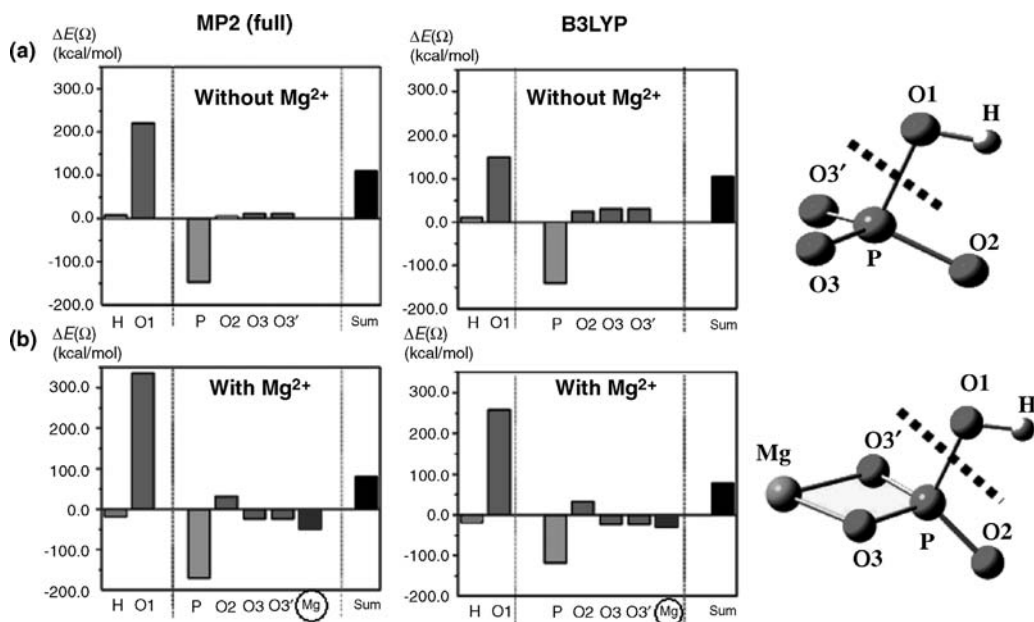


Figure 15.4 The atomic contributions to the BDE [$\Delta E(\Omega)$ in kcal mol⁻¹] of the P–O1 bond in the magnesium-free HPO_4^{2-} (a) and in the MgHPO_4 complex (b) calculated at (U)MP2 (full)/6-311++G(d,p)/(U)MP2(full)/6-311++G(d,p) (left) and the (U)B3LYP/6-311++G(d,p)/(U)B3LYP/6-311++G(d,p) (right) levels of theory. The atomic contribution to the BDE is positive/negative if the atom is destabilized/stabilized in the products of bond

breaking. The sum of all contributions (the column out the far right of the plots) is equal to the bond dissociation energy (BDE). The atom labeling is indicated on a ball-and-stick representation of the optimized geometries of the free and metal-complexed HPO_4^{2-} . A dashed line is drawn across the bond being broken homolytically. (Adapted from Ref. [16] with permission of the American Chemical Society).

where $T(\Omega)$ is the electronic kinetic energy of the atom and $V_{\text{ne}}(\Omega)$, $V_{\text{ee}}(\Omega)$, and $V_{\text{nn}}(\Omega)$ are the nuclear–electron attraction, electron–electron repulsion, and nuclear–nuclear repulsion potential energy contributions from the atom. While $T(\Omega)$, $V_{\text{ne}}(\Omega)$, and $V_{\text{ee}}(\Omega)$ have expressions similar to the corresponding molecular expressions, the nuclear repulsion contribution $V_{\text{nn}}(\Omega)$ is a less obvious origin-independent (for exact wavefunctions) sum of three origin-dependent terms (see Section 6.3.4 of Ref. [17] for explicit expressions for these terms):

$$V_{\text{nn}}(\Omega) = \sum_{A=1}^{n_{\text{atoms}}} \mathbf{R}_A \cdot \mathbf{F}_A(\Omega) + V(\Omega, \Omega') + \mathcal{V}_S(\Omega) \quad (15.7)$$

where $\mathbf{R}_A$ is the position vector of nucleus A, $\mathbf{F}_A(\Omega)$ is the force on that nucleus, and $\mathcal{V}_S(\Omega)$ is the virial of the Ehrenfest forces exerted on the surface bounding the atom.

The contributions $V_{\text{nn}}(\Omega)$ are additive to give the molecular V_{nn} because of the molecular Hellman–Feynman electrostatic theorem and because the terms $V(\Omega, \Omega')$ sum to zero for the molecule, as do the $\mathcal{V}_S(\Omega)$ terms.

Equation 15.6 is unique to atoms in molecules (and groups of atoms in molecules) both in its variational derivation and because only for atoms in molecules is the kinetic energy always well defined [17].

Similar to the molecular energy E , an atomic energy $E(\Omega)$ is defined as the sum of kinetic and potential contributions as follows [17]:

$$E(\Omega) = T(\Omega) + V_{ne}(\Omega) + V_{ee}(\Omega) + V_{nn}(\Omega) = T(\Omega) + V(\Omega) \quad (15.8)$$

Combining Equations 15.6 and 15.8, one obtains the following simple relationship between $E(\Omega)$ and $T(\Omega)$ and between $E(\Omega)$ and $V(\Omega)$:

$$E(\Omega) = -T(\Omega) = \frac{1}{2} V(\Omega) \quad (15.9)$$

Equations 15.6, 15.8, and 15.9 are applicable not only to atoms in molecules but also to groups of atoms in molecules and, of course, the molecule as a whole. When referring to energetic terms for the molecule as a whole, the “ (Ω) ” suffix is omitted.

Being able to express the total atomic energy $E(\Omega)$ in terms of the atomic kinetic energy $T(\Omega)$ drastically simplifies its calculation and, in some ways, its interpretation.

For a typical *approximate* wavefunction $\Psi_{\text{approx.}}$, however, the atomic and molecular virial theorems will *not* be exactly satisfied. The consequence of this is that calculating atomic energies using Equation 15.9 will *not* result in energy additivity, that is,

$$\sum_{\Omega=1}^{n_{\text{atoms}}} [-T(\Omega)] = -T \neq E \quad (\text{for approximate, noncoordinate-scaled wavefunctions}) \quad (15.10)$$

Another consequence of typical approximate wavefunctions is that the atomic energy contribution $V_{nn}(\Omega)$ defined in Equation 15.7 will be origin dependent [21], thus making the direct evaluation of atomic energies $E(\Omega)$ using Equation 15.8 ambiguous. In addition, the $V_{nn}(\Omega)$ contributions will *not* be additive to give the molecular value V_{nn} unless the wavefunction satisfies the molecular Hellman–Feynman electrostatic theorem for all nuclei [21], a stringent requirement not satisfied by typical approximate wavefunctions. The direct evaluation of atomic energies using Equation 15.8 thus does *not* guarantee energy additivity for typical approximate wavefunctions. Even in cases where the Hellman–Feynman electrostatic theorem is satisfied for all nuclei and energy additivity is obtained using Equation 15.8, each atomic energy $E(\Omega)$ will still be origin dependent due to the $V_{nn}(\Omega)$ term, in addition to being difficult and costly to calculate. The origin independence of the atomic kinetic energy is another good reason for using Equation 15.9 to calculate the atomic energy, assuming the problem of energy additivity expressed by Equation 15.10 is addressed.

Energy additivity for atomic energies defined by atomic kinetic energies using Equation 15.9 *can* be obtained if the coordinates of the wavefunction are scaled using the following factor ς [22, 23]:

$$\xi = -\frac{1}{2} \frac{V}{T} = -\frac{1}{2} \left(\frac{E}{T} - 1 \right) \quad (15.11)$$

It proves enlightening to express ζ in terms of 1 plus a (small) correction term ϵ , which vanishes for wavefunctions satisfying the molecular virial theorem, as follows:

$$\xi = \frac{1}{2} - \frac{1}{2} \frac{E}{T} = 1 - \frac{1}{2} \frac{E}{T} - \frac{1}{2} = 1 + \epsilon \quad (15.12)$$

where

$$\epsilon = -\frac{1}{2} \frac{E}{T} - \frac{1}{2} = -\frac{1}{2} \frac{V}{T} - 1 \quad (15.13)$$

Using a (renormalized) wavefunction Ψ_ζ , whose coordinates have been scaled by ζ , the molecular kinetic energy T_ζ , potential energy V_ζ , and total energy E_ζ are given by:

$$T_\xi = \xi^2 T = T + 2\epsilon T + \epsilon^2 T = \frac{(E-T)^2}{4T} \quad (15.14)$$

$$V_\xi = \xi V = V + \epsilon V = -\frac{(E-T)^2}{2T} \quad (15.15)$$

and

$$\begin{aligned} E_\xi = T_\xi + V_\xi &= T + 2\epsilon T + \epsilon^2 T + V + \epsilon V \\ &= T + 2\epsilon T + \epsilon^2 T - \epsilon(1 + \epsilon)2T \\ &= T - T\epsilon^2 \\ &= -\frac{(E-T)^2}{4T} \\ &= 2V_\xi = -T_\xi \end{aligned} \quad (15.16)$$

These equations show that the energies T_ζ , V_ζ , and E_ζ from the coordinate-scaled wavefunction Ψ_ζ satisfy the molecular virial theorem and, equally important, that the energy E_ζ is quadratic in the (small) correction ϵ , while both T_ζ and V_ζ are linear in the (small) correction ϵ . In other words, coordinate scaling of the wavefunction to satisfy the molecular virial theorem will change the kinetic and potential energy components T and V much more than the total energy E .

Unfortunately, such a coordinate scaling of the wavefunction will also lead to forces on the nuclei and make the energy nonstationary with respect to the variational parameters in the wavefunction [22, 23]. In addition, some atomic and molecular properties calculated using the unscaled wavefunction Ψ will be inconsistent with the energies calculated from the scaled wavefunction. Thus, ideally, coordinate scaling of the wavefunction to satisfy the molecular virial theorem should be done *self-consistently* with geometry optimization and the wavefunction determination, leading to a valid variational and/or perturbational wavefunction, satisfaction of the molecular virial theorem, a true equilibrium geometry, and a consistent set of atomic and molecular properties.

It should be noted that coordinate scaling of the wavefunction to satisfy the molecular virial theorem does *not* guarantee satisfaction of the individual atomic virial theorems [21].

In many cases, a computationally simpler and commonly used [17] procedure for obtaining energy additivity for atomic energies calculated using Equation 15.9 – when the wavefunction does not satisfy the molecular virial theorem – is simply to scale the atomic kinetic energies $T(\Omega)$ by E/T . This simpler procedure does *not* correspond to a coordinate scaling of the wavefunction but it is employed in the present work to obtain the *ab initio* MP2 energies:

$$E(\Omega) = \frac{E}{T} T(\Omega) \quad (15.17)$$

Equation 15.17 is a valid approximation to a coordinate scaling result if the change in total molecular energy E brought about by coordinate scaling is relatively small and if the change in each atomic kinetic energy brought about by coordinate scaling is directly proportional – by the same factor for all atoms, and hence for the molecule – to the corresponding unscaled atomic kinetic energy.

15.3.3.2 Atomic Energies from Kohn–Sham Density Functional Theory Methods

For Kohn–Sham DFT (KS-DFT) methods [24], such as the B3LYP [25, 26] method employed in the present work, the definition and calculation of atomic energies is less clear than for *ab initio*, that is, Hamiltonian-based, wavefunction methods such as Hartree–Fock or MP2. However, if one views KS-DFT theory as a semiempirical variant of Hartree–Fock theory, then one can follow a similar procedure to that given in Section 3.3.1, albeit with a somewhat different interpretation. The atomic virial theorem corresponding to Equation 15.6 for KS-DFT methods is [27]:

$$2T_s(\Omega) + V_{ne}(\Omega) + V_{ee,H}(\Omega) + V_{nn}(\Omega) + E_{xc}(\Omega) + T_c(\Omega) = 0 \quad (15.18)$$

where $T_s(\Omega)$ is the so-called “noninteracting” kinetic energy of atom Ω , $V_{ne}(\Omega)$ is the nuclear–electron attraction energy contribution from atom Ω , $V_{ee,H}(\Omega)$ is the “Hartree” (i.e., electrostatic) contribution of atom Ω to the electron–electron potential energy, and $V_{nn}(\Omega)$ has the same expression as given in Equation 15.7. A possible “starting point” for defining an atomic “exchange correlation” energy $E_{xc}(\Omega)$ and correlation kinetic energy $T_c(\Omega)$ is to relate them to the virial of the exchange–correlation potential $v_{xc}(\mathbf{r})$ as follows [28]:

$$E_{xc}(\Omega) + T_c(\Omega) = - \int_{\Omega} d\mathbf{r} Q(\mathbf{r}) \mathbf{r} \cdot \nabla v_{xc}(\mathbf{r}) \quad (15.19)$$

an *origin-dependent* expression that constitutes a generalization of the corresponding *origin-independent* molecular expression to an atom in a molecule. The definition of $v_{xc}(\mathbf{r})$ depends, of course, on the particular KS-DFT method used.

If one defines the atomic energy $E(\Omega)$ in a KS-DFT method as:

$$E(\Omega) = T_s(\Omega) + V_{ne}(\Omega) + V_{ee,H}(\Omega) + V_{nn}(\Omega) + E_{xc}(\Omega) \quad (15.20)$$

then one gets energy additivity, assuming that “wavefunction” of the KS-DFT method satisfies the Hellman–Feynman electrostatic theorem for all nuclei and thus that $V_{nn}(\Omega)$ are additive to give V_{nn} at equilibrium geometries.

Combining Equations 15.18 and 15.20, one gets the following relationship:

$$E(\Omega) = -[T_s(\Omega) + T_c(\Omega)] = -T(\Omega) \quad (15.21)$$

As for *ab initio* methods, this relationship will not be satisfied either at the atomic or molecular levels by typical approximate KS-DFT “wavefunctions.” However, just as for *ab initio* methods, coordinate scaling of the KS-DFT “wavefunction” can be done to satisfy $E = -T$ for the molecule and additivity of the atomic energies calculated using Equation 15.21.

The KS-DFT relationship between the atomic energy and total atomic kinetic energy is the same as for *ab initio* methods, but the atomic kinetic energy now consists of two contributions, the readily accessible “noninteracting” kinetic energy, whose expression is the same as for Hartree–Fock, and the “correlation” kinetic energy, which can in principle be determined from Equation 15.21, if $E_{xc}(\Omega)$ is determined separately.

The molecular correlation kinetic energy T_c is believed to be on the order of the correlation energy itself, $T_c \sim -E_c$ [29], and therefore much smaller than the molecular “noninteracting” kinetic energy T_s . If one simply ignores T_c and $T_c(\Omega)$, then one may calculate an atomic energy from $T_s(\Omega)$ by simply scaling $T_s(\Omega)$ by the factor E/T_s :

$$E(\Omega) = \frac{E}{T_s} T_s(\Omega) \quad (15.22)$$

This is the definition of the B3LYP atomic energies employed here and it is similar to the definition used for the *ab initio* MP2 atomic energies used in Equation 15.17. The validity of this expression, compared to using the full kinetic energies T and $T(\Omega)$, requires that either $T_c \ll T_s$ and $T_c(\Omega) \ll T_s(\Omega)$ or $T_c = aT_s$ and $T_c(\Omega) = aT_s(\Omega)$, as shown below:

$$\frac{1}{(T_s + T_c)} = \frac{1}{T_s} - \frac{T_c}{(T_s + T_c)^2} + \frac{T_c^2}{(T_s + T_c)^3} + \dots \quad (15.23)$$

$$\begin{aligned} E(\Omega) &= \frac{E}{T} T(\Omega) = \left[\frac{E}{(T_s + T_c)} \right] [T_s(\Omega) + T_c(\Omega)] \\ &= \left[\frac{T_s(\Omega)}{T_s} - \frac{T_s(\Omega)T_c}{(T_s + T_c)^2} + \frac{T_s(\Omega)T_c^2}{(T_s + T_c)^3} + \dots \right] \\ &\quad + \left[\frac{T_c(\Omega)}{T_s} - \frac{T_c(\Omega)T_c}{(T_s + T_c)^2} + \frac{T_c(\Omega)T_c^2}{(T_s + T_c)^3} + \dots \right] \\ &= \frac{E}{T_s} T_s(\Omega) \quad [\text{if } T_c \ll T_s \text{ and } T_c(\Omega) \ll T_s(\Omega)] \end{aligned} \quad (15.24)$$

$$\text{If: } T_c = aT_s, T_c(\Omega) = aT_s(\Omega) \quad (15.25)$$

$$\begin{aligned} \text{Then: } E(\Omega) &= \frac{E}{T} T(\Omega) = \left[\frac{E}{(T_s + aT_s)} \right] [T_s(\Omega) + aT_s(\Omega)] \\ &= \left[\frac{E}{(1+a)T_s} \right] [1+a]T_s(\Omega) = \frac{E}{T_s} T_s(\Omega) \end{aligned} \quad (15.26)$$

15.3.3.3 Atomic Contributions to the Energy of Reaction

An atomic contribution $\Delta E(\Omega)$ to the electronic energy of reaction ΔE is obtained also from Equation 15.4 except that now reactants and products are those of multiple bond breaking and making events rather than a fission of a single chemical bond, leading to an additivity similar to the one described by Equation 15.5:

$$\Delta E = \sum_{\Omega} \Delta E(\Omega) \quad (15.27)$$

Atomic energies used in Equation 15.4 to obtain $\Delta E(\Omega)$ are calculated either from Equation 15.17 in the case of MP2 or from Equation 15.22 in the case of DFT.

15.4

Computational Details

Since DFT's B3LYP hybrid functional has been shown to recover the trends in the atomic contributions to the BDE of the P–O bond calculated at the MP2 level of theory, it is used along with the 6-31 + G(*d*, *p*) basis set in the elucidation of the atomic contribution to the energy of hydrolysis of ATP. The diffuse functions on the nonhydrogen atoms (denoted by “+”) are included in the basis set to improve the description of the diffuse electron density of anionic species that are involved in this hydrolysis reaction. Geometry optimizations were performed at the same level of theory to ensure that (gradient) forces on all nuclei vanish. Electronic structure calculations were performed using Gaussian 03 [30] followed by a QTAIM analysis using AIMALL [31, 32], while molecular graphs were obtained with AIM2000 [33].

The chosen model for the ATP molecule consists of its triphosphate tail (the primary site of energy storage) capped with a methyl group to eliminate the possibility of forming hydrogen bonds between a terminal hydrogen and the phosphate tail that may favor nonrealistic buckled conformations. Methyl triphosphate and methyl diphosphate will be referred to as ATP and ADP, respectively, throughout the remainder of this chapter.

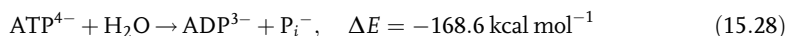
³¹P-NMR experiments demonstrate that the hydrolysis of ATP is brought about by an in-line nucleophilic attack of a water molecule on the terminal phosphate followed by an inversion of configuration of P γ [34] that entails the formation of two bonds (O–P γ and O γ 1–H') and the breaking of two other (O–H' and P γ –O3) (see Figure 15.3). Other mechanisms involving more than one water molecule have

also been proposed, for example, the “multicenter proton relay” mechanism [35–37] but are not considered here.

15.5

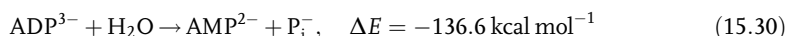
(Global) Energies of the Hydrolysis of ATP in the Absence and Presence of Mg^{2+}

The vacuum-phase electronic (vibrationless, 0 K) energies of hydrolysis of ATP to ADP calculated at B3LYP/6-31+G(*d*, *p*)/B3LYP/6-31+G(*d*, *p*) in the absence and presence of the metal, respectively, are [20]:

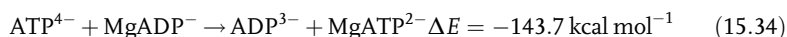
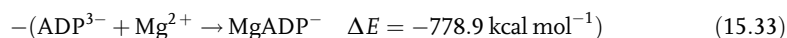
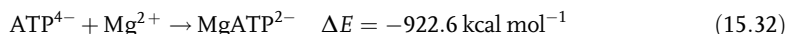


These two reaction energies are the subject of atomic partitioning in this work [20].

Comparing the energies of reactions (15.28) and (15.29) indicates that the metal reduces the magnitude of the energy of hydrolysis in the vacuum-phase dramatically, a fact that has been noted previously for the hydrolysis of ADP (to AMP) at the same level of theory [9]:



The metal-induced reduction in the magnitude of energy of the hydrolysis reactions as written above indicates a preferential binding of Mg^{2+} to the reactant with respect to the products by $143.7 \text{ kcal mol}^{-1}$ (Equation 15.29) and by $121.0 \text{ kcal mol}^{-1}$ (Equation 15.31). Equivalently, the difference in the binding of the metal to the reactant (ATP^{4-}) and to the product (ADP^{3-}) is $143.7 \text{ kcal mol}^{-1}$, that is [20],



The stronger binding of Mg^{2+} to ATP^{4-} is often attributed the larger negative electrostatic charge.

15.6

How “(De)Localized” is the Energy of Hydrolysis of ATP?

15.6.1

Phosphate Group Energies and Modified Lipmann’s Group Transfer Potentials

The total energy of the three phosphate groups in ATP exhibits an interesting trend: there is a gradual increase in the energy of the phosphate on going from the

α -phosphate group to the terminal γ -phosphate group, the γ -phosphate being the least stable in the ATP molecule. The total energies of these PO_3 groups in atomic units (and relative energies in kcal mol^{-1}) are $E(\alpha\text{-PO}_3) = -567.1094 \text{ au}$ (0 kcal mol^{-1}) $< E(\beta\text{-PO}_3) = -567.0781 \text{ au}$ ($19.7 \text{ kcal mol}^{-1}$) $< E(\gamma\text{-PO}_3) = -566.8707 \text{ au}$ ($149.8 \text{ kcal mol}^{-1}$), a trend in the same direction as the corresponding group volumes that are (in atomic units) 457.2, 482.2, and 553.4, respectively.

Upon complexation with magnesium, the trends in group volumes and energies parallel those without magnesium, but with a considerably less spread between the energy of the most-stable (α) and least-stable (γ) phosphate groups. There is a marked stabilization of all phosphate groups in the complex when compared to the corresponding groups in the free ATP molecule, an observation that can be ascribed to the favorable interaction of the negatively charged phosphate tail with the field of the positive metal ion in the complex. The total energies of the PO_3 groups in atomic units (and relative energies in kcal mol^{-1} with respect to the $\alpha\text{-PO}_3$ of free uncomplexed ATP) are $E(\alpha\text{-PO}_3) = -567.2430 \text{ au}$ ($-83.8 \text{ kcal mol}^{-1}$) $< E(\beta\text{-PO}_3) = -567.1849 \text{ au}$ ($-47.4 \text{ kcal mol}^{-1}$) $< E(\gamma\text{-PO}_3) = -567.1394 \text{ au}$ ($-18.8 \text{ kcal mol}^{-1}$), a trend that, again, is in the same direction as the corresponding group volumes that are (in atomic units) 434.1, 436.7, and 500.1, respectively. Thus, the metal (a) dampens and evens out the variations in group energies along the phosphate tail of ATP and (b) lowers the energies of all these groups when compared to free uncomplexed ATP. Similar observations are found in the case of ADP.

Lipmann defines the “group potential” [1] as a measure of the degree of *activation of a group in a certain binding, comparing it to what might be called the ‘ground state’ or the ‘free compound’*, quoting Albert Pullman and Bernard Pullman [38]. Since, the quantum theory of atoms in molecules provides an unambiguous definition of the energy of an atom or a group *within* a molecule, we may propose a possible modification of Lipmann’s definition by considering the atom/group potential as the energy of the atom or group in a parent molecule (e.g., reactants) minus the energy of that atom or group in a reference compound (e.g., products). The group potential (G.P.) of the terminal γ -phosphate group in ATP, for example, is given by [20]:

$$\begin{aligned} \text{G.P.}(\gamma\text{-PO}_3)_{\text{ATP}^{4-}} &= \sum_{\gamma\text{-PO}_3} (\text{ATP}^{4-}) - \sum_{\gamma\text{-PO}_3} (\text{H}_2\text{PO}_3^-) \\ &= -566.8707 + 567.2559 \text{ (au)} \\ &= +241.7 \text{ (kcal mol}^{-1}\text{)} \end{aligned} \quad (15.35)$$

In the magnesium complex, this group’s potential is [20]:

$$\begin{aligned} \text{G.P.}(\gamma\text{-PO}_3)_{\text{MgATP}^{2-}} &= \sum_{\gamma\text{-PO}_3} (\text{MgATP}^{2-}) - \sum_{\gamma\text{-PO}_3} (\text{H}_2\text{PO}_3^-) \\ &= -567.1394 + 567.2559 \text{ (au)} \\ &= +73.1 \text{ (kcal mol}^{-1}\text{)} \end{aligned} \quad (15.36)$$

The change in the $\gamma\text{-PO}_3$ group potential due to complexation is $241.7 - 73.1 = 168.6 \text{ kcal mol}^{-1}$, signifying that the γ -phosphate groups has a much lower transfer tendency from the MgATP^{2-} complex than from the metal-free ATP^{4-} .

15.6.2

Atomic Contributions to the Energy of Hydrolysis of ATP in the Absence and Presence of Mg^{2+}

In free uncomplexed ATP, ten atoms contribute in excess of 10 kcal mol^{-1} in magnitude to the energy of hydrolysis, as can be seen in Figure 15.5a (details in Table 3 of Ref. [20]). From this figure, six atoms are more stable in the product of this reaction (i.e., ADP and P_i); these are $\text{P}\alpha$, $\text{P}\beta$, $\text{P}\gamma$, $\text{O}\gamma 1$, $\text{O}\gamma 2$, and $\text{O}\gamma 3$. On the other hand, four atoms, $\text{O}\beta 1$, $\text{O}\beta 2$, $\text{O}3/\text{O}\beta 3$, and O , are more stable in the reactants (i.e., ATP^{4-} and water). The sum of the contributions of these atoms account for the bulk of the energy of reaction (the atomic contributions of the remaining atoms have a resultant of approximately -2 kcal mol^{-1}).

The γ -phosphate group has a dominant contribution favoring hydrolysis equal to the negative of the modified Lipmann's group transfer potential (Equation 15.35), that is, $\Delta E(\gamma\text{-PO}_3) = -241.7 \text{ kcal mol}^{-1}$. The β -phosphate as well as the incoming water molecule disfavor the reaction by 61.9 and $29.5 \text{ kcal mol}^{-1}$, respectively. The destabilization of the incoming water molecule occurs principally in the oxygen atom, the energies of the two hydrogen atoms being almost unaffected by hydrolysis to within $0.5 \text{ kcal mol}^{-1}$. The destabilization of the β -phosphate group upon reaction is the resultant of $-54.9 \text{ kcal mol}^{-1}$ from the $\text{P}\gamma$ (more stable in the products) and an overwhelmingly opposite contribution from the three oxygen atoms $\text{O}\beta 1$, $\text{O}\beta 2$, and $\text{O}3/\text{O}\beta 3$, which together contribute $+116.8 \text{ kcal mol}^{-1}$ (more stable in the reactants). We mention here that the group contributions to the energy of hydrolysis alternate in sign: positive for water and β -phosphate, and negative for γ - and α -phosphates.

A comparison of the bar graphs in Figure 15.5a and b that are plotted to the same scale reveals the marked dampening effect Mg^{2+} has on all atomic contributions to the energy of reaction (in both direction, favoring and disfavoring reactions) and on the overall energy of reaction ΔE (the rightmost bar).

15.7

Other Changes upon Hydrolysis of ATP in the Presence and Absence of Mg^{2+}

15.7.1

Bond Properties and Molecular Graphs

Figure 15.6 displays the molecular graphs of the species involved in Equations 15.28 and 15.29. The graph of the metal-ATP complex shows that the metal is tetracoordinated, being linked by bond paths to four oxygens, namely, $\text{O}\alpha 2$, $\text{O}\beta 1$, $\text{O}\gamma 2$, and $\text{O}3$. The metal is only tricoordinated in the metal-ADP complex (to $\text{O}\alpha 2$, $\text{O}\beta 1$, and $\text{O}3/\text{O}\beta 2$). The electron density at the BCP, ρ_{BCP} , involving metal in the MgADP complex ranges from 0.052 to 0.059 au , consistent with a bonding of approximately equal strength. In the metal-ATP complex, the $\text{Mg}-\text{O}3$ is significantly longer and exhibits lower density at the BCP than the others ($\text{BL} = 2.133 \text{ \AA}$, $\rho_{\text{BCP}} = 0.034 \text{ au}$),

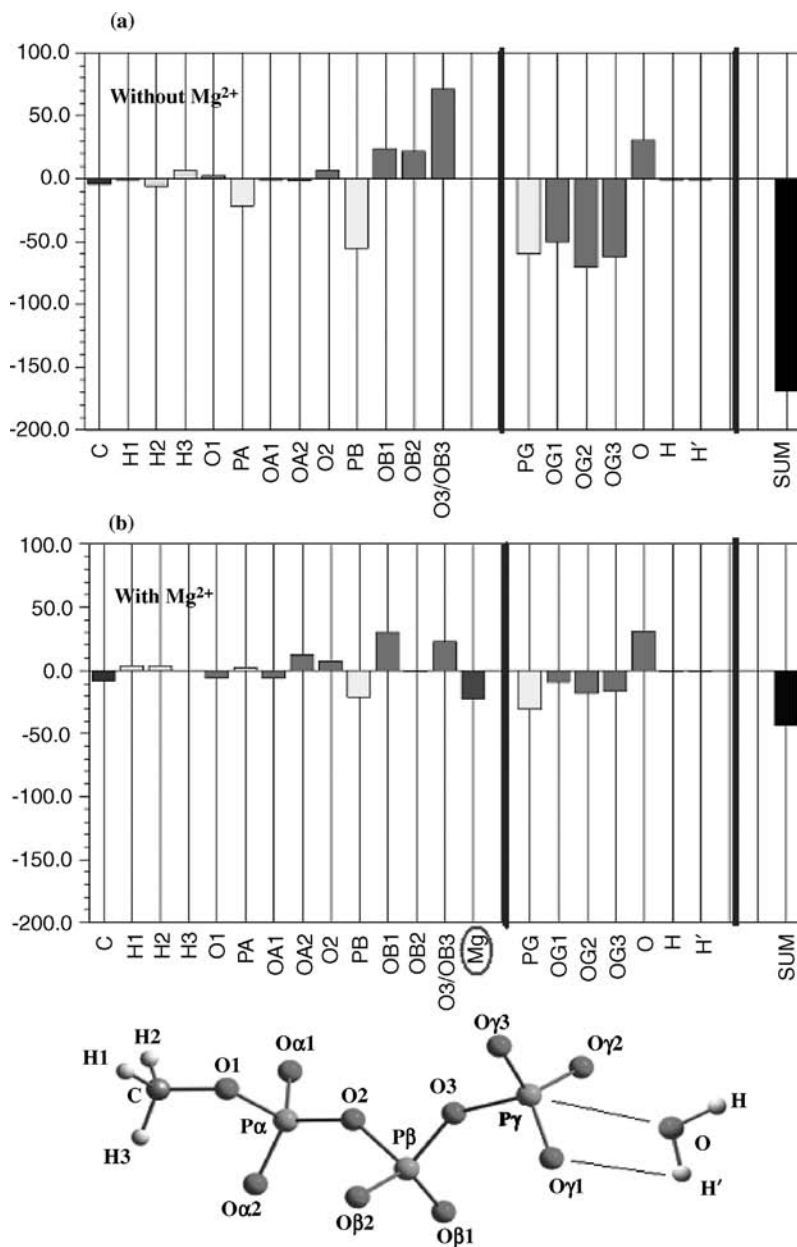


Figure 15.5 Atomic contributions to the energy of hydrolysis of ATP, $\Delta E(\Omega)$, (a) in the absence and (b) presence of Mg^{2+} , along with the atom labeling scheme. The heavy vertical lines partition each bar graph into three regions: the left section corresponds to ADP, the middle to P_i , and the right section is the sum of the

atomic contributions to the energy of reaction (i.e., the energy of reaction). When $\Delta E(\Omega) > 0$, the atom is more stable in the products and when $\Delta E(\Omega) < 0$, the atom is more stable in the reactants. (Reproduced from Ref. [20] with permission of the American Chemical Society).

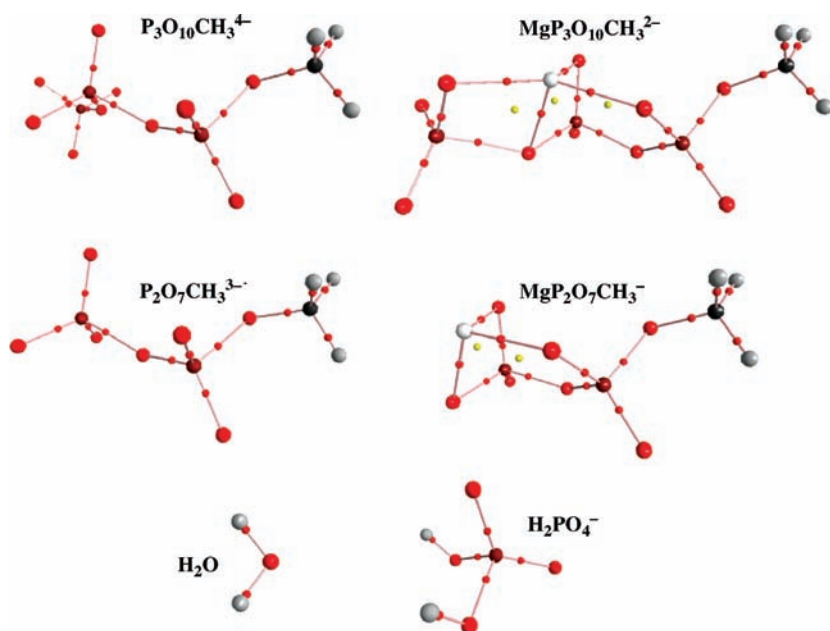


Figure 15.6 Molecular graphs of the reactant and products of hydrolysis of the ATP model used in this study, in the presence and absence of Mg^{2+} : $\text{P}_3\text{O}_{10}\text{CH}_3^{4-}$, $\text{MgP}_3\text{O}_{10}\text{CH}_3^{2-}$, $\text{P}_2\text{O}_7\text{CH}_3^{3-}$, and $\text{MgP}_2\text{O}_7\text{CH}_3^-$ are, respectively, the models of ATP, ATP complexed with Mg^{2+} , ADP, and ADP complexed with Mg^{2+} and H_2PO_4^- is the inorganic phosphate. The

positions of the bond critical points (BCP) are indicated by the small red dots on the bond paths, and those of the ring critical points by the yellow dots. The positions of the nuclei are indicated by the spheres with the following color code: P = dark red, O = red, C = black, H = gray, and Mg = white. (Reproduced from Ref. [16] with permission of the American Chemical Society).

while the $\text{Mg}-\text{O}\gamma 2$ is considerably shorter ($\text{BL} = 1.905 \text{ \AA}$) and exhibits a $q_{\text{BCP}} = 0.057 \text{ au}$. All the metal–oxygen bonds in both complexes can be primarily classified as closed shell (ionic) [39] because they exhibit relatively small values of q_{BCP} , $\nabla^2 q_{\text{BCP}}$, potential energy density at the BCP (V_{BCP}), and total energy density at the BCP (H_{BCP}), and are characterized by $\nabla^2 q_{\text{BCP}} > 0$ and $H_{\text{BCP}} > 0$.

Table 15.1 lists bond lengths and the electron density at the bond critical point (q_{BCP}) for the bonds along the triphosphate chain. The values listed in the table show the effect of Mg^{2+} on the bond lengths (BL) and q_{BCP} values along the $\text{O}-\text{P}-\text{O}$ backbone. The effect of the metal on the last bond in the chain, the $\text{O}3-\text{P}\gamma$, is particularly significant where it elongates it by $0.135 \text{ \AA}$ and decreases the q_{BCP} of this bond from 0.105 au to 0.085 . These effects are consistent with a significant weakening of this bond (preparing it for hydrolysis), a conclusion consistent with that of other investigations [40]. Interestingly, complexation with the metal appears to strengthen (rather than weaken) the second “high-energy phosphate bond,” that is, the $\text{O}2-\text{P}\beta$, by shortening it to $0.092 \text{ \AA}$ with a marked accompanying increase in q_{BCP} (from 0.109 to 0.134 au). These observations are consistent with increasing the preference for the hydrolysis of the terminal (γ) high-energy bond and lowering that

Table 15.1 Triphosphate chain bond lengths (BL) and electron densities at the bond critical point (ρ_{BCP}) in ATP and Mg-ATP complex.

Bond	BL (Å)	ϱ_{BCP} (au)	BL (Å)	ϱ_{BCP} (au)
	ATP ^{4−}		MgATP ^{2−}	
C-O1	1.390	0.281	1.416	0.257
O1-P α	1.748	0.132	1.664	0.157
P α -O2	1.581	0.179	1.615	0.170
O2-P β	1.812	0.109	1.720	0.134
P β -O3	1.594	0.175	1.619	0.171
O3-P γ	1.826	0.105	1.961	0.085

tendency for the inner (β) high-energy bond in the complex when compared to the free ATP molecule. The magnesium cation also accentuates the bond length alternation that preexists in the free uncomplexed molecule, an alternation that is already well documented (see, for example, Refs [10, 35]).

Three-dimensional representations are sometimes useful in providing a visual image accompanying numerical results. Figure 15.7a is a representation of two isodensity envelopes for methyl triphosphate (the model for ATP), and Figure 15.7b and c display similar envelopes for the magnesium complex of methyl triphosphate

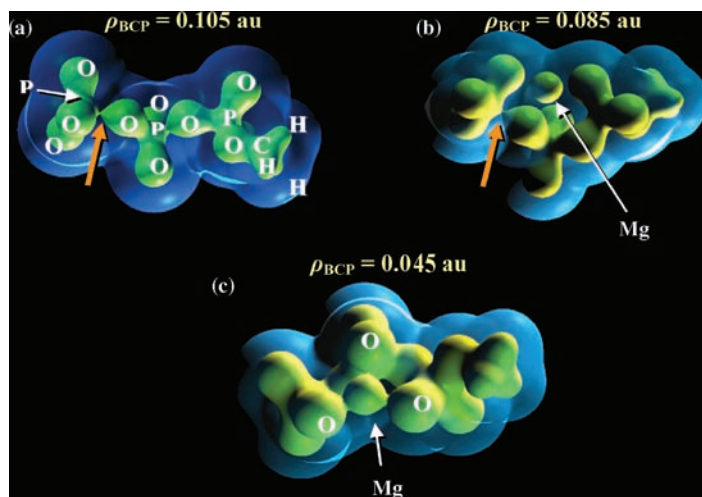
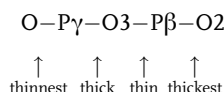


Figure 15.7 Electron density envelopes of (a) free methyl triphosphate and (b and c) Mg complex of methyl triphosphate. The outer transparent envelope (blue) is the van der Waals envelope ($\rho = 0.001$ au isodensity envelope) that corresponds to the empirical “outer surface” of the molecule. The inner solid

(yellow) surfaces have an isodensity indicated in the figure: In (a) and (b) corresponding the ρ_{BCP} of the terminal P γ –O3 bond in (a) free and (b) complexed methyl triphosphate. (c) is a rotated methyl triphosphate molecule to show the metal and three of its four oxygen ligands (O α 2 on the right, O γ 2 on the left, and O β 1 to the top).

(the model of Mg–ATP complex). In all the three representations, the outer transparent blue envelope is that corresponding to $\varrho = 0.001$ au, the so-called van der Waals envelope that is strongly correlated with experimental effective molecular sizes. The inner solid envelope (yellow) has ϱ values of (a) 0.105, (b) 0.085, and (c) 0.045 au. These inner isodensity surfaces correspond to the values of ϱ equal to the bond critical point density (ϱ_{BCP}) of a chosen bond in each of the three cases, respectively: (a) ϱ_{BCP} of the terminal O3–P γ in the free uncomplexed ATP, (b) ϱ_{BCP} of the terminal O3–P γ in the Mg–ATP complex, and (c) ϱ_{BCP} of the weakest Mg–oxygen other than O3. These isodensity envelopes encompass in a continuous sheath of density all nuclei bonded by a bond path with $\varrho_{\text{BCP}}' > \varrho_{\text{BCP}}$; the thicker the encompassing sheath in the bonding region, the higher the ϱ_{BCP} of that particular bond. Those atoms bonded by a bond path characterized by $\varrho_{\text{BCP}}' = \varrho_{\text{BCP}}$ will be just touching at one point (at the BCP), and those bonded with weaker bonds (i.e., with $\varrho_{\text{BCP}}' < \varrho_{\text{BCP}}$) will be surrounded by discontinuous surface surrounding each nucleus separately [41, 42]. Finally, the shape of these inner ϱ_{BCP} surfaces also provide a fast visual indicator of the ionicity of a chemical bonding interaction where it is close to spherical in ionic bonding and is distorted in more covalent interactions.

Figure 15.7a and b shows the marked alternation of the thickness of the envelope encompassing the backbone:



with the pattern in the case of Mg complex (Figure 15.7b) emerging at a significantly lower value of ϱ_{BCP} indicating the weakening of the terminal P γ –O bond. Figure 15.7c shows the gradual increase in strength on going from the Mg–O β 1 bond (top) going clockwise in the figure to the Mg–O α 2 (right), and finally to the strongest bond on the left, namely, the Mg–O γ 2 bond. Figure 15.7b and c shows how spherical the metal appears, which is not surprising given the highly ionic nature of the bonding to this atom in the complex, the magnesium having lost almost completely the two electrons in its valence M-shell (see Section 15.7.2).

15.7.2

Group Charges in ATP in the Absence and Presence of Mg^{2+}

Figure 15.8 illustrates the group charges in ATP (free and complexed with magnesium). From this figure it is clear that the charge of ATP, $q(\text{ATP}) = -4$ au, is spread out among the three phosphate groups in addition to the terminal methoxy groups, with each group carrying nearly a unit of negative charge with a magnitude ranging between about 1.2–0.8 au. The most negative group is the terminal γ -phosphate group (carrying ~ 1.2 au). This charge distribution is at variance with the one given in typical biochemistry textbooks where the terminal γ -phosphate group is assigned a charge of -2 au. The small departure of the total charge of ATP from -4 (by 0.002 au)

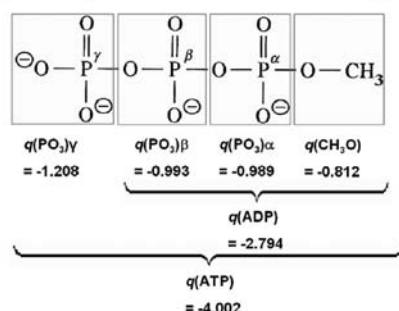
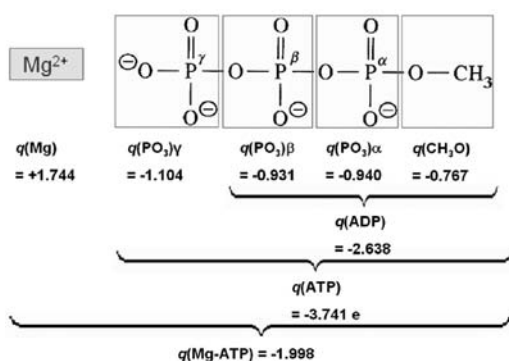
(a) Group Charges in the Absence of Mg^{2+} (b) Group Charges in the Presence of Mg^{2+} 

Figure 15.8 Group charge distribution in ATP and in the absence (a) and presence (b) of Mg^{2+} . (Reproduced from Ref. [20] with permission of the American Chemical Society).

and other similar deviations from molecular values are due to small cumulative numerical integration errors.

Figure 15.8b shows that the metal remains to a large extent a doubly charged cation with a net charge of $+1.744 \text{ au}$. In the metal complex, the phosphate groups and the capping methoxy group retain, each, a negative charge that is close to unity, not dissimilar to those in free uncomplexed ATP. A comparison of the charge of the terminal γ -phosphate between free and metal-complexed ATP shows that this group loses electron population in the metal-complex, this group has a charge of -1.208 au in free ATP^{4-} and -1.104 au in MgATP^{2-} .

15.7.3

Molecular Electrostatic Potential in the Absence and Presence of Mg^{2+}

The molecular electrostatic potential (MEP), $V(\mathbf{r})$, is obtained from

$$V(\mathbf{r}) = \sum_A \frac{Z_A}{|\mathbf{R}_A - \mathbf{r}|} - \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}' \quad (15.37)$$

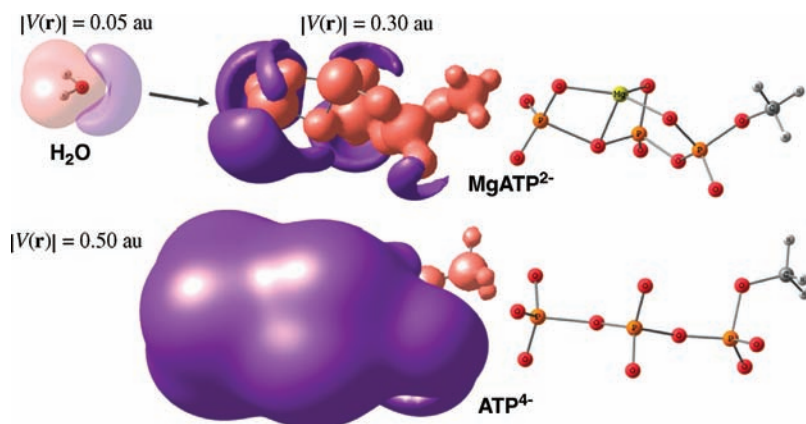


Figure 15.9 Selected isovalue envelopes of molecular electrostatic potential (MEP) of (top) magnesium–methyl triphosphate complex and (bottom) free, uncomplexed methyl triphosphate. The magnitude of the MPE is given in atomic unit where violet envelopes are for $V(\mathbf{r}) < 0$, while the pale red envelopes are

for $V(\mathbf{r}) > 0$. On the far right is a ball-and-stick model of the structure for which the MEP is displayed in the same orientation, and on the far left (in transparent envelopes) is the MEP of the attacking water with an arrow suggesting a direction of attack.

where Z_A is the charge of nucleus A at a location given by the position vector $\mathbf{R}_A$ and $\rho(\mathbf{r}')$ is the electron density. Figure 15.9 shows selected isopotential envelopes of the calculated molecular electrostatic potential maps of the ATP analogue (methyl triphosphate) in the presence (a) and absence (b) of magnesium along with the molecular skeleton of the respective species in exactly the same orientation (to the right). In view of its large negative charge, the ATP molecule is surrounded by a strongly negative electrostatic potential that disfavors the approach of a nucleophilic species necessary to trigger the hydrolysis reaction. Complexation with the metal cation reduces the spatial extent and magnitude of the negative regions of the MEP around the triphosphate chain and opens electrophilic regions in the potential field, that is, regions of positive $V(\mathbf{r})$. These regions, notably the region between the three terminal oxygen atoms $O\gamma_1$, $O\gamma_2$, and $O\gamma_3$, may suggest a direction of approach of the attacking nucleophile (e.g., H_2O) oriented to expose its negative side of $V(\mathbf{r})$ toward the positive “hole” punctured in the potential surrounding $P\gamma$. This is consistent with the in-line nucleophilic attack of the water on the terminal phosphate followed by an inversion of configuration of the $P\gamma$ proposed on the basis of ^{31}P -NMR evidence [34].

15.8 Conclusions

The quantum theory of atoms in molecules provides a partitioning of the molecular space in nonoverlapping atomic regions with well-defined boundaries.

This partitioning of space allows the partitioning of molecular properties into atomic contributions. A notable example is the partitioning of the total molecular energy into additive atomic energies through the application of the atomic statement of the virial theorem. We describe here the physical meaning of atomic energies within *ab initio* theory and within DFT. Since atomic energies are well defined for an intact molecule as well as for its dissociation products, the bond dissociation energy, and more generally the energy of reaction, can be partitioned into atomic contributions. The difference in the energy of an atom before and after the reaction is the contribution of *that* atom to the energy of reaction (whether the reaction is a simple bond dissociation or involves the breaking and making of several chemical bonds). Thus, this chapter addresses the question:

What is the contribution of every atom in a reacting system to the energy of reaction?

The energy of hydrolysis in the vacuum phase and without statistical mechanical corrections was found to be -168.6 to -24.9 kcal mol $^{-1}$ for the metal-free and the Mg-complex case, respectively. The atomic partitioning of the energy of hydrolysis of ATP in the presence and absence of complexation was studied along with a number of other properties and their change due to the hydrolysis.

The metal has a considerable dampening effect on the individual atomic contributions to the energy of hydrolysis, particularly the atoms constituting the terminal (γ) phosphate, the group released in the products of hydrolysis. The terminal phosphate group (γ -PO $_3$) is the dominant contributor to the energy of reaction and is, therefore, the region in the ATP molecule from which the dominant fraction of the electronic energy is released upon hydrolysis. The “energy richness” of the terminal phosphate is much more considerable in the metal-free case. The values of the proposed modified definition of Lipmann’s group transfer potential are $+241.7$ and $+73.1$ kcal mol $^{-1}$ for the terminal phosphates in the metal-free and the metal-complexed ATP, respectively, indicating that the terminal phosphate is a significantly better leaving group in the metal-free case. (The modified Lipmann’s group transfer potential is the difference between the QTAIM group energy in the reactants and the energy of the same group in a product reference molecule).

The molecular graph of the metal complexes with the reactants and products of hydrolysis reveals that the metal shares four bond paths in its complex with ATP but only three in ADP. Furthermore, and as already noted by other workers [35], Mg $^{2+}$ induces a large alternation in the bond length and bond strength (reflected in Q_{BCP}) of the P–O–P backbone. In particular, the metal considerably lengthens and weakens the terminal P–O bond, one of two of the so-called “energy-rich phosphate bonds” of ATP and the terminal P–O bond in ADP (that is molecule’s only energy-rich bond). At the same time, the metal strengthens the inner high-energy phosphate bond in ATP (so favoring the hydrolysis of the terminal bond and reducing the hydrolysis tendency of the inner bond).

An examination of the atomic and group charges reveals that the negative charge on each phosphate group in free and metal-complexed ADP and ATP is close to

unity and that there is a significant charge carried by the methoxy oxygen of approximately -0.8 au, indicating a greater spread of the negative charge than is typically reported based on classical formal charges in biochemistry textbooks. Further, the complexation with the metal does not alter the atomic and group charges in ATP or ADP much and the Mg^{2+} remains essentially a doubly charged cation bearing a net charge of approximately $+1.7e$ in both complexes, having lost its outer M-shell completely.

This study will be extended in the future by the (nontrivial) inclusion of solvation effects and statistical mechanical and thermochemical corrections, effects and corrections indispensable for a comparison with experimentally determined Gibbs energy of reaction. We also plan to investigate the atomic contributions to the activation energy barrier to the hydrolysis reaction to shed light on the origin of the barrier and how it is affected through complexation with metal ions (magnesium as well as other biologically relevant cations, for example, Ca^{2+}). The future plans also include extending the calculation to the full ATP molecule and to other possible positions for the interaction with the metal ion, positions that may be affected by the presence of the sugar and the adenine base.

To the authors' knowledge, the question about the atomic contributions to the energies of reactions appears to have never been addressed in a quantitative manner in the literature. The quest to further investigate this question is motivated by a fundamental interest in view of the central role played by bond making and breaking in (bio)chemistry. Moreover, this approach may also lead to a number of "practical" applications such as in the improvement of the accuracy of calculated BDEs by targeting the use of the locally dense basis set approach (LDBS) [43–46] instead of arbitrarily placing the "dense" part of the basis functions on atoms *presumed* to be the main contributors to the BDE. Instead, the dense part of the basis set can be placed on those atoms that contribute most significantly to the BDE *in addition* to those atoms directly involved in the reaction.

We close this chapter by a foreseeing and visionary statement by Bader and Nguyen-Dang made almost three decades ago [47]:

Through the definition of an atom's average energy, one may isolate those *spatial* regions of a reacting system in which potential energy is at first accumulated and then later released, either to drive the same reaction to completion or to initiate a subsequent one. This ability to spatially identify the "energy-rich" atoms of a molecular system can be used to understand in a detailed way the mechanics of an enzyme-substrate interaction and to quantify the concept of "high-energy bonds" and the role they are assigned in biochemical reactions. Related concepts such as steric acceleration could also be tested in a quantitative manner.

Acknowledgments

The authors are indebted to Professor Lou Massa (Hunter College, The City University of New York) and Professor James Pincode (Dalhousie University) for

useful suggestions. The authors thank the American Chemical Society and the American Institute of Physics for permissions to reproduce copyrighted material. Alya Arabi acknowledges the Natural Sciences and Engineering Research Council of Canada (NSERC) for financial support in the form of an NSERC-PGS-D Fellowship, and the Killam Trusts for an honorary graduate scholarship. CFM acknowledges NSERC for a Discovery Grant, Canada Foundation of Innovation (CFI) for a Leaders Opportunity Fund, and Mount Saint Vincent University for an internal research grant.

References

- 1 Lipmann, F. (1941) Metabolic generation and utilization of phosphate bond energy. *Adv. Enzymol.*, **1**, 99–162.
- 2 Kalckar, H.M. (1941) The nature of energetic coupling in biological syntheses. *Chem. Rev.*, **28**, 71–178.
- 3 Szent-Györgyi, A. (1957) *Bioenergetics*, Academic Press, New York.
- 4 McClare, C.W.F. (1972) In defence of the high energy phosphate bond. *J. Theor. Biol.*, **35**, 233–246.
- 5 Ramasarma, T. (1998) A profile of adenosine triphosphate. *Curr. Sci.*, **74**, 953–966.
- 6 Guérin, B. (2004) *Bioénergétique*, EDP Sciences, Les Ulis, France.
- 7 Admiraal, S.J. and Herschlag, D. (1995) Mapping the transition state for ATP hydrolysis: implication for enzymatic catalysis. *Chem. Biol.*, **2**, 729–739.
- 8 Saint-Martin, H., Ruiz-Vicent, L.E., Ramirez-Solis, A., and Ortega-Blake, I. (1996) Toward an understanding of the hydrolysis of Mg-PP_i. An *ab initio* study of the isomerization reactions of neutral and anionic Mg–pyrophosphate complexes. *J. Am. Chem. Soc.*, **118**, 12167–12173.
- 9 Franzini, E., Fantucci, P., and Gioia, L. De. (2003) Density functional theory investigation of guanosine triphosphate models: catalytic role of Mg²⁺ ions in phosphate ester hydrolysis. *J. Mol. Catal. A: Chem.*, **204–205**, 409–417.
- 10 Akola, J. and Jones, R.O. (2003) ATP hydrolysis in water: a density functional study. *J. Phys. Chem. B*, **107**, 11774–11783.
- 11 Williams, N.H. (2004) Models for biological phosphoryl transfer. *Biochem. Biophys. Acta*, **1697**, 279–287.
- 12 Mao, L., Wang, Y., Liu, Y., and Hu, X. (2004) Molecular determinants for ATP-binding to proteins: a data mining and quantum chemical analysis. *J. Mol. Biol.*, **336**, 787–807.
- 13 Herschlag, D. and Jencks, W.P. (1989) Phosphoryl transfer to anionic oxygen nucleophiles. Nature of the transition state and electrostatic repulsion. *J. Am. Chem. Soc.*, **111**, 7587–7596.
- 14 Luo, Y.-R. (2003) *Handbook of Bond Dissociation Energies in Organic Compounds*, CRC Press, New York.
- 15 Matta, C.F., Castillo, N., and Boyd, R.J. (2006) Atomic contributions to bond dissociation energies in aliphatic hydrocarbons. *J. Chem. Phys.*, **125**, 204103_1–204103_13.
- 16 Matta, C.F., Arabi, A.A., and Keith, T.A. (2007) Atomic partitioning of the dissociation energy of the PO(H) bond in hydrogen phosphate anion (HPO₄²⁻): disentangling the effect of Mg²⁺. *J. Phys. Chem. A*, **111**, 8864–8872.
- 17 Bader, R.F.W. (1990) *Atoms in Molecules: A Quantum Theory*, Oxford University Press, Oxford, UK.
- 18 Popelier, P.L.A. (2000) *Atoms in Molecules: An Introduction*, Prentice Hall, London.
- 19 Matta, C.F. and Boyd, R.J. (Eds.) (2007) *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*, Wiley-VCH, Weinheim.
- 20 Arabi, A.A. and Matta, C.F. (2009) Where is energy stored in adenosine

- triphosphate? *J. Phys. Chem. A*, **113**, 3360–3368.
- 21 Keith, T.A. to be published.
 - 22 Löwdin, P.-O. (1959) Scaling problem, virial theorem, and connected relations in quantum mechanics. *J. Mol. Spectrosc.*, **3**, 46–66.
 - 23 Magnoli, D.E. and Murdoch, J.R. (1982) Obtaining self-consistent wave functions which satisfy the virial theorem. *Int. J. Quantum Chem.*, **22**, 1249–1262.
 - 24 Kohn, W. and Sham, L.J. (1965) Self consistent equations including exchange and correlation effects. *Phys. Rev. A*, **140** (4A), 1133–1138.
 - 25 Becke, A.D. (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, **38**, 3098–3100.
 - 26 Becke, A. (1993) A new mixing of Hartree–Fock and local density-functional theories. *J. Chem. Phys.*, **98**, 1372–1377.
 - 27 Nagy, A. (1992) Regional virial theorem in density-functional theory. *Phys. Rev. A*, **46**, 5417–5419.
 - 28 Levy, M. and Perdew, J.P. (1985) Hellmann–Feynman, virial, and scaling requisites for the exact universal density functionals. Shape of the correlation potential and diamagnetic susceptibility for atoms. *Phys. Rev. A*, **32**, 2010–2021.
 - 29 Sule, P. (1996) Kinetic contribution to the correlation energy density: benchmark to $T_c[n]$ energy functionals. *Chem. Phys. Lett.*, **259**, 69–80.
 - 30 Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Montgomery, J.A., Jr., Vreven, T., Kudin, K.N., Burant, J.C., Millam, J.M., Iyengar, S.S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G.A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J.E., Hratchian, H.P., Cross, J.B., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev, O., Austin, A.J., Cammi, R., Pomelli, C., Ochterski, J.W., Ayala, P.Y., Morokuma, K., Voth, G.A., Salvador, P., Dannenberg, J.J., Zakrzewski, V.G., Dapprich, S., Daniels, A.D., Strain, M.C., Farkas, O., Malick, D.K., Rabuck, A.D., Raghavachari, K., Foresman, J.B., Ortiz, J.V., Cui, Q., Baboul, A.G., Clifford, S., Cioslowski, J., Stefanov, B.B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R.L., Fox, D.J., Keith, T., Al-Laham, M.A., Peng, C.Y., Nanayakkara, A., Challacombe, M., Gill, P.M.W., Johnson, B. Chen, W. Wong, M.W. Gonzalez, C., and Pople, J.A. (2003) *Gaussian 03*, Gaussian Inc, Pittsburgh PA.
 - 31 Biegler-König, F.W., Bader, R.F.W., and Tang, T.-H. (1982) Calculation of the average properties of atoms in molecules. II. *J. Comput. Chem.*, **13**, 317–328.
 - 32 Keith, T.A. (2009) AIMALL (aim@tkgristmill.com).
 - 33 Biegler-König, F.W., Schönbohm, J., and Bayles, D. (2001) AIM2000: a program to analyze and visualize atoms in molecules. *J. Comput. Chem.*, **22**, 545–559.
 - 34 Senter, P., Eckstein, F., and Kagawa, Y. (1983) Substrate metal–adenosine 5′-triphosphate chelate structure and stereochemical course of reaction catalyzed by the adenosine triphosphatase from the thermophilic bacterium PS3. *Biochemistry*, **22**, 5514–5518.
 - 35 Dittrich, M., Hayashi, S., and Schulten, K. (2003) On the mechanism of ATP hydrolysis in F_1 -ATPase. *Biophys. J.*, **85**, 2253–2266.
 - 36 Dittrich, M., Hayashi, S., and Schulten, K. (2004) ATP hydrolysis in the β_{TP} and β_{DP} catalytic sites of F_1 -ATPase. *Biophys. J.*, **87**, 2954–2967.
 - 37 Dittrich, M. and Schulten, K. (2005) Zooming in on ATP hydrolysis in F_1 . *J. Bioenerg. Biomembr.*, **37**, 441–444.
 - 38 Pullman, B. and Pullman, A. (1963) *Quantum Biochemistry*, Interscience Publishers, New York.
 - 39 Bianchi, R., Gervasio, G., and Marabello, D. (2000) Experimental electron density analysis of $Mn_2(CO)_{10}$: metal–metal and metal–ligand bond characterization. *Inorg. Chem.*, **39**, 2360–2366.
 - 40 Yoshikawa, K., Shinohara, Y., Terada, H., and Kato, S. (1987) Why is Mg^{2+} necessary for specific cleavage of the terminal phosphor group of ATP? *Biophys. Chem.*, **27**, 251–254.

- 41 Matta, C.F. and Hernández-Trujillo, J. (2003) Bonding in polycyclic aromatic hydrocarbons in terms of the electron density and of electron delocalization. *J. Phys. Chem. A*, **107**, 7496–7504 (Correction: *J. Phys. Chem A*, 2005, **109**, 10798).
- 42 Matta, C.F. and Gillespie, R.J. (2002) Understanding and interpreting electron density distributions. *J. Chem. Educ.*, **79**, 1141–1152.
- 43 Wright, J.S., Rowley, C.N., and Chepelev, L.L. (2005) A ‘universal’ B3LYP-based method for gas-phase molecular properties: bond dissociation enthalpy, ionization potential, electron and proton affinity and gas-phase acidity. *Mol. Phys.*, **103**, 815–823.
- 44 DiLabio, G.A. (1999) Using locally dense basis sets for the determination of molecular properties. *J. Phys. Chem. A*, **103**, 11414–11424.
- 45 Pratt, D.A., Wright, J.S., and Ingold, K.U. (1999) Theoretical study of carbon–halogen bond dissociation enthalpies of substituted benzyl halides. How important are polar effects? *J. Am. Chem. Soc.*, **121**, 4877–4882.
- 46 DiLabio, G.A. and Wright, J.S. (1998) Calculation of bond dissociation energies for large molecules using locally dense basis sets. *Chem. Phys. Lett.*, **297**, 181–186.
- 47 Bader, R.F.W. and Nguyen-Dang, T.T. (1981) Quantum theory of atoms in molecules-Dalton revisited. *Adv. Quantum Chem.*, **14**, 63–124 (p. 118).

Part Three

Reactivity, Enzyme Catalysis, Biochemical Reaction Paths and Mechanisms

16

Quantum Transition State for Peptide Bond Formation in the Ribosome

Lou Massa, Chérif F. Matta, Ada Yonath, and Jerome Karle

16.1

Introduction

Crystallography is the principal method used to determine the structure of a ribosome, and consequently for understanding its functions, including formation of the peptide bond by ribozyme catalysis, and the decoding of the genetic code [1–5]. As shown in Figure 16.1, the ribosome is made of two subunits. It was found that the mRNA is decoded at the small subunit. The peptide bond is formed on the larger subunit within a cavity, hosting the peptidyl transferase center (PTC), composed mainly of ribosomal RNA [1–8].

Of importance to the quantum calculations reviewed in this chapter, a region of “pseudo twofold symmetry,” which was detected in all known ribosome structures in and around the PTC [1,3b,3c,3d,3e], is associated with the translocation of the aminoacylated tRNA through the ribosome, as peptide bond formation occurs [2, 3], navigated by the ribosome architecture (Figure 16.2). The nascent proteins move out of the ribosome via an exit tunnel whose opening lies adjacent to the PTC and receives thereby each successive peptide bond as the protein elongates. Thus the architecture of the ribosome is consistent with the requirements of peptide bond catalysis and protein formation [2, 3, 5, 9, 10].

Given the structural architecture of the ribosome, quantum crystallography (QCr) [11] may be applied to study the transition state (TS) for peptide bond formation. (The foundations and applications of QCr are reviewed in Chapter 1.) QCr combines crystallographic structural information with quantum mechanical theory. This facilitates theoretical calculations and adds an energetic aspect to crystallography. The crystallographic structure is a starting point, constraint and anchor for the quantum calculations. In QCr the molecular system is mathematically divided into computationally tractable pieces. Subsequently, this may be followed by a quantum investigation of their mutual interactions, and thus in a step-by-step manner one may rebuild the entire quantum mechanism as a whole. This approach has been applied to the investigation of the peptide bond TS. The first step here was to

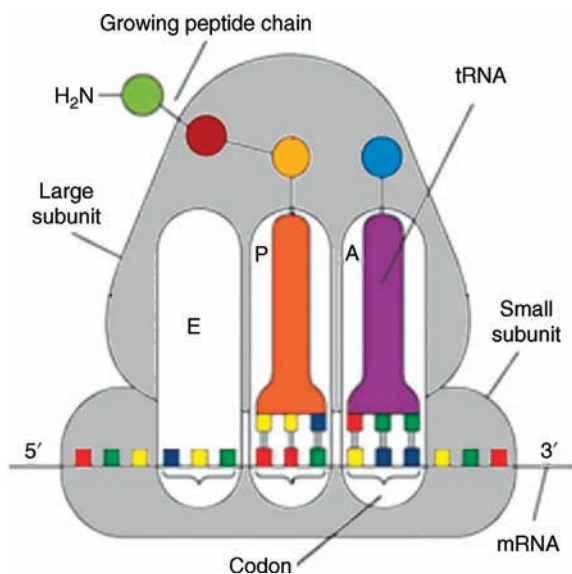


Figure 16.1 Protein synthesis [6]. Production line during protein synthesis, incoming tRNA (purple) carrying the next amino acid (blue circle) enters the A site if its anticodon (three “teeth” on its bottom) is complementary in sequence to the codon on mRNA. The reaction (not shown) between A-site tRNA and P-site tRNA (orange) extends the peptide chain by one amino acid unit. (Reproduced with the permission of the American Chemical Society from Reference [6].)

choose those atoms most likely to be importantly involved in the mechanism of peptide bond formation. This choice is small enough to be rigorously treated in density functional theory (DFT) quantum mechanics, but presumably large enough to represent the TS mechanism of peptide bond formation in the ribosome. Of course, this first choice can be successively expanded in future investigations.

A quantum mechanical TS for formation of the peptide bond has been found [12]. It is characterized by means of geometry, activation energy, thermodynamic parameters and quantum topology. The relevance of all this to peptide bond formation in the ribosome is discussed.

16.2

Methodology: Searching for the Transition State and Calculating its Properties

In this work, the Kohn–Sham equations of DFT were used to obtain the transition state for the peptide bond formation within the ribosome. The calculations included those 50 atoms assumed to be essential to peptide bond formation in the ribosome. Quantum mechanical calculations were carried out with the Mulliken program package [13]. The Becke three-parameter-hybrid (B3) [14] was used in conjunction with the Lee–Yang–Parr (LYP) functional [15] in all calculations, and a gaussian-type

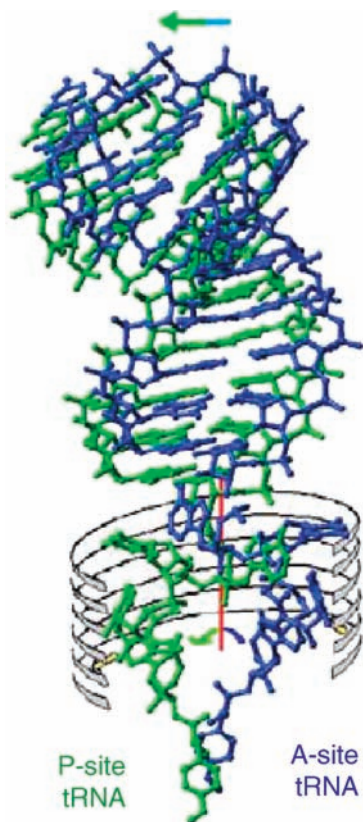


Figure 16.2 Schematic indication of the combined linear and rotational motions associated with the movements of the aminoacylated tRNA through the ribosome. The twofold axis, shown in red, points towards the exit tunnel through which the elongating proteins escapes the

ribosome. The apparent overlap of the two tRNA stems is a result of the specific view chosen to show best the concerted motions. (Reproduced with the permission of the National Academy of Science of the United States of America from Reference [12].)

basis set, 6-31+G(d,p) was used. In this manner, geometries of all reactants, products and transition states have been optimized at the DFT-B3LYP/6-31+G(d,p) level of theory.

Harmonic vibrational frequencies have been calculated using the same approximation for characterization of the nature of stationary points and zero-point vibrational energy (ZPVE) corrections. All the stationary points have been positively identified as minima with no imaginary frequencies, and the TS as a saddle point on the energy surface with one imaginary frequency. The bonds that are at the point of making and breaking in a transition state structure are consistent with a transformation connecting the desired reactants and products associated with peptide bond formation. The Cartesian coordinates of all atom positions in the optimized TS and calculated values of the vibrational frequencies are provided in Reference [12].

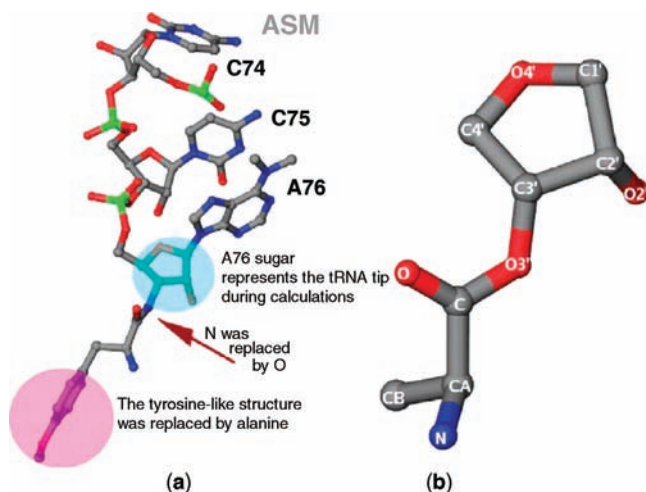


Figure 16.3 The 3'-end of the tRNA analog. (a) Tip of the tRNA ASM taken from the experimental crystal structure of its complex with D50S (Protein Data Bank ID code 1NJF), as used for the quantum mechanics calculations. The modified regions are highlighted by cyan and

magenta. Hydrogen atoms are not shown. (b) Sugar moiety at tip of tRNA, charged with alanine. (Reproduced with the permission of the National Academy of Science of the United States of America from Reference [12].)

The crystal structure of a 50S large ribosomal subunit from *Deinococcus radiodurans* complexed with a tRNA acceptor stem mimic (ASM) was used ([2], 1NJF in Protein Data Bank). Figure 16.3a shows a small part of this structure, the 3' end of the aminoacylated tRNA analog (ending with the highlighted sugar ring) attached via nitrogen to a tyrosine-like molecule (taking advantage of the non-hydrolysable nitrogen of the tRNA 3' end analog, puromycin). Because, in protein synthesis, amino acids attach to tRNA via ester-type bond we replaced N with an O at that location in the image shown. The highlighted region of the image contains atoms that have been judged to be of importance to the formation of the peptide bond. There are two analogous sets of such atoms, one is located in the A-site of the PTC and the other in the P-site, which was derived from A-site tRNA by rotation around the twofold axis. Both sets together constitute the 50 atoms chosen to represent the formation of the TS.

As shown in Figure 16.3a and b (hydrogen atoms not shown), we used the sugar moiety to represent the tip of tRNA, to investigate the actual reaction and, because of computational considerations, replaced the tyrosine-like bound amino acid structure with an alanine.

The TS results from a search that, except for initial conditions, is an automatic search, which only stops at a convergence satisfying stringent mathematical criteria. That occurs for a geometry that is at an energy minimum for every direction of displacement except one, for which it is at an energy maximum along a displacement toward products and away from reactants. A TS is a saddle point on the potential

energy surface, at which there occurs exactly one imaginary vibrational frequency, with all others real. The DFT quantum computations allow all 50 atoms to move freely, until a mathematically well-characterized TS is found. In terms of corroborating a TS, the bonds that are “making” and “breaking” must be consistent with the chemistry of the reaction.

The geometry of the TS, together with the twofold symmetry of the PTC [2, 3, 5, 10], has been used to estimate the angle of rotation of the A-site tRNA at the point of peptide bond formation. We made the estimate of rotation to the point of peptide bond TS formation by using coordinates of simulated ASM rotation every 15° about the twofold axis of the PTC. Superimposing our TS sugar moiety corresponding to the A site onto the acceptor stem mimic (ASM) sugar moiety, we let the TS ride around the twofold axis, looking for an angle that brought the second sugar moiety of the TS into best coincidence with its analog ring at the P site. We assumed that the position of the tRNA at the P-site is fixed, and it is the motion of the A site tRNA in its swing about the twofold axis that brings the reacting amino acids into coincidence. At each 15° increment of rotation we optimized the superposition of a TS sugar moiety onto that of the A site. In an analogous way we optimized a superposition of the TS sugar moiety onto that of the P site. Because the 50 atoms of the TS have been optimized independently of the tRNAs at A and P sites, it is not possible for the TS to fit them both simultaneously. Thus, we defined a best “average” position of the TS as occurring at the midpoint along a linear transformation between the optimal superpositions on the A and P sites. Using an objective error measure method [12], based upon the distance between analogous atoms at the average position of the TS and the A and P sites, we found the best match of our TS to the positions of the A and P sites to occur at a rotation angle of approximately 45° .

The thermodynamic parameters of the reaction that leads to TS in the ribosome have been measured in experiments. The corresponding parameters have been estimated for our theoretical TS, and they are found to be in qualitative proximal agreement with the experimental results [16].

A particular hydrogen atom, originally attached to the nitrogen of the A-site amino acid, has been suggested to participate in a “shuttle mechanism” (referenced below) during peptide bond formation. To study this proposed mechanism we carried out a topological study with methods of the quantum theory of atoms in molecules (QTAIM) (referenced below). Using the optimized geometry of the transition state [12], the Kohn–Sham (KS) [17] density functional theory (DFT)-B3LYP/6-31+G(d,p) level [14, 15] was used to define the origin of the intrinsic reaction coordinate axis (IRC) [18, 19]. The IRC is defined as “the minimum energy reaction pathway (MERP) in mass-weighted Cartesian coordinates between the transition state of a reaction and its reactants and products” [20]. (Gaussian 03 [21] was used in all electronic structure calculations.)

The evolution of the reaction was followed along the paths of steepest descent from the saddle point on both sides leading to the reactants and products, respectively. The initial direction of descent from the TS was that of the vibrational mode exhibiting an imaginary frequency. The reaction path was calculated without geometrical constraints using the algorithm of Gonzalez and Schlegel [22, 23]

sampling the path at 15 points on each side of the barrier at steps of $0.1 \text{ amu}^{1/2} \text{ bohr}$. Single point calculations were performed using 31 optimized geometries along the IRC (15 before and 15 after the TS in addition to the TS geometry). The resulting KS electron densities were subsequently analyzed according to the QTAIM [24–26] using the automated Windows implementation of AIMPAC [27, 28], AIMALL (T.A. Keith, personal communication 2009) and AIM2000 ([29–31]). The Poincaré–Hopf topological relationship [23] was verified for all points on the potential energy surface to ensure that no critical point had been missed.

16.3

Results: The Quantum Mechanical Transition State

Figure 16.4 shows the image of the optimized TS geometry for the formation of the peptide bond in the ribosome, including key geometrical parameters. The optimized TS bond distances are labeled according to whether they are in the act of breaking or forming, to achieve the transition from reactants to products. The end result is that the peptide bond N–C is formed, which results in elongating the nascent protein attached to the rotating A-site tRNA. The new O–H bond that is formed on the P-site tRNA saturates the open valence of the oxygen atom, which would occur as the C–O

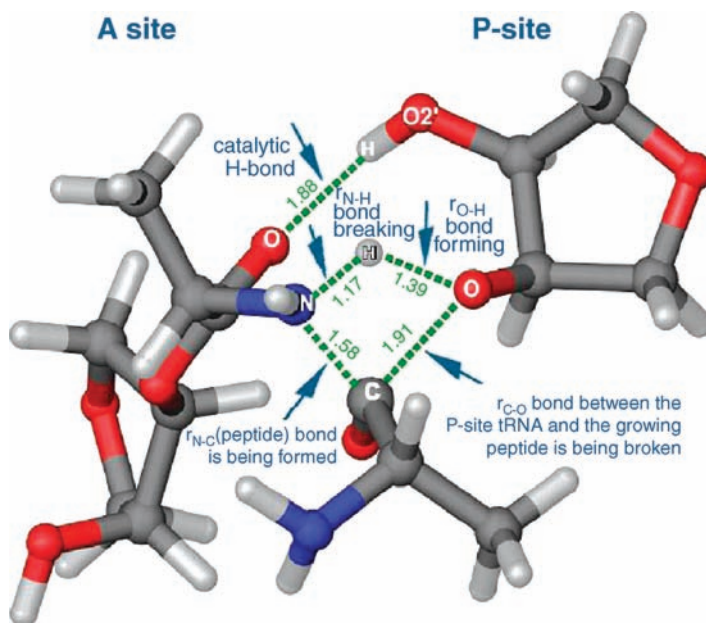


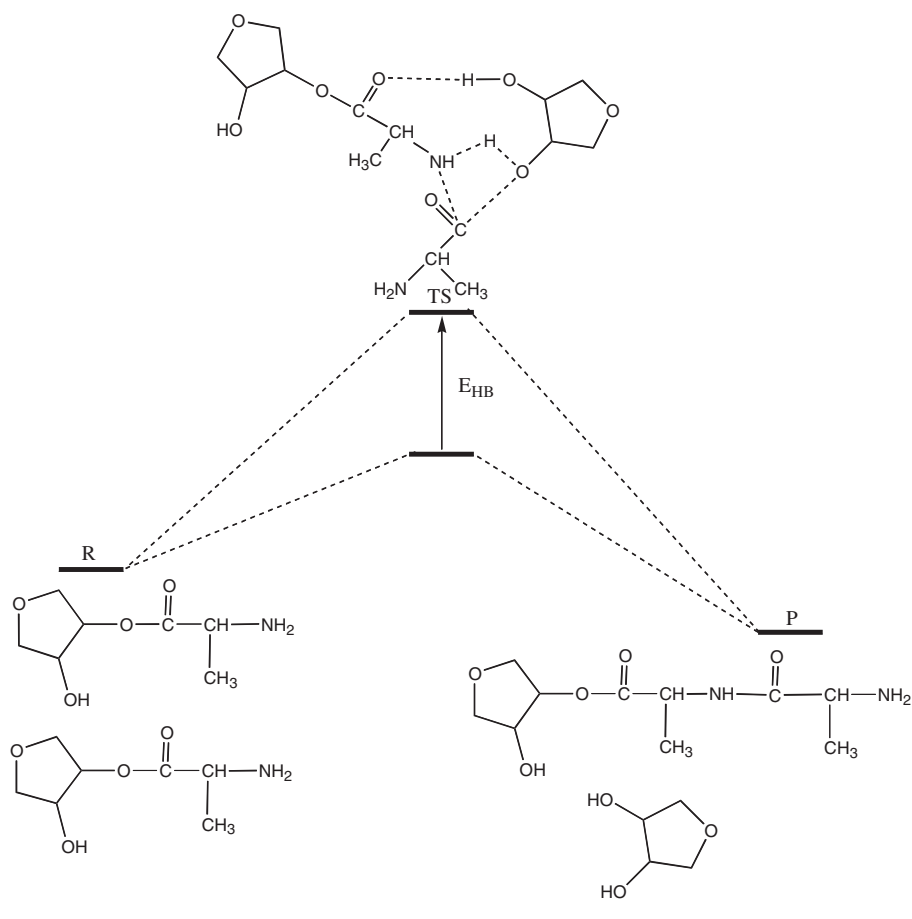
Figure 16.4 Peptide bond transition state in the ribosome. The amino acids are alanine. (Reproduced with permission of the National Academy of Science of the United States of America from Reference [12].)

bond breaks to allow release of the amino acid transferred to the nascent protein. A bond that is breaking in the TS, namely, N–H, completes the release of the P-site tRNA. Hence, given such bond making and breaking, the former A-site tRNA can occupy the P-site which becomes available by the former P-site tRNA release.

The TS geometry of Figure 16.4 shows the 2'OH of the P-site forming a hydrogen bond with the carboxyl oxygen of the A-site amino acid. That hydrogen bond is formed in the TS, having a bond length 1.879 Å. Such hydrogen bonding, perhaps serving as an anchor holding reactants in place at the TS, is consistent with the catalytic role that has been ascribed to the tRNA A76 2'OH group based on biochemical experiments [9]. Careful examination of Figure 16.4, which conveys something of the three-dimensional arrangements of the atoms in the TS, allows one to perceive how the peptide bond is being formed, and how the P site tRNA is allowed to break away after the peptide bond is being made.

Our TS has a calculated activation energy, E_a , of 35.5 kcal mol⁻¹ (1 kcal = 4.184 kJ). However, we found that in the ribosome the number of hydrogen bonds, between the rotating moiety of the tRNA aminoacylated 3'end and the surrounding nucleotides of the PTC, increases as the reactants move toward the transition state, resulting in lower activation energy. The number of hydrogen bonds, based upon a distance criterion that considers a hydrogen bond “cut off” at 4 Å, as a function of the angle of rotation about the twofold axis of symmetry in the PTC shows an increase of three hydrogen bonds as the transition state forms [12]. Assuming, on qualitative grounds, that such hydrogen bonds might vary in strength over the range 2–10 kcal mol⁻¹ [32], an average value of 6 kcal mol⁻¹ is adopted for each of the three newly formed hydrogen bonds. This implies a net stabilization of 18 kcal mol⁻¹ that would reduce the calculated activation energy to a qualitatively estimated value of approximately 18 kcal mol⁻¹ (Table 16.1).

The amino acids that are the reactants in the TS reaction are attached to large tRNA molecules, which suppress their translation and rotation degrees of freedom. The electronic levels are assumed to be too widely spaced to contribute to entropy change. Therefore, we take the electronic, translational and rotational contributions to entropy to be zero [16]. Consequently, the conditions of the ribosome environment reduce the change of entropy to that associated only with the vibrational degrees of freedom; that is, only the vibrational frequencies of the normal modes at the optimized geometries for the TS and reactants are required to obtain the entropies. These have been obtained using the Gaussian program package [21]. For the non-catalyzed reaction the calculated entropy contribution to the free energy change is $T\Delta S_{\text{total}}^\ddagger = -14.6$ kcal mol⁻¹, corresponding to an enormous and unfavorable decrease in entropy [16]. This may be compared to the catalyzed reaction in which the TS is stabilized by the formation of three hydrogen bonds to the ribosome nucleotides, and in which the translational and rotational degrees of freedom are suppressed by the ribosome. In this case the estimated overall entropy contribution to the free energy change is $T\Delta S_{\text{vib} + 3\text{HB}}^\ddagger = 1.5$ kcal mol⁻¹, which corresponds to a favorable increase in entropy [16]. The calculated enthalpy changes ($\Delta H^\ddagger$) for the non-catalyzed and catalyzed reactions are known from previous work [12, 16] to be 34.3 and 16.3 kcal mol⁻¹, respectively. These enthalpies are obtained from the corresponding

Table 16.1 Calculated energies (using B3LYP/6-31+G(d,p) method) along the peptide bond formation pathway.^{a)}

Chemical species	Energy (au)	Relative (to reactants) energy (kcal mol ⁻¹)
R	-1259.78 590	0.0
TS	-1259.72 318	35.5
P	-1259.79 099	-3.2

a) ΔE_{HB} represents the qualitative reduction in our calculated transition state activation energy that would be expected to occur because of increased hydrogen bonding concomitant with the reaction's progress towards the transition state. An increase of three hydrogen bonds, of average magnitude 6 kcal mol⁻¹, would be consistent with a qualitatively estimated transition state barrier of 18 kcal mol⁻¹. (Reproduced with the permission of the National Academy of Science of the United States of America from Reference [12].)

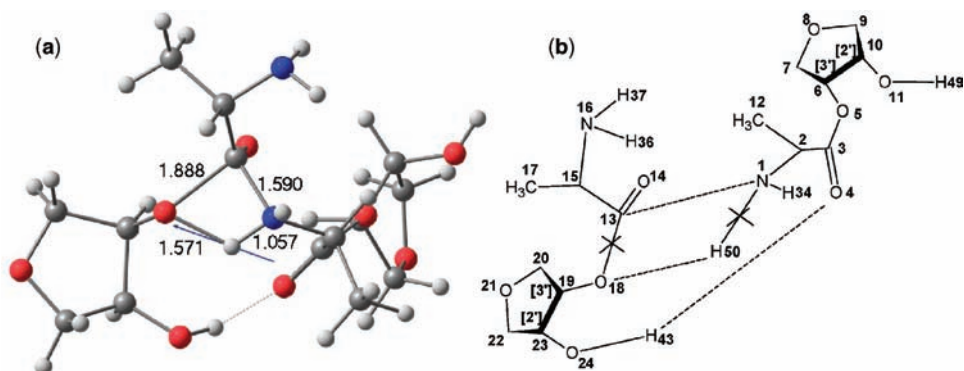


Figure 16.5 (a) Ball-and-stick model of the transition state with an arrow representation of the eigenvector of the single imaginary frequency ($\nu_{\text{im}} = -1084.13i \text{ cm}^{-1}$). The arrow clearly indicates the transfer of hydrogen from the amine nitrogen to the oxygen O3' (O18) of the P-site ribose sugar. The O2' hydroxyl group of

the P-site tRNA (O24–H43) forms a stable hydrogen bond, indicated by a dashed line, to the ester carbonyl group of the tRNA at the A-site (O4). (b) Diagram of the TS, showing the atom numbering scheme adopted in the discussion.

calculated activation energies E_a , which are 35.5 and $17.5 \text{ kcal mol}^{-1}$, respectively. Thus, the TS reaction within the ribosome is enhanced by both enthalpy and entropy relative to what would be the case for the same reaction in the gas phase. As regards entropy, its identification with noise allows the conclusion that the ribosome, by suppressing noise, contributes to catalysis of the peptide bond.

As can be seen from Figure 16.5, the eigenvector associated with the imaginary frequency ($\nu_{\text{im}} = -1084.13i \text{ cm}^{-1}$, in the harmonic approximation) is centered on H50, the hydrogen atom being transferred from the $-\text{NH}_2$ ($-\text{N1-H50-H34}$) to O3' (O18) of the P-site ribose sugar. This vector points in the direction of the reaction path when the system is at the TS point on the PES, which clearly indicates the transfer of H50 from the amine nitrogen to the oxygen. Figure 16.5 also provides the interatomic distances of the bonds that are forming (C–N and O–H) and breaking (N–H and O–C).

Figure 16.6 displays the molecular graph (the collection of bond paths) of the TS along with the number and type of critical points that satisfy the Poincaré–Hopf relationship. The lines of maximum electron density linking the (bonded) nuclei are the bond paths and the saddle points on those paths, indicated by the small red dots, are the bond critical points (BCPs). The yellow dots are the ring critical points. Each nuclear critical point is color-coded in the figure to reflect the identity of the atomic element. The Poincaré–Hopf relationship is (Equation 16.1):

$$n_{\text{NCP}} - n_{\text{BCP}} + n_{\text{RCP}} - n_{\text{CCP}} = 1 \quad (16.1)$$

where n_{NCP} is the number of nuclear critical points (50, in total), n_{BCP} the number of bond critical points (56), n_{RCP} the number of ring critical points (7) and n_{CCP} the number of cage critical points (none were found in the molecular graph of the TS).

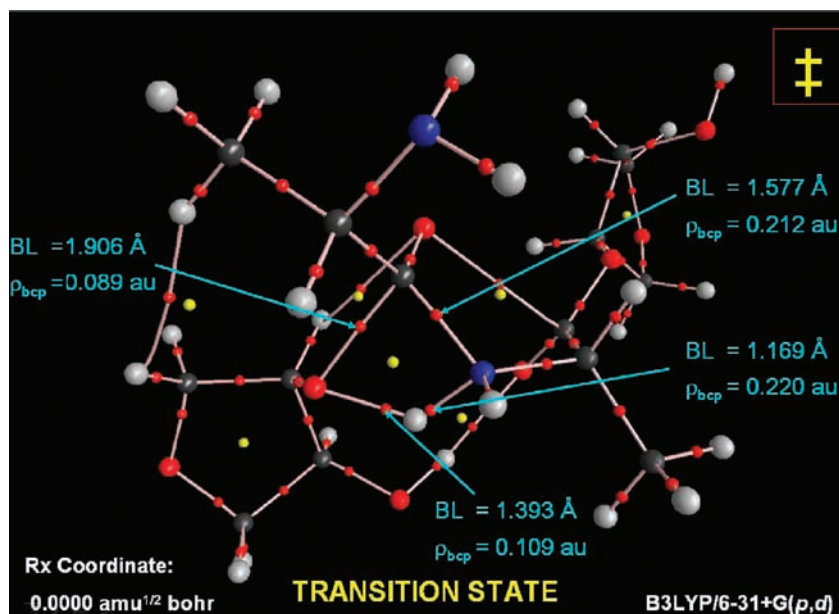


Figure 16.6 Molecular graph of the transition state. The large dark spheres indicate the nuclear critical points of carbon atoms, the large red sphere those of the oxygen nuclei, the blue spheres nitrogen nuclei and the large light gray spheres indicate the position of the hydrogen nuclear critical points. The lines of maximum electron density linking the (bonded) nuclei are

the bond paths and the saddle points on those paths, indicated by the small red dots, are the bond critical points (BCPs). The yellow dots are the ring critical points. (BL stands for “bond length.” The Poincaré–Hopf relationship (Equation 16.1) is satisfied by the molecular graph ($n_{\text{NCP}}(50) - n_{\text{BCP}}(56) + n_{\text{RCP}}(7) - n_{\text{CCP}}(0) = 1$).

The molecular graph in Figure 16.6 satisfies the Poincaré–Hopf relationship. In Figure 16.6 attention is drawn to the lines of maximum electron density linking the nuclei. These bond paths connect the hydrogen atom (H50) referred to in the shuttle mechanism to the oxygen O3', and *not* to oxygen O2'. This same bond path connection is preserved for all 15 points that have been calculated along the IRC beyond the TS moving towards products. The TS we characterized and the sequence of bond paths making and breaking precludes a shuttle mechanism in the present Ala-Ala system, and supports direct mechanism described above. Consistent with this, Figure 16.6 shows that the O2' hydroxyl group of the P-site tRNA (O24–H43) exhibits a bond path indicative of a hydrogen bond to the ester carbonyl group of the tRNA at the A-site (O4). The estimated energy of this hydrogen bond from the Espinosa–Molins–Lecomte (EML) empirical topological formula [33] is around 7 kcal mol^{-1} and remains constant in the segment of the reaction path we have studied. Its role appears to be to hold the reacting system in place for optimum orientation of the reacting groups. This hydrogen bond would be broken at later stages along the reaction path for the A-site and P-site reacting fragments to detach after formation of the peptide bond.

16.4

Discussion

The potential energy surface of 50 atoms considered to be most important in peptide bond formation has been calculated. Within the quantum mechanics of DFT (B3LYP) we have computed a molecular structure and energy that satisfies the mathematical criteria for a TS, including a frequency spectrum with all but one frequency real. The TS makes good chemical sense, in terms of what the amino acid molecules must do, namely, form a peptide bond, attach an elongating peptide to A-site tRNA as it moves to P-site, and have P-site tRNA separate from A-site tRNA. The chemical sense, after the mathematical criteria, is what corroborates the TS.

The calculated E_a of 35.5 kcal mol⁻¹ for our TS applies only to the barrier associated with those 50 atoms considered in the DFT calculation. However, qualitative considerations make clear how such an activation energy would be stabilized in the ribosome. During elongation the A-site tRNA carries out a “linear motion.” At the same time its 3' end, namely the amino acid attached to its CCA end, executes a rotational twofold motion. The combined linear and rotational motions of the full tRNA are indicated schematically in Figure 16.2. The number of hydrogen bonds associated with the rotating moiety of the tRNA 3' end within the PTC appears to increase by as much as three hydrogen bonds between 0 and 45° [12]. Adopting a reasonable average energy for such hydrogen bonds allows a qualitative estimate of the stabilization of the transition state that would occur. If every hydrogen bond confers ~6 kcal mol⁻¹, three such bonds would confer 18 kcal mol⁻¹ of stabilization. Thus, a qualitative estimate for the activation energy barrier for formation of the peptide bond in the ribosome would be approximately 18 kcal mol⁻¹. This qualitative estimate for the approximate E_a may be compared to the related (but different) experimental measurement [34], which has $E_a = 17.5$ kcal mol⁻¹; see also the related theoretical calculations of [35–37], all of which, however, deal with mechanisms different than our own.

Interestingly, the TS geometry is achieved after a modest rotation, which we estimate as 45°. At that stage the P-site O2' hydroxyl group forms a hydrogen-bond within the TS. Such an H-bond can stabilize the TS geometry, as recently suggested by biochemical studies [38]. This means that the TS for the peptide bond is made rather early in the rotation. However, the final bonds made and broken that result from the TS will achieve their equilibrium values after further rotation along the guiding reaction pathway associated with the twofold axis of the PTC.

We conclude that it is satisfactory that the DFT quantum computations, allowing all 50 atoms to move freely, have found a mathematically, well-characterized TS. In addition, the fact that the OH group at the P site ends up making a catalyzing hydrogen bond, in accordance with experiments that are generally agreed to be credible, underlines again the chemical sense our TS conveys.

The non-catalyzed reaction of the amino acids we have considered is associated with an enormous and unfavorable decrease in entropy related to translation and rotation degrees of freedom. However, in the ribosome these degrees of freedom are suppressed, by the tRNA attachments to the reactant amino acids, and because of

that suppression the catalyzed reaction shows a favorable increase in entropy [16]. Using the gas-phase non-catalyzed reaction as a standard of comparison, the ribosome environment enhances the formation of the peptide bond from both an enthalpy and an entropy point of view. The activation energy for formation of the TS is reduced by formation of three external hydrogen bonds. The entropy is increased by the suppression of translational and rotational degrees of freedom. The remaining vibrational degrees of freedom contribute to an increase of entropy. This is counteracted by a decrease of entropy associated with the formation of three hydrogen bonds. On balance, there remains an overall increase of entropy as the TS is formed. Both enthalpy and entropy contribute to ribosome amino acid reaction catalysis.

A so-called “proton shuttle” mechanism [36, 37] has been suggested in the ribosome literature. As may be seen in Figure 16.4, once the C–O bond is broken, between the P-site sugar and the growing peptide, the valence of the oxygen atom, that is O3', remains to be satisfied by bonding to a hydrogen atom. The proton shuttle proposal is one that entails the amino hydrogen from the A-site being passed over to the P-site O2' oxygen, which passes its own hydrogen within the P-site for valence satisfaction of its O3' oxygen. But with regard to our quantum mechanical TS, the shuttle mechanism is precluded. As shown in Figure 16.5, the vibrational motion of the amino hydrogen projects it directly towards the P-site O3', and away from O2'. Moreover, quantum topological arguments inveigh against the shuttle mechanism for the TS, as indicated in Figure 16.6, which shows the molecular graph for the TS. At equilibrium the molecular graph is made up of bond paths, that is, universal indicators of which atoms are bonded to one another [39]. Other than for geometries at stationary points on the potential energy surface [e.g., equilibrium geometries or first-order saddle points (TS)] these lines are called atomic interaction lines. A bond path is a ridge of electron density linking chemically-interacting nuclei and contributing to the stability (i.e., lowering) of the electronic potential energy for any nuclear configuration, at equilibrium or not. Consequently, the ridge of maximal density is the line along which the density contributes maximally to electronic potential energy stability. Following the TS, in the direction of the final chemical products, before a new equilibrium geometry is reached, a bonding interaction must develop along the reaction path (the IRC) that is indicative of the direction in which equilibrium lies. There is a continuity of meaning for atomic interaction lines and bond paths. If at a stationary point geometry there is to be a bond path between nuclei, its precursor is an interaction line that develops along the IRC. As regards the shuttle mechanism in the ribosome TS of Figure 16.6, the interaction line that develops along the IRC is the quantum topological definitive proof that there is no shuttle mechanism. This is because the proposed shuttle is inconsistent with the molecular graph drawn by interaction lines along the IRC that follows the TS. Once beyond the TS, that is for all 15 IRC energy calculations we have examined for this reaction, there is a consistent atomic interaction line between the amino hydrogen and the P-site O3' analogous to that shown in Figure 16.6 at the TS. Such a line connecting to the P-site O2' would be required for the shuttle mechanism to hold, but no such line occurs. Moreover, the O2' hydroxyl group of the P-site tRNA (O24–H43) exhibits a remarkably stable hydrogen bond path to the ester carbonyl group of the tRNA at the

A-site (O4). The estimated energy of this hydrogen bond from the EML formula of Reference [33] is around 7 kcal mol^{-1} and remains constant in the segment of the reaction path we have studied explicitly. This too is inconsistent with a proton shuttle involving O2'. Instead, the role of the O2' hydroxyl group appears to be, through formation of a hydrogen bond, to hold the reacting system in place for optimum reaction orientation. This hydrogen bond is broken at later stages along the reaction path to allow the for P-site tRNA to exit the ribosome. The mechanism presented here is simpler than the popular proton shuttle mechanism, inasmuch as it involves a direct transfer of hydrogen from the attacking $-\text{NH}_2$ group to the ester oxygen at the 3' carbon of the P-site sugar.

16.5

Summary and Conclusions

Quantum mechanics and crystallography have been joined to study the formation of the peptide bond as it occurs in the ribosome's peptidyl transferase center (PTC). Quantum calculations were based upon a choice of 50 atoms assumed to be important in the mechanism. Density functional theory (DFT) was used to optimize the geometry and energy of the transition state (TS) for peptide bond formation. The calculated transition state activation energy, E_a , is $35.5 \text{ kcal mol}^{-1}$. However, an increase in hydrogen bonding occurs between A-site tRNA and ribosome nucleotides during the twofold rotation from the A-site towards the P-site as the TS forms. The activation energy is stabilized by the increase in hydrogen bonding to a value qualitatively estimated to be approximately 18 kcal mol^{-1} . The optimized geometry of the TS corresponds to a structure in which the peptide bond is being formed as other bonds are being broken, in just such a manner as to release the P-site tRNA so that it may exit as a free molecule, and be replaced by its A-site analog attached to an elongating nascent protein. The entropy increase of the TS is estimated. The calculated thermal parameters of the TS are in qualitative agreement with corresponding experimental values. At TS formation the 2'OH group of the P-site tRNA A76 forms a hydrogen bond with the oxygen atom of the carboxyl group of the amino acid attached to the A-site tRNA, suggestive of a catalytic role, which is consistent with experimental findings. The estimated magnitude of the rotation angle about the ribosomal twofold pseudo-symmetrical axis, between the A-site starting position and the place at which the TS occurs, is approximately 45° . Using quantum topology we investigated a "shuttle mechanism," which has often been suggested in the literature to describe hydrogen atom transfer associated with peptide bond formation. The inconsistency between this mechanism and the quantum mechanical transition state is discussed.

Acknowledgments

Thanks are due to Professor Richard F. W. Bader for suggesting that a QTAIM analysis would shed light on the nature of the TS. We acknowledge ribosome studies

in collaboration with Asta Gindulyte, Anat Bashan and Ilana Agmon, [12]. L.M.'s studies were funded by U.S. Army, breast cancer award, W81XWH-06-1-0658, US National Institute of Health (NIGMS MBRS SCORE 5S06GM606654) and the National Center for Research Resources (RR-03037). C.M. acknowledges the Natural Sciences and Engineering Research Council of Canada (NSERC), Canada Foundation for Innovation (CFI) and Mount Saint Vincent University for funding. A.Y. was supported by National Institutes of Health Grant GM34360, Human Frontier Science Program Organization Grant RGP0076_2003 and the Kimmelman Center for Macromolecular Assemblies. A.Y. holds the Martin and Helen Kimmel Professorial Chair. The research at The Naval Research Laboratory was supported by the Office of Naval Research. The authors thank the National Academy of Science of the United States of America and the American Chemical Society for permissions to reproduce copyrighted material.

References

- Harms, J., Schlutzenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F., and Yonath, A. (2001) *Cell*, **107**, 679–688.
- Bashan, A., Agmon, I., Zarivach, R., Schlutzenzen, F., Harms, J., Berisio, R., Bartels, H., Franceschi, F., Auerbach, T., Hansen, H.A.S., Kossoy, E., Kessler, M., and Yonath, A. (2003) *Mol. Cell.*, **11**, 91–102.
- (a) Agmon, I., Bashan, A., Zarivach, R., and Yonath, A. (2005) *Biol. Chem.*, **386**, 833–844; (b) Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000) *Science*, **289**, 905–920; (c) Schuwirth, B.S., Borovinskaya, M.A., Hau, C.W., Zhang, W., Vila-Sanjurjo, A., Holton, J.M., and Cate, J.H.D. (2005) *Science*, **310**, 827–834; (d) Selmer, M., Dunham, C.M., Murphy Iv, F.V., Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R., and Ramakrishnan, V. (2006) *Science*, **313**, 1935–1942; (e) Korostelev, A., Trakhanov, S., Laurberg, M., and Noller, H.F. (2006) *Cell*, **126**, 1065–1077.
- Yonath, A. (2003) *Biol. Chem.*, **384**, 1411–1419.
- Yonath, A. (2005) *Mol. Cell*, **20**, 1–16.
- Borman, S. (2007) *Chem. Eng. News*, **85**(8), 13–16.
- Youngman, E.M., Brunelle, J.L., Kochaniak, A.B., and Green, R. (2004) *Cell*, **117**, 589–599.
- Brunelle, J.L., Youngman, E.M., Sharma, D., and Green, R. (2006) *RNA*, **12**, 33–39.
- Weinger, J.S., Parnell, K.M., Dorner, S., Green, R., and Strobel, S.A. (2004) *Nat. Struct. Mol. Biol.*, **11**, 1101–1106.
- Bashan, A. and Yonath, A. (2005) *Biochem. Soc. Trans.*, **33**, 488–492.
- Huang, L., Massa, L., and Karle, J. (2001) *IBM J. Res. Dev.*, **45**, 409–415.
- Gindulyte, A., Bashan, A., Agmon, I., Massa, L., Yonath, A., and Karle, J. (2006) *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 13327–13332.
- IBM, MULLIKEN. MULLIKEN is IBM proprietary software package that implements *ab initio* quantum chemical calculations on the IBM SP/2 supercomputer (The Laboratory for Quantum Crystallography, Hunter College, CUNY) (1995).
- Becke, A.D. (1993) *J. Chem. Phys.*, **98**, 5648–5652.
- Lee, C., Yang, W., and Parr, R.G. (1988) *Phys. Rev. B*, **37**, 785–789.
- Massa, L. (2007) Comment on the suppression of noise by the ribosome. The SPIE Symposium on Fluctuations and Noise, 20–24 May at the La Pietra Center in Florence, Italy.
- Kohn, W. and Sham, L.J. (1965) *Phys. Rev. A*, **140**, 1133–1138.

- 18 Fukui, K. (1970) *J. Phys. Chem.*, **74**, 4161–4163.
- 19 Fukui, K. (1981) *Acc. Chem. Res.*, **14**, 363–368.
- 20 Zipse, H. (2008) Following the intrinsic reaction coordinate, <http://www.cup.uni-muenchen.de/oc/zipse/compchem/geom/irc1.html>.
- 21 Frisch, M.J., Trucks, G.W., Schlegel, H.B. *et al.* (2003) Gaussian 03, Gaussian, Inc., Pittsburgh, PA.
- 22 Gonzalez, C. and Schlegel, H.B. (1989) *J. Chem. Phys.*, **90**, 2154.
- 23 Gonzalez, C. and Schlegel, H.B. (1990) *J. Phys. Chem.*, **94**, 5523–5527.
- 24 Bader, R.F.W. (1990) *Atoms in Molecules: A Quantum Theory*, Oxford University Press, Oxford, UK.
- 25 Popelier, P.L.A. (2000) *Atoms in Molecules: An Introduction*, Prentice Hall, London.
- 26 Matta, C.F. and Boyd, R.J. (eds) (2007) *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*, Wiley-VCH Verlag GmbH, Weinheim.
- 27 Biegler-König, F.W., Bader, R.F.W., and Tang, T.-H. (1982) *J. Comput. Chem.*, **13**, 317–328.
- 28 Bader, R.F.W. AIMPAC, <http://www.chemistry.mcmaster.ca/aimpac/>.
- 29 Biegler-König, F.W., Schönbohm, J., and Bayles, D. (2000) AIM2000, <http://gauss.fh-bielefeld.de/aim2000>.
- 30 Biegler-König, F.W., Schönbohm, J., and Bayles, D. (2001) *J. Comput. Chem.*, **22**, 545–559.
- 31 Biegler-König, F.W. (2000) *J. Comput. Chem.*, **21**, 1040–1048.
- 32 Lii, J.-H. (1998) in *Encyclopedia of Computational Chemistry* (ed. P.R. Schleyer), John Wiley & Sons, Ltd, pp. 1271–1283.
- 33 Espinosa, E., Molins, E., and Lecomte, C. (1998) *Chem. Phys. Lett.*, **285**, 170–173.
- 34 Sievers, A., Beringer, M., Rodnina, M.V., and Wolfenden, R. (2004) *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 7897–7901.
- 35 Das, S.R. and Piccirilli, J.A. (2005) *Nat. Chem. Biol.*, **1**, 45–52.
- 36 Sharma, P.K., Xiang, Y., Kato, M., and Warshel, A. (2005) *Biochemistry*, **44**, 11307–11314.
- 37 Trobro, S. and Aqvist, J. (2005) *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 12395–12400.
- 38 Huang, K.S., Weinger, J.S., Butler, E.B., and Strobel, S.A. (2006) *J. Am. Chem. Soc.*, **128**, 3108–3109.
- 39 Bader, R.F.W. (1998) *J. Phys. Chem. A*, **102**, 7314–7323.

17

Hybrid QM/MM Simulations of Enzyme-Catalyzed DNA Repair Reactions

Denis Bucher, Fanny Masson, J. Samuel Arey, and Ursula Röthlisberger

17.1

Introduction

Deoxyribonucleic acid (DNA) is a fundamental molecule of life, since it contains the information that makes each species unique, and the biological instructions needed to construct all other components of cells, such as proteins and RNA molecules. The genomic integrity is subject to deterioration by reactive oxygen species produced during normal metabolism, radiation from the environment, toxic chemicals and natural degradation [1–3]. Once the exact sequence is lost no replacement is possible, since there are only two copies of each chromosome in the cell. For this reason, all cellular life forms and many viruses encode a multitude of proteins that function to faithfully repair the lesions inflicted on DNA. The DNA repair machinery has been classified into several broad pathways: direct damage reversal, base excision repair (BER), nucleotide excision repair (NER), mismatch repair and double-strand break repair, as detailed in recent reviews [4, 5].

The study of the structures and mechanisms of DNA repair enzymes is interesting for several reasons. First, DNA repair enzymes are outstandingly efficient natural catalysts when it comes to editing DNA, and their ability to extract and manipulate gene sequences is paving the way for important new applications in biotechnology. Recent examples include the design of engineered enzymes that can operate on specific gene sequences [6], and biomimetic catalysts that are modeled on natural systems [7]. Second, insights into DNA repair enzymes can lead to the development of new drugs that can either assist directly the repair of chemical alterations to the genetic code, or enhance cancer therapy by inhibiting the function of DNA repair enzymes in malignant cells [8–10]. In particular, today's emerging knowledge of mutations and polymorphisms in key human DNA-repair genes, coupled with the in-depth knowledge of DNA repair mechanisms, is likely to provide a rational basis for improved strategies for therapeutic interventions on several tumors and degenerative disorders [11].

Over the past 20 years, our knowledge of DNA repair mechanisms has dramatically improved. More than 150 human genes associated with DNA repair have been identified (www.cgal.icnet.uk/DNA_Repair_Genes.html), and many of these genes have been associated with increased longevity in various organisms [12]. In addition, ~700 structures of DNA repair enzymes have been solved and deposited in the Protein Data Bank (PDB). Although crystallographic studies have provided important structural information about DNA repair enzymes, in many cases the reaction mechanisms of DNA repair enzymes are not known. The complexity and large size of DNA repair enzymes makes the experimental determination of the mechanisms very difficult.

In this chapter, examples of current applications are described to illustrate the possibilities of first-principles simulations to tackle DNA repair issues. Twenty years ago, most *ab initio* methods were only capable of modeling systems consisting of a few atoms – hence the applicability to study real enzymatic systems was very limited at that point. Today, first-principles simulations on large systems can be performed if several simplifications and approximations are used. These approximations are reviewed in Section 17.2, where some aspects of the methodology are detailed. In Section 17.3, we discuss applications that illustrate how computational methods can be used to single out the role of different residues in catalysis, and to compare different mechanistic hypotheses by computing the free energy paths along well-chosen reaction coordinates. In the first example, we describe the repair of thymine dimers by DNA photolyase. The simulations shed light on the kinetics of the reaction, assisting the interpretation of experiments. In the second example, we discuss the reaction mechanism, and the catalytic role of the metal center, in the DNA repair enzyme Endonuclease IV. Finally, in the last example, we discuss the BER enzyme MutY. Particular focus is devoted to evaluating the role of structured waters in the catalytic mechanism of MutY. In Section 17.4, we conclude with some general remarks.

17.2

Theoretical Background

Typical enzymatic processes, such as the repair of DNA by enzymes, involve system sizes of thousands of atoms in aqueous solution and can span time scales from millisecond to seconds. The size and the complexity of these systems is such that the use of quantum mechanics in studying DNA repair has been limited. However, important theoretical developments have revitalized the field and made recent applications possible. These developments include: (i) the use of quantum mechanics/classical mechanics (QM/MM) schemes to extend the size of the systems that can be studied, (ii) density functional theory (DFT) to model the electron–electron interactions, (iii) pseudopotential theory to model the electron–ion interactions and (iv) thermodynamic integration techniques to compute the free energy along possible reaction coordinates.

The QM/MM scheme, introduced by Warshel and Levitt in 1976 [14], has become widely used in recent years to investigate chemical reactions that occur in a complex

and heterogeneous environment (see also Chapters 2–4). In the QM/MM scheme, the chemically active part of an enzymatic system is described using QM methods, while the rest of the system (the solvent, counter-ions and the rest of the protein) is described using empirical force fields. Such a hierarchical approach has the advantage that the computational effort can be concentrated on the part of the system where it is most needed, whereas the effects of the surroundings are taken into account with a more expedient model.

A QM/MM description can be coupled with first-principles molecular dynamics to obtain better equilibrium structures, and to estimate the kinetic and thermodynamic properties of the systems. The implementation of the QM/MM scheme [16] used here is designed to work in conjunction with the first-principles molecular dynamics code CPMD [15]. Here, we limit ourselves to a very brief description of the method, since an in-depth description can be found elsewhere [13]. In this QM/MM scheme, the total energy of the system is described as the sum of three contributions:

$$E_{\text{tot}} = E_{\text{QM}} + E_{\text{MM}} + E_{\text{QM/MM}} \quad (17.1)$$

which, in the language of operators, becomes:

$$\hat{H}_{\text{tot}} = \hat{H}_{\text{QM}} + \hat{H}_{\text{MM}} + \hat{H}_{\text{QM/MM}} \quad (17.2)$$

The QM/MM Hamiltonian can be expressed as:

$$\hat{H}_{\text{QM/MM}} = \hat{H}_{\text{el}} + \sum_{i \in \text{MM}} \sum_{j \in \text{QM}} \frac{Z_j q_i}{r_{ij}} + \sum_{i \in \text{MM}} \sum_{j \in \text{QM}} v_{\text{vdw}}(r_{ij}) + \hat{H}_{\text{bonded}} \quad (17.3)$$

where the subscripts i and j refer to classical interactions sites and QM nuclei, respectively. The basic equations are relatively simple. However, the description of the interface region can be non-trivial and, to complete the valence of QM atoms, capping hydrogen atoms or optimized carbon pseudopotentials [17] are used. In addition, the form for the electrostatic interaction Hamiltonian ($\hat{H}_{\text{el}}$) poses serious theoretical and technical problems, related to both its short-range and its long-range behavior. In its simplest form $\hat{H}_{\text{el}}$ can be written as:

$$\hat{H}_{\text{el}} = \sum_{i \in \text{MM}} q_i \int dr \rho_e(r) \quad (17.4)$$

A first issue is related to the fact that positively charged classical atoms can act as traps for electrons if the basis set is flexible enough to allow for this. The Pauli repulsion from the electron cloud that would surround the classical atoms is absent and, therefore, the electron density is overpolarized, at short range, by an incorrect purely attractive potential, giving rise to the so-called electron spill-out problem. This effect is particularly pronounced in a plane-wave basis-set approach, in which the electrons are fully free to delocalize, but can be of relevance also in schemes using localized basis sets, especially if extended basis sets with diffuse functions are used. To overcome this problem, the QM/MM implementation employs a

Coulomb potential that is suitably modified at short range [16]. A second problem is related to the computational cost to compute $\hat{H}_{\text{el}}$ (17.4) within a plane-wave scheme. This is resolved by using a multipolar expansion of the QM charge density to compute the long-range Coulomb interaction with the classical environment, thereby drastically reducing the number of operations to be performed.

First-principles simulations can be carried out by using the Born–Oppenheimer molecular dynamics approach. In this case, the many-body problem is reduced to the solution of the dynamics of the electrons in some frozen configuration of the nuclei. The nuclear forces are computed from the Hellmann–Feynman theorem by solving the DFT problem [18] (i.e., minimizing the Kohn–Sham energy functional) at each nuclear configuration. Alternatively, first-principles simulations can be carried out by using the Car–Parrinello (CP) extended Lagrangian approach [19]. Born–Oppenheimer and Car–Parrinello dynamics differ in the way the electronic variables are obtained along the nuclear dynamics. In the latter approach, the Kohn–Sham orbitals are imbued with a fictitious time dependence, that is, a classical dynamics for the orbitals is introduced that propagates an initially fully minimized set of orbitals to subsequent minima corresponding to each new nuclear configuration. This task is accomplished by designing the orbital dynamics in such a way that the orbitals are maintained at a “temperature” T_e that is much smaller than the real nuclear temperature T . The fictitious temperature T_e of the orbitals and the adiabatic decoupling from the nuclear dynamics is controlled by the choice of μ . The fictitious mass μ is chosen specifically to be as small as is feasible for accurate integration of the equations of motions with a reasonably large time step, thus allowing the orbitals to relax quickly in response to the nuclear motion. To solve the electronic structure problem, density functional theory (DFT) is used, exploiting the fact that DFT has a favorable scaling ($\sim N^3$) over the more accurate methods of quantum chemistry (N^5 to N^8). At present, DFT gives accurate results for relatively large systems (~ 100 – 1000 atoms) at a reasonable computational cost, which is ideal for the study of enzymatic mechanisms. Although DFT is formally exact, the form of the exchange–correlation energy functional, $E(\text{xc})[\rho]$, is unknown and must be approximated. The most common classes of approximations are the local-density, generalized-gradient and meta-generalized gradient approximations. One of the limitations of such approximations is the treatment of dispersion; however, new hybrid meta-GGA exchange–correlation functionals [20] and dispersion corrected atom-centered potentials [17] can improve this shortcoming significantly.

In most enzymatic systems, reaction barriers are of the order of 10 kcal mol^{-1} or higher ($1 \text{ kcal} = 4.184 \text{ kJ}$). As a consequence, enzymatic reactions often do not occur spontaneously on the time scale accessible with QM/MM simulations (picosecond timescale). Hence, various techniques have been developed to enhance the sampling of rare reactive events. Enhanced sampling techniques are often designed to: (i) accelerate the sampling of relevant regions of the free energy surface and (ii) enable integration of the potential energy to give the change in free energy between two or more distinct states. For the purposes of this chapter, one commonly used enhanced sampling method is briefly described: free energy

calculations based on thermodynamic integration along a constraint reaction coordinate [21].

In many enzymatic systems, a one-dimensional partial reaction coordinate can be proposed from visual inspection of the structures along with the available experimental data. Integration of the potential of mean force along the reaction coordinate becomes the method of choice. We briefly describe this thermodynamic integration method; however, a potential of mean force (PMF) may also be derived from a restraint such as an umbrella bias or other bias potentials or by occurrence averaging [22]. In the thermodynamic integration approach, the system is constrained such that the reaction coordinate is fixed at a given value, and all other degrees of freedom are allowed to propagate freely by molecular dynamics. For a given configuration of the system, a certain force is required to maintain the reaction coordinate constraint, and this constraint force may be calculated from the known QM/MM potential. If the system configuration space is sufficiently sampled at a given reaction constraint value, then the average constraint force will eventually converge to an ergodic limit. Hence, the mean constraint force can be estimated along different values of the reaction coordinate by using molecular dynamics methods to sample the corresponding configuration space. Finally, by integrating the mean constraint force with respect to the reaction coordinate of interest, the free energy profile of the corresponding reaction pathway is obtained.

A recent study has found that, compared to other approaches to estimate the PMF, applying a constraint to integrate the average force along a reaction coordinate is in fact the most efficient method to converge the PMF for the separation of two aqueously-dissolved methane molecules [22]. Note that it is possible to estimate the convergence of the mean of the constraint force at each reaction coordinate value using established statistical analyses [23].

17.3 Applications

17.3.1

Thymine Dimer Splitting Catalyzed by DNA Photolyase

The *cis-syn* pyrimidine dimer (cyclobutane pyrimidine dimer, CPD) is the major product induced by UV irradiation and is one of the principal causes of skin cancer [24, 25]. DNA photolyase – which is found in prokaryotes, plants and various animals, including frogs, fish and snakes – is a highly efficient light-driven enzyme that can recognize and repair the CPD lesion [26].

According to the most recent experimental study, the overall process can be outlined as follows: a quantum of light energy in the blue or near-UV range is initially absorbed by an antenna pigment (8-hydroxy-5-deazaflavinHDF or methenyltetrahydrofolate, MTHF) and transferred to a reduced flavin coenzyme (FADH^-) (Figure 17.1). The excited FADH^- donates an electron to the CPD lesion, leading to a

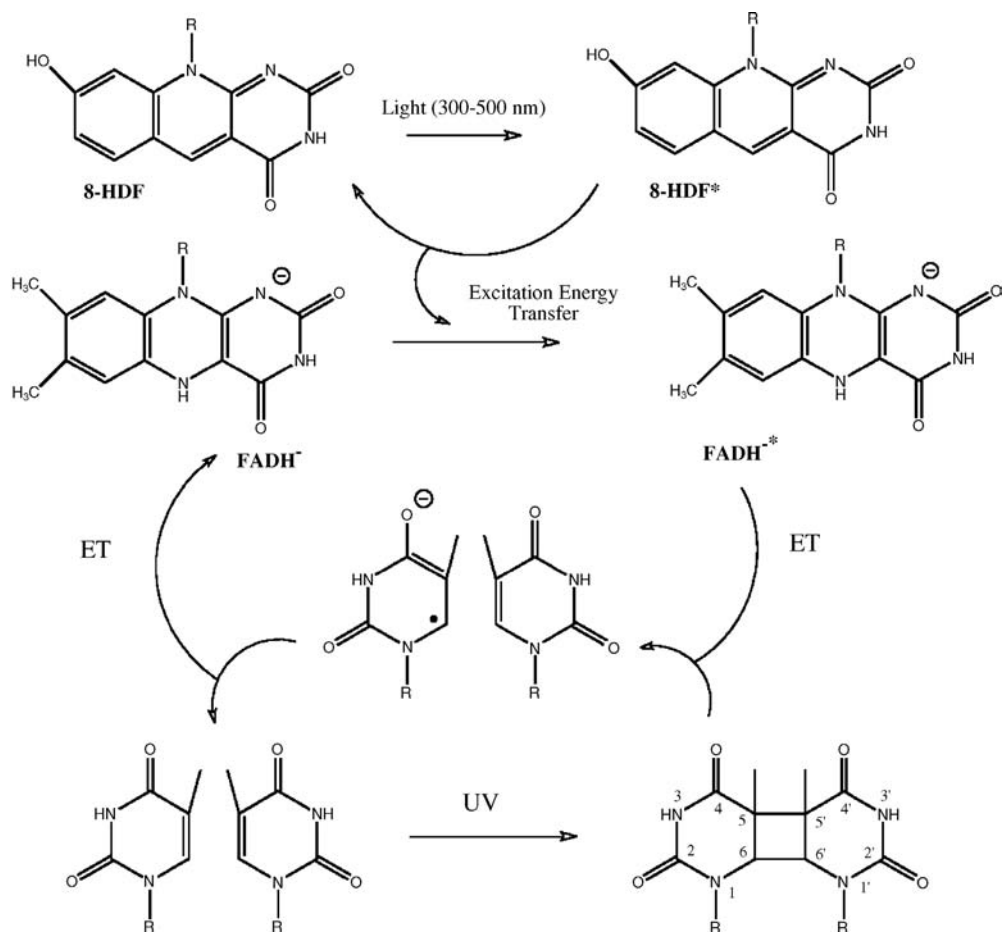


Figure 17.1 Mechanism of repair of cyclobutane pyrimidine dimers (CPD) by DNA photolyase; 8-HDF: 8-hydroxy-5-deazaflavin, FADH⁻: reduced and deprotonated flavin cofactor, ET: electron transfer. The atom numbering scheme is illustrated for the thymine dimer.

destabilization of the C5–C5' and C6–C6' bonds and to the conversion of the thymine dimer (T<T) into the original bases. Formation of the repaired thymine monomers is followed by an electron back-transfer to the FADH^{*} radical, restoring the catalytic species FADH⁻. The repair process can therefore be schematically divided into three characteristic steps: (1) the transient reduction of the CPD lesion by a reduced flavin cofactor FADH⁻, (2) the splitting reaction of the thymine dimer radical anion and (3) the electron back-transfer process to the FADH radical. Open questions relate to the detailed mechanism of the repair process: (i) the role of the active-site residues and water molecules in promoting the splitting reaction, (ii) the sequential or concerted nature of the splitting reaction of the thymine dimer radical

anion and (iii) the kinetics of the repair process. A significant number of transient absorption studies have been performed [27–30], showing partial disagreements, which reflects the difficulties associated with the measurement of the kinetic parameters of the photochemical repair process.

A computational study of the bond-breaking process can help clarify the few unresolved issues and offer a uniform interpretation of the available experimental data. From a theoretical point of view, investigation of these questions requires a mixed quantum/classical approach, because the splitting mechanism of the thymine dimer radical anion (in our simulation system located between T7 and T8, and containing about 30 atoms) is a quantum chemical process, but the description of the whole solvated enzyme–DNA complex (about 72 000 atoms) is only feasible within a classical framework. We performed a statistical analysis based on seven independent QM/MM trajectories (CPD1–CPD7) [31]. These simulations identified the enzyme-catalyzed repair reaction as an asynchronous concerted mechanism, in which the breaking of the C5–C5' bond is spontaneous upon electron uptake and is subsequently followed by barrier-less C6–C6' cleavage. The breaking process occurred spontaneously after C5–C5' bond cleavage (within 400 fs) in all reactive trajectories but one (CPD4), where the C6–C6' bond broke about 2800 fs afterwards. In the case of CPD4, we performed a metadynamics simulation [32] to estimate an upper limit to the free energy barrier characterizing the basin sampled by this configuration. A low energy barrier of $2.5 \text{ kcal mol}^{-1}$ was obtained. Careful inspection of the CPD4 configuration showed that important hydrogen bond and salt bridge interactions, present in the other six configurations, were missing in the non-reactive trajectory (see below). Therefore, the system in CPD4 is trapped in an unfavorable free energy basin characterized by an unusual hydrogen-bond pattern. Thus, the value of $2.5 \text{ kcal mol}^{-1}$, which can be easily overcome at room temperature, can be considered as the energy necessary to escape this local free energy minimum.

The atomic picture given by the QM/MM simulations can also provide new insights into the role of specific conserved DNA photolyase residues in promoting the splitting reaction. Glu283 is thought to stabilize the radical anion CPD by transferring a proton to O4(T7) (Figure 17.2) [33], and its mutation to alanine impairs enzyme activity by diminishing the quantum yield for the repair reaction by 60% [34]. Indeed, the simulations show a proton transfer from Glu283 to the C4 (T7) carbonyl oxygen of the thymine dimer radical anion in five out of seven trajectories. The observation that the ring splittings in trajectories CPD6 and CPD7 occur without protonation prompted us to also consider the roles of the positively charged side-chains of Arg232 and Arg350. For CPD6, the two side-chains are close enough to directly interact with O4(T7) and O2(T8), whereas Arg232 interacts directly with O2(T8) in CPD7. These observations suggest that the electrostatic contributions of the cationic side-chains of Arg232 and Arg350 are sufficient to stabilize the dimer radical anion. Interestingly, alanine substitution at Arg350 also demonstrated a 60% decrease in quantum yield, indicating that Arg350 plays a key role in stabilizing the dimer [34]. In fact, it seems that a tight (water mediated or direct) interaction between T8 and Arg232, or Arg350, is necessary to trigger the ring splitting as

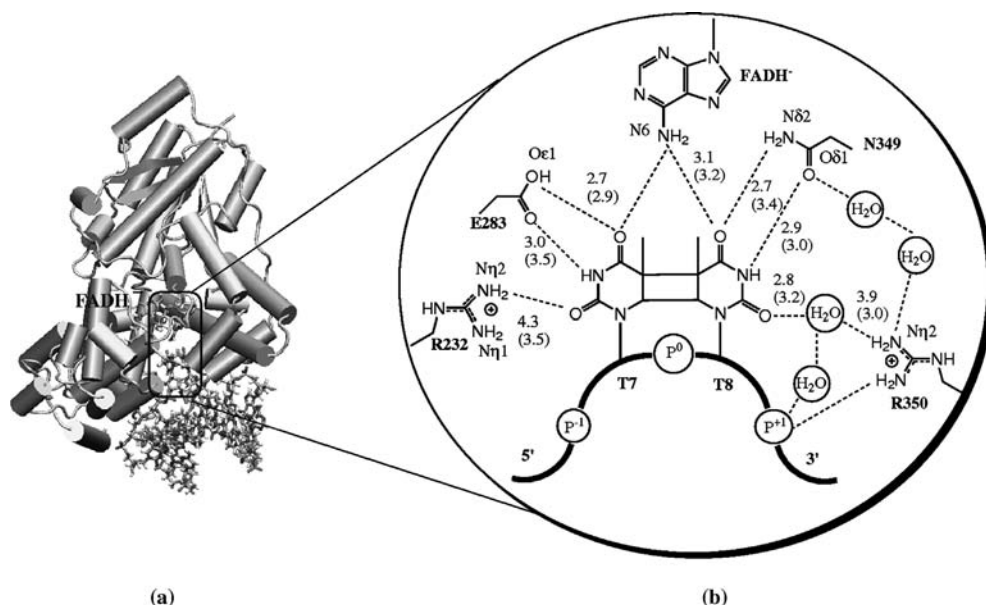


Figure 17.2 (a) DNA photolyase from *Anacystis nidulans* bound to double-stranded DNA with a CPD lesion (PDB code 1TEZ) [35]; (b) characteristic interaction distances (Å) between the cis-syn thymine dimer and the active site in the classical optimized structure. For comparison, interaction distances (Å) revealed by the X-ray crystallography are provided in brackets for the repaired thymine dinucleotide.

the electron is found to be localized on T8 when the cleavage of the C6–C6' bond occurs [31]. We observed that each O2 carbonyl group on T8 is tightly hydrogen-bonded either to water molecules or directly to the arginine side-chains. The only exception being CPD4, for which the O2 carbonyl group distances to water and arginine residues are above 3 Å most of the time during the simulations. This may explain why we could not observe the C6–C6' bond breaking in CPD4 within the QM/MM simulation time scale.

Asn349 and the flavin cofactor appear to mediate the repair reaction by anchoring the dimer through hydrogen bonds that are preserved during the entire process. The van der Waals interactions between the conserved tryptophans (W286 and W392) and the thymine dinucleotide are maintained during the whole course of the reaction, suggesting that these π -stacking effects also contribute to the stabilization of the dimer radical anion.

In summary, the QM/MM calculations enabled us to describe the dynamics of the DNA photolyase catalyzed splitting reaction, and to identify the bond-breaking process as an ultrafast reaction. The picture provided shed some light on apparent experimental discrepancies and offers a uniform and alternative interpretation of these data.

17.3.2

Reaction Mechanism of Endonuclease IV

Apurinic and apyrimidinic (AP) sites are the most frequent DNA lesions occurring *in vivo* [36], since they can result both from the natural loss of DNA base pairs and from the action of DNA glycosylases during the base excision repair pathway [37]. AP sites, when left unrepaired, can promote mutagenesis and result in substitution or frameshift mutations [38, 39]. For this reason, an important class of enzymes, named AP endonucleases, exists to catalyze the incision of DNA at AP sites, preparing the DNA for subsequent synthesis and ligation. In the bacteria *Escherichia coli*, Exonuclease III (Exo III), accounts for approx 90% of the AP-endonuclease activity, while Endonuclease IV (Endo IV) normally contributes to <10% [40]. Most AP endonucleases use Mg^{2+} ions to bind to DNA and catalyze phosphodiester cleavage at AP sites. However, Endo IV uses a trinuclear Zn center to catalyze DNA hydrolysis. Both Zn-based and Mg-based endonuclease families are thought to represent an example of convergent evolution, in which two enzymes have evolved independently to catalyze a similar reaction. Because Endo IV is undetected in mammalian cells but is present in pathogens, including *Mycobacterium tuberculosis*, *Candida albicans* and *Plasmodium falciparum*, it is an attractive target for antibacterial, antifungal and antimalarial agents [41–43].

A structure-based mechanistic hypothesis has been proposed for Endo IV based on the available experimental information, which includes several X-ray structures in the presence, and in absence, of a DNA substrate [44]. In the proposed reaction mechanism, a water molecule in the enzyme active site is activated into a hydroxide ion, and used to carry out the phosphodiester hydrolysis. In principle, both associative and dissociative mechanisms are possible for DNA hydrolysis. However, the reaction mechanism of Endo IV occurs through a synchronous bimolecular ($\text{A}_\text{N}\text{D}_\text{N}$) mechanism, in which a pentacoordinated phosphorus transition state is created (Figure 17.3). Recently, Ivanov *et al.* [45] have carried out a QM/MM MD study of the reaction, which confirmed the existence of a synchronous bimolecular ($\text{A}_\text{N}\text{D}_\text{N}$) mechanism and shed light on some aspects of DNA hydrolysis at AP sites. A similar computational study has been carried out in our group that is qualitatively, and quantitatively, in excellent agreement with the previous experimental and theoretical works. A computational approach is very attractive for Endo IV, since the transition state of the reaction cannot be isolated experimentally and, therefore, the detailed role of the trinuclear Zn^{2+} center during catalysis is difficult to assess.

The minimum energy pathway from the reactants to the product was obtained by computing the PMF along the reaction coordinate ϵ , which is defined as: $\epsilon = \text{P-OP}' - \text{O}^*-\text{P}$, where O^*-P and $\text{P-OP}'$ are the distances from the electrophilic P atom, O^* is the oxygen atom of the nucleophilic hydroxide and OP' is the displaced oxygen atom of the substrate (Figure 17.4a). The reaction is found to proceed through a synchronous bimolecular ($\text{A}_\text{N}\text{D}_\text{N}$) mechanism. At the transition state, the pentacoordinated phosphorus is in a trigonal bipyramidal geometry. A free energy barrier of $20.4 \text{ kcal mol}^{-1}$ ($20.6 \text{ kcal mol}^{-1}$ [45]) was computed at the BLYP level of theory,

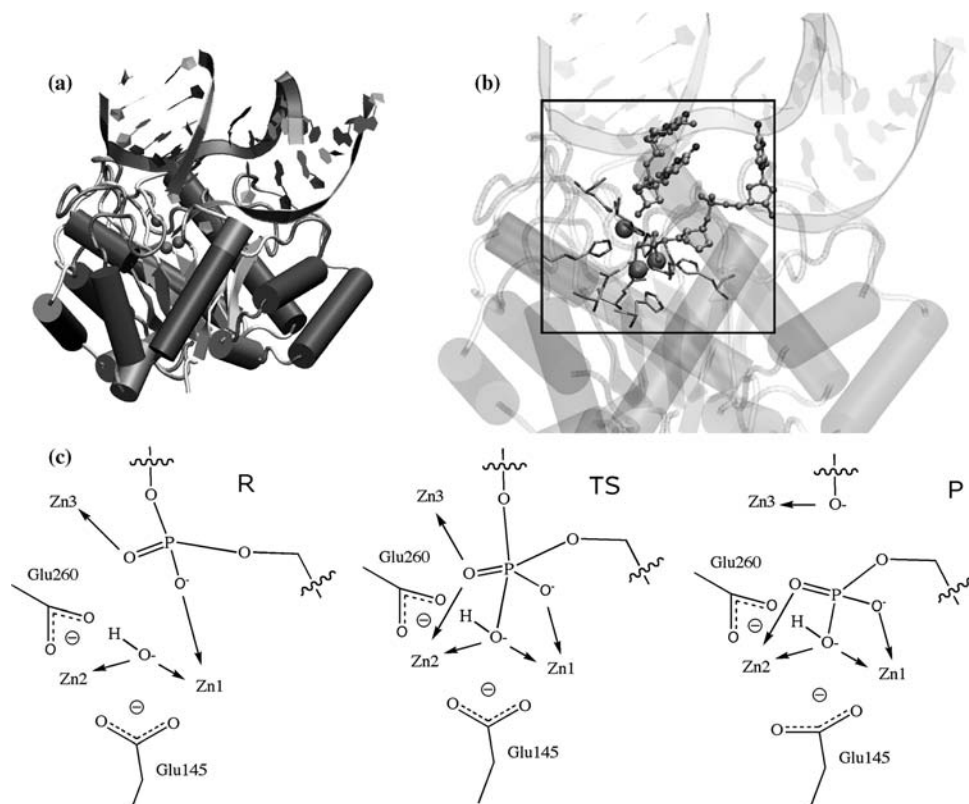


Figure 17.3 (a) Endo IV bound to double-stranded DNA with AP lesion (PDB code 1QUM); (b) active site residues and approximate position of the QM box; (c) proposed reaction mechanism for DNA hydrolysis by Endo IV: the reactive pathway connecting the reactants (R) and the cleaved DNA products (P) involves a pentacoordinated phosphorus atom at the transition state (TS) [44].

using thermodynamic integration. Experimental k_{cat} values for Endo IV-mediated phosphodiester hydrolysis are in the range $5.6\text{--}18\text{ min}^{-1}$ [46], which translates into $\Delta G \approx 19\text{ kcal mol}^{-1}$. The deviation of $<1.5\text{ kcal mol}^{-1}$ between the computational and experimental estimates is well within the range of error expected from DFT and the use of the BLYP functional.

Inspection of the trajectories shows that the transition state structure is stabilized by all three Zn^{2+} ions. This can be achieved by a contraction in the Zn2–Zn3 distance, from 4.7 to 4.2 Å, occurring at the transition state. After the reaction is completed, the Zn2–Zn3 distance increases again, which may promote the complete separation of the products and prevent re-crossing of the barrier. The first-shell aspartate and glutamate, apart from serving as a scaffold for the catalytic metal centers, accommodate the changes in substrate coordination by minimal compen-

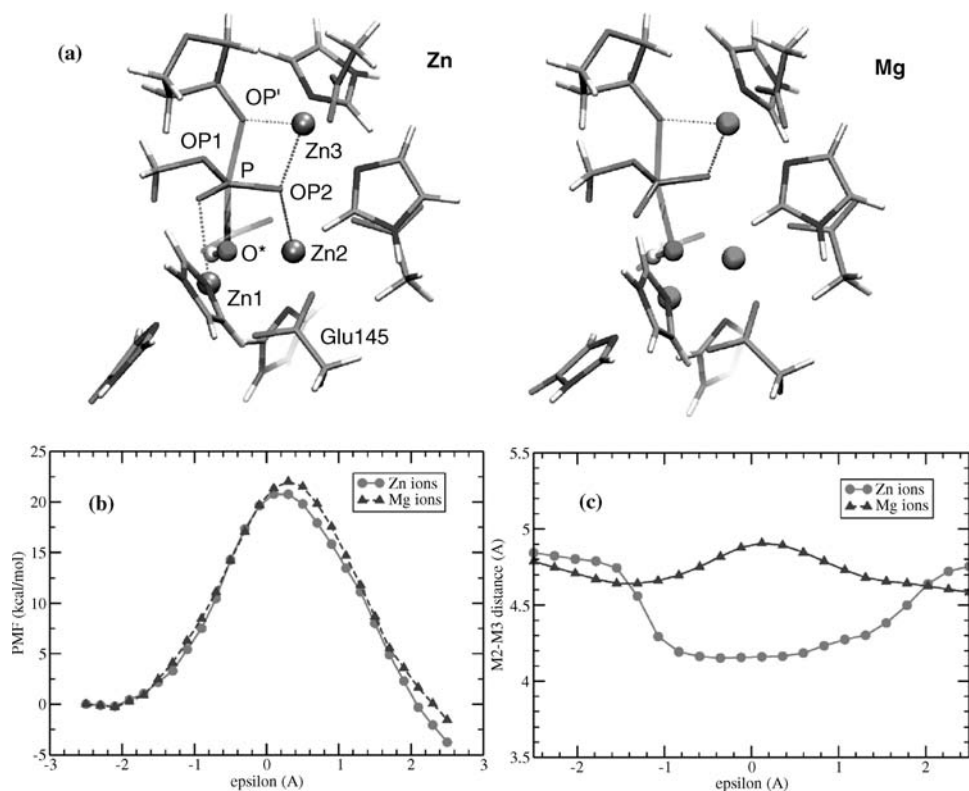


Figure 17.4 (a) Transition state of the reaction; (b) free energy profile; (c) distance between metals 2 and 3 during the reaction.

satory moves. In particular, Glu145 shifts from a bidentate (μ -1,3 bridging) to a monodentate coordination mode.

A natural question arising in the case of Endo IV is the particular choice of Zn^{2+} for the metal center. To a large extent, magnesium is the metal ion found in AP endonucleases, mainly as a result of its high natural abundance and availability of appropriate hydration states, ligand exchange rates, redox inertness and high charge density [47]. The choice of Zn^{2+} in Endo IV suggests that other factors such as the metal cluster flexibility also influence the catalytic efficiency.

The importance of the Zn flexibility for rapid catalytic turnover was investigated here by applying an external constraint on the Zn2–Zn3 distance, to keep the Zn2–Zn3 distance fixed at 4.7 Å during the simulations. In that case, the penta-coordinated phosphorus transition state could no longer be stabilized by all three Zn atoms. The energy required to carry out DNA hydrolysis along this reactive pathway could be estimated to be at least 10 kcal mol^{−1} higher. This suggests that the flexibility of the Zn center plays a crucial role in the stabilization of the transition state in the ($A_N D_N$) mechanism.

A computational experiment was performed to compare the catalytic efficiency of Endo IV when the trinuclear Zn^{2+} center is substituted by a trinuclear Mg^{2+} center. Experiments suggest that Endo IV has no Mg^{2+} requirement [40], but that it can be activated by Mg^{2+} in the absence of Zn^{2+} [48]. The energy barrier for the reaction was computed and found to be similar in the case of a trinuclear Mg^{2+} cluster, $<22 \text{ kcal mol}^{-1}$ (Figure 17.4b). Interestingly, however, inspection of the trajectories reveals that the transition state of the reaction is not stabilized by all three Mg^{2+} ions (Figure 17.4a and c). Therefore, Mg^{2+} ions appear to offer no advantage over Zn^{2+} ions in Endo IV. In addition, in Endo IV the product release has been proposed as the rate-determining step [49, 50]. In that case, a Mg^{2+} center may lead to a slower release of the products.

The role of the active site residue Glu261 was also investigated in the simulations. In a recent functional study, mutation of Glu261 into the neutral glutamine residue was shown to compromise the catalytic activity of the enzyme, while conserving the ability of the Glu-261-Gln mutant to favorably bind to a damaged DNA substrate [49]. It suggests that Glu261 assists catalysis by positioning the hydroxide ion in a near attack conformation prior to the reaction. To support this mechanistic hypothesis and help rationalize the lack of catalytic activity of the Glu-261-Gln mutant, the charge on Glu261 was neutralized in the computational model by substituting one of the Glu261 oxygens with a hydrogen. Indeed, the simulations showed that the favorable orientation of the hydroxide is lost in the mutant. The reaction free energy profile was computed for the mutant and found to be at least 10 kcal mol^{-1} higher in energy. This result is consistent with the Glu-261-Gln X-ray structure, which indicates that the catalytic efficiency is lost in the mutant without any apparent structural changes in the active site [49]. Glu261 is also found to play the role of a proton acceptor at the end of the reaction. In three out of six QM/MM MD simulations, a proton transfer was observed from the phosphate group to the nearby Glu261 residue.

In summary, the simulations could shed light on the different biophysical principles that are used to catalyze DNA hydrolysis at AP sites. In particular, a key element of phosphodiester cleavage catalysis is the efficient charge neutralization at the transition state, which is assisted by the metal ions. In water, the activation energy for an uncatalyzed DNA hydrolysis would be as high as $<38\text{--}40 \text{ kcal mol}^{-1}$ [51, 52], because both reactants are negatively charged. To perform hydrolysis within minutes, Endo IV needs to display rate enhancements of 10^{17} -fold. This is achieved in the enzyme by creating an altered electrostatic environment that neutralizes effectively the negatively charged reactants. The simulations reveal that the flexibility of Zn coordination is an advantage of Zn-based AP endonucleases that is not shared by Mg-based AP endonucleases. The flexibility is used in Endo IV to stabilize the penta-coordinated phosphorus transition state with all three Zn^{2+} atoms. In the case of Mg^{2+} , only two metal ions can participate in the stabilization of the transition state. Interestingly, a similar explanation has been used to account for the different behavior of magnesium (a co-catalyst) and calcium (an inhibitor) in the BamHI restriction enzyme [53].

Apart from an electrostatic stabilization of the transition state, another important environmental change in the enzyme is the creation of basic conditions. This role is

played by Glu261, which has a twofold contribution: first it orients the hydroxide nucleophile in an optimal geometry for the attack and second it can accept a H^+ from the phosphate group after the nucleophilic attack is completed.

QM/MM simulations of Endo IV illustrate how computer simulations can help highlight the relationships between the structures of di- and multinuclear metal complexes and their functions. In particular, a better understanding of the cooperativities between metal and ligands, and between metal sites in the course of DNA hydrolysis, can inspire the design of di- and multinuclear metal-based artificial restriction enzymes. These systems are likely to be of increasing importance in biotechnology and medicine.

17.3.3

Role of Water in the Catalysis Mechanism of DNA Repair Enzyme, MutY

Among the 20 different types of nucleobase damage due to oxidative stress, one of the most common and most stable is oxidation of guanine (G) to 7,8-dihydro-8-oxoguanine (8-oxoG) [4]. 8-oxoG pairs preferentially with adenine (A) over cytosine (C), resulting in a high rate of C to A transversions during replication of the DNA strand containing the 8oxoG lesion [54, 55]. The 8-oxoG lesion and resulting mispairs are handled by a suite of interacting proteins described as the GO system, originally characterized in *Escherichia coli* [56].

The GO system protein MutY removes the A nucleobase from 8-oxoG: A mispairs present in the DNA double-helix [56]. MutY is a monofunctional base excision repair enzyme, one among a class of DNA glycosylases that hydrolyze N glycosidic bonds, thereby generating a (normal or damaged) free nucleobase plus a DNA abasic (AP) site [57]. MutY was first described in *Escherichia coli* (eMutY) [56], but homologs have been documented in many organisms, such as *Bacillus stearothermophilus* (bMutY), yeast *Schizosaccharomyces pombe* (SpMYH), *Salmonella typhimurium*, *Deinococcus radiodurans*, mouse (mMYH), rat, calf thymus and human (hMYH) cells [58–64].

Numerous studies have elucidated structural features that play an important role in MutY catalysis. These studies indicate that MutY is composed of two sub-domains: the C-terminal domain, which appears responsible for recognition of the oxoG lesion [58, 65–68], and the N-terminal domain, which contains the catalytic machinery responsible for excision of A from the DNA duplex [65, 66, 69, 70]. In studies of eMutY, two residues of the N-terminal domain are found to be catalytically essential: Glu37 [71, 72] and Asp138 [72, 73]. Notably, both of these residues are conserved in the MutY homologs of *B. stearothermophilus* (corresponding to Glu43 and Asp144 in bMutY), *S. pombe* and humans [58]. A common mutation of Tyr165 in hMYH (Y165C) has been linked to human colorectal cancers. In studies of mMYH and eMutY, the corresponding mutations (Y150C and Y82C) resulted in decreased substrate binding affinities and decreased rates of the catalytic A excision step [74]. In the bMutY crystal structure, the corresponding residue (Tyr88) intercalates into the DNA duplex in between the damaged oxoG and the nucleobase 5' to oxoG [58]. The results of these studies suggest that Tyr88 has a role in both oxoG recognition and A excision catalysis [74].

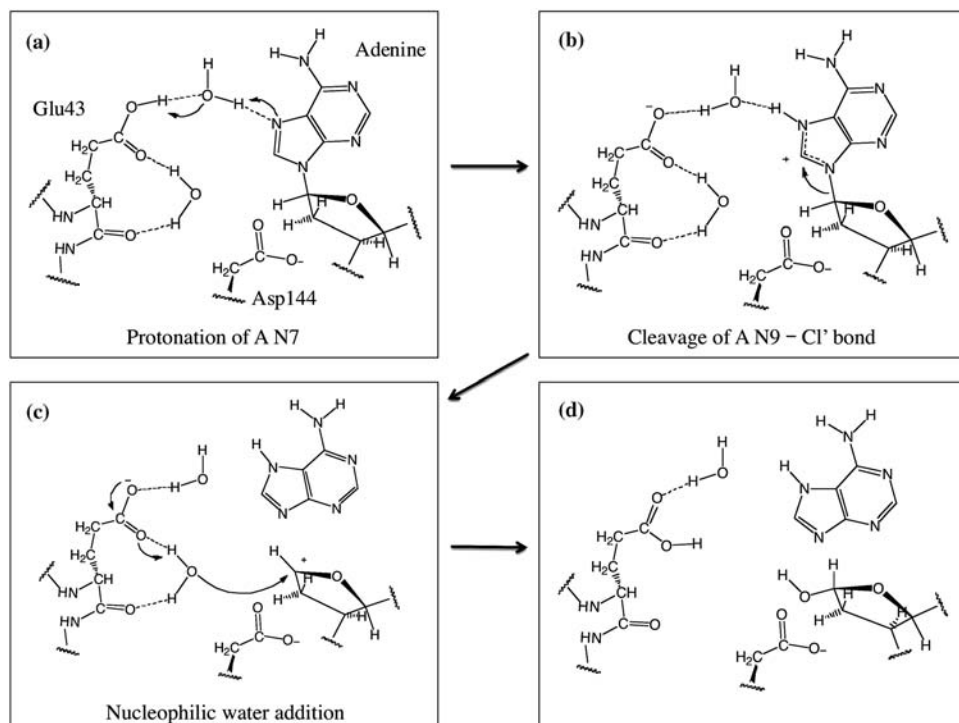


Figure 17.5 Previously proposed reaction steps of the MutY catalysis mechanism.

A reaction mechanism has been proposed for MutY with the following steps. First, a nearby acidic group, probably Glu43 (in bMutY), protonates A N7 (Figure 17.5a), and the protonated A cleaves from the glycosyl ring at the N9–C1' bond, generating a discrete oxocarbenium intermediate (Figure 17.5b). Then, a (possibly activated) water nucleophile attacks the glycosyl ring at C1', and proton exchanges complete the N-glycosylide hydrolysis (Figure 17.5c and d). MutY and/or solvent interactions with A N3 and A N1 are believed to help enable departure of the leaving group [57], but it is unclear how this occurs.

Water plays numerous proposed roles in the MutY catalysis mechanism. The published crystal structure of bMutY Asp144Asn*(GO: A) [58] reveals a structured water situated within van der Waals contact of both the Glu43 carboxylic oxygen and the glycosyl C1' carbon (Figure 17.6a). The crystal structure of the eMutY_Lys20Ala*A catalytic domain [75] exhibits a similarly positioned structured water. This water is a presumed candidate for nucleophilic attack of the oxocarbenium intermediate [58, 76], but it is unclear whether nearby residues (e.g., Glu43 or Asp144) activate or otherwise mediate the nucleophilic attack. In the bMutY_Asp144Asn*(GO: A) crystal structure, a structured water bridges Glu43 with A N7 [58] (Figure 17.6a). Consistent with solvent deuterium kinetic isotope effect (KIE) evidence [76], this water is proposed to relay a proton between Glu43 and

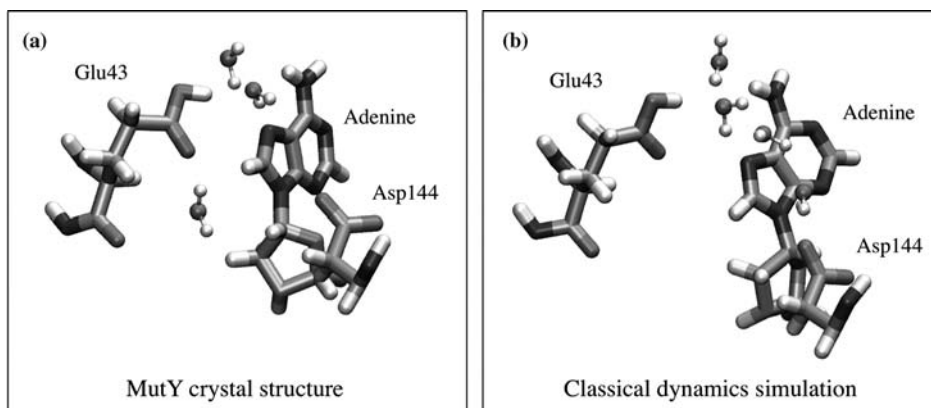


Figure 17.6 Structured waters in the catalytic region: (a) according to the bMutY crystal structure [58]; (b) according to classical molecular dynamics simulations (snapshot taken at ~ 13 ns, 330 K).

A N7, thereby activating the A leaving group [57, 58, 76]. In addition, solvent and/or MutY interactions with A N1, A N3 and A N6 may be important, but these mechanistic details remain uncertain.

Here a simulation study was designed to address open questions that relate to: (i) the detailed catalytic mechanism, (ii) the catalytic role of water and (iii) the catalytic role of residues proximate to the active site, especially Asp144.

We conducted the following simulations, using MutY crystal structure data to set the initial coordinates of the system. Fromme *et al.*'s [58] crystal structure of bMutY bound to a 10-mer of duplex DNA containing an oxoG: A lesion site (Protein Data Bank accession code 1RRQ) was used to generate the initial atomic coordinates. After a restrained annealing to warm the system, unrestrained classical dynamics simulations of the precatalytic bMutY structure were performed in explicit TIP3P water and using the AMBER parm99 force field. Drawing upon equilibrated structures generated by classical simulations, we used QM/MM to conduct a free energy calculation of the proposed reaction mechanism, applying thermodynamic integration. The BLYP [77, 78] density functional method was used together with Troullier–Martins pseudo-potentials [79] and dispersion-corrected atom-centered potentials [17] to treat the catalytic region. Based on ProPka [80] predictions, Poisson–Boltzmann calculations and previous experimental evidence [58, 76], we assigned residue Glu43 as protonated. All other residues were assigned the default protonation states given in the AMBER force field library.

Classical molecular dynamics simulations conveyed new information about the role of structured waters in the active site region, including some aspects that the crystal structure experiment was unable to diagnose. Consistent with the crystal structure, classical simulations predicted a stable structure of two waters bridging Glu43, A N7 and A N6. In 10-ns simulations at three different temperatures (300, 330 and 350 K) two waters maintained this structure reliably and reversibly, displaying

occasional disassembly, exchange with the solvent and spontaneous reformation. This constitutes an important corroboration of the stable water structures that were observed in the crystal structure, because the crystal X-ray measurement had been conducted in liquid nitrogen and, thus, at 77 K, far from biological relevance. Conversely, molecular dynamics simulations predicted a disordered state for the lone water, which according to the crystal structure is bound between Glu43 and the catalytic target glycosyl. This water is currently proposed as the nucleophilic water involved in the glycoside hydrolysis step [58, 76]. However, in all three simulations (300, 330 and 350 K) of 10 ns each, this water departs into the solvent within 1–2 ns, and no water molecule subsequently returns to the Glu43 water bridging position observed in the crystal structure. It is unclear whether the disparity between the simulation prediction and the crystal structure data results from failure of the classical force field, insufficient sampling of simulation dynamics or an experimental artifact of the crystal preparation or measurement. The outcome clearly has ramifications for interpretations of the reaction mechanism, as discussed further below. According to our molecular dynamics simulations, Asp144 significantly influences the conformations of structured water in the catalytic region. However, the existing published bMutY crystal structure is unable to provide much insight into this possible role for Asp144. The bMutY crystal structure was prepared as an Asp144Asn mutant, precisely so that the resulting protein–DNA complex would be catalytically inactive and therefore crystallizable. Molecular dynamics simulations predict a spontaneously forming and stable three-water bridge connecting Asp144 to Glu43 and A N7 (Figure 17.6b). In fact, this was the most favored structured water conformation that we detected in the entire catalytic region, exhibiting several events of reversible assembly and disassembly in all three of the 10-ns classical molecular dynamics simulations (300, 330 and 350 K). This raises new hypotheses for the possible role of Asp144. Through bridging waters, this residue could stabilize the water-mediated proton exchange between Glu43 and A N7. Perhaps more importantly, Asp144 and Glu43 could jointly tether the structured, activated nucleophilic water that attacks the oxacarbenium. This could lend insight into the crucial catalytic function of Asp144. Although Asp144 is catalytically required, currently published evidence leaves its role ambiguous. Taken together, data from molecular dynamics simulations suggest that structured waters may participate in catalysis differently than what has been inferred from the bMutY crystal structure. The crystal structure data reflect the limitations of non-biological temperature (77 K), the absence of a catalytically required residue, Asp144, and only narrow insight into system dynamics. Molecular dynamics simulations allow us to address all three of these information gaps. The simulations corroborate the experimental evidence that structured waters bridge Glu43 to A N7 and A N6, thereby facilitating the Glu43 A N7 proton exchange during catalysis. Additionally, classical molecular dynamics simulations suggest the new hypotheses that Asp144 may participate in positioning and/or stabilizing the waters necessary for both the A N7 protonation step and the nucleophilic attack on the oxacarbenium ion.

In currently ongoing work, we hope to address some of these newly developed hypotheses. Preliminary results from QM/MM simulations are consistent with

existing experimental data and suggest new insights into the MutY catalysis mechanism. Since experimental evidence supports an S_N1 mechanism for the cleavage of A N9 from C1', we assumed that the A N9–C1' distance was the principal reaction coordinate for the first reaction step. We used thermodynamic integration along this constrained reaction coordinate to systematically explore the system's behavior with varying N9–C1' distance. Remarkably, elongation of the N9–C1' bond induced spontaneous deprotonation of Glu43 and protonation of A N7, via a proton exchange through the bridging water. At a N9–C1' bond length constrained near 1.95 Å, this proton exchange reaction transpired and reversed spontaneously several times. The calculated free energy profile derived from constraint force statistics shows that the reversible proton hop occurs at a shorter N9–C1' bond length than the transition state of the S_N1 A cleavage step. Hence, the proton hop would usually occur before the system reaches a transition structure for the S_N1 reaction. The calculated free energy barrier for the S_N1 reaction was about 10 kcal mol⁻¹. However, it is difficult to judge the quality of this prediction against experimentally observed rate constants for the catalysis reaction until simulations of the subsequent reaction step (nucleophilic attack) have been taken into account since, according to experimental KIE evidence [78], the rate-determining transition state of the glycolyside hydrolysis reaction occurs during the nucleophilic water attack step. Simulations of the nucleophilic reaction step, and associated proton exchanges, are currently underway. In upcoming calculations, we hope to further distinguish the function of Asp144. Based on the current and previous work, our working hypotheses are that Asp144 may stabilize the oxacarbenium intermediate, it may mediate the protonation of A N7 as well as mediate the nucleophilic water attack of the oxacarbenium at C1', or it may participate in multiple roles.

17.4

Conclusions

We have discussed the application of QM/MM simulation techniques to study problems in enzymology, and improve our understanding of DNA repair mechanisms. Currently, the main limitations of the method are: (i) the accuracy of the DFT functional, (ii) the convergence of sampling properties and (iii) the match between the computational model and the real system. Although, these limitations often become important for complex systems, such as DNA repair enzymes, QM/MM studies can already provide valuable additional mechanistic insights that could not have been obtained by experiments alone.

Acknowledgment

F.M. would like to thank Teodoro Laino and Professor Jürg Hutter for fruitful discussion. J.S.A. gratefully acknowledges U.S. National Science Foundation MPS-DRF Award 0502600 and the Swiss National Supercomputing Centre (CSCS) for their support.

References

- 1 Lindahl, T. (1993) *Nature*, **362**, 709.
- 2 Poulsen, H.E. and Loft, S. (1998) *Acta Biochim. Pol.*, **45**, 133.
- 3 Slupphaug, G., Markussen, F.H., Olsen, L.C., Aasland, R., Aarsaether, N., Bakke, O., Krokan, H.E., and Helland, D.E. (1993) *Nucleic Acids Res.*, **21**, 2579.
- 4 Slupphaug, G., Kavli, B., and Krokan, H.E. (2003) *Mutat. Res.*, **531**, 231.
- 5 Hakem, R. (2008) *EMBO J.*, **27**, 589.
- 6 Sheppard, T.L., Ordoukhanian, P., and Joyce, G.F. (2000) *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 7802.
- 7 Wiest, O., Harrison, C.B., Saetell, N.J., Cibulka, R., Sax, M., and Konig, B. (2004) *J. Org. Chem.*, **69**, 8183.
- 8 Bentle, M.S., Bey, E.A., Dong, Y., Reinicke, K.E., and Boothman, D.A. (2006) *J. Mol. Histol.*, **37**, 203.
- 9 Lieberman, H.B. (2008) *Curr. Med. Chem.*, **15**, 360.
- 10 Kelley, M.R. and Fishel, M.L. (2008) *Anticancer Agents Med. Chem.*, **8**, 417.
- 11 Altieri, F., Grillo, C., Maceroni, M., and Chichiarelli, S. (2008) *Antioxid. Redox Signal.*, **10**, 891.
- 12 Hyun, M., Lee, J., Lee, K., May, A., Bohr, V.A., and Ahn, B. (2008) *Nucl. Acids Res.*, **36**, 1380.
- 13 Rothlisberger, U. and Carloni, P. (eds) (2006) *Computer Simulations in Condensed Matter Systems*, Springer.
- 14 Warshel, A. and Levitt, M. (1976) *J. Mol. Biol.*, **103**, 227.
- 15 IBM (1990) CPMD Copyright IBM Corp 1990-2006. Copyright MPI Festkörperforschung Stuttgart (1997–2001).
- 16 Laio, A., Van de Vondelle, J., and Rothlisberger, U. (2002) *J. Chem. Phys.*, **116**, 6941–6947.
- 17 Von Lilienfeld, O.A., Tavernelli, I., Sebastiani, D., and Rothlisberger, U. (2005) *J. Chem. Phys.*, **122**, 014113.
- 18 Kohn, W. and Sham, L. (1965) *J. Phys. Rev.*, **140**, A1133–A1138.
- 19 Car, R. and Parrinello, M. (1985) *Phys. Rev. Lett.*, **55**, 2471–2474.
- 20 Zhao, Y. and Truhlar, D.G. (2008) *J. Chem. Theory Comput.*, **4**, 1849.
- 21 Carter, E.A., Ciccotti, G., Hynes, J.T., and Kapral, R. (1989) *Chem. Phys. Lett.*, **156**, 472.
- 22 Trzesniak, D., Kunz, A.P.E., and van Gunsteren, W.F. (2007) *ChemPhysChem*, **8**, 162.
- 23 Schiffrl, S.K. and Wallace, D.C. (1985) *J. Chem. Phys.*, **83**, 5203.
- 24 Cadet, J. and Vigny, P. (1990) *The Photochemistry of Nucleic Acids in Bioorganic Photochemistry: Photochemistry and the Nucleic Acids*, John Wiley & Sons, Inc., New York.
- 25 Stege, H., Roza, L., Vink, A.A., Grewe, M., Ruzicka, T., Grether-Beck, S., and Krutmann, J. (2000) *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 1790.
- 26 Sancar, A. (2003) *Chem. Rev.*, **103**, 2203.
- 27 Okamura, T., Sancar, A., Heelis, P.F., Begley, T.P., Hirata, Y., and Mataga, N. (1991) *J. Am. Chem. Soc.*, **113**, 3143.
- 28 Langenbacher, T., Zhao, X., Bieser, G., Heelis, P.F., Sancar, A., and Michel-Beyerle, M.E. (1991) *J. Am. Chem. Soc.*, **119**, 10532.
- 29 MacFarlane, A.W. and Stanley, R.J. (2003) *Biochem.*, **42**, 8558.
- 30 Kao, Y.T., Saxena, C., Wang, L., Sancar, A., and Zhong, D. (2005) *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 16128.
- 31 Masson, F., Laino, T., Rothlisberger, U., and Hutter, J. (2009) *ChemPhysChem*, **20**, 400.
- 32 Laio, A. and Parinello, M. (2002) *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 12562.
- 33 Essen, L.O. and Klar, T. (2006) *Cell. Mol. Life Sci.*, **63**, 1266.
- 34 Vande Berg, B.J. and Sancar, G.B. (1998) *J. Biol. Chem.*, **273**, 20276.
- 35 Mees, A., Klar, T., Gnau, P., Hennecke, U., Eker, A.P.M., Carell, T., and Essen, L.-O. (2004) *Science*, **306**, 1789.
- 36 Otterlei, M., Kavli, B., Standal, R., Skjellbred, C., Bharati, S., and Krokan, H.E. (2000) *EMBO J.*, **19**, 5542.
- 37 Friedberg, E.C., Walker, G.C., and Siede, W. (eds) (1995) *DNA repair and mutagenesis*, ASM Press, Washington, DC.
- 38 Loeb, L.A. and Preston, B.D. (1986) *Annu. Rev. Genet.*, **20**, 201.
- 39 Woodgate, R. and Levine, A.S. (1996) *Cancer Surv.*, **28**, 117.
- 40 Demple, B., Johnson, A., and Fung, D. (1998) *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 7731.

- 41 Hosfield, D.J., Guan, Y., Haas, B.J., Cunningham, R.P., and Tainer, J.A. (1999) *Cell*, **98**, 397.
- 42 Tsuji, A., Kodaira, K., Inoue, M., and Yasukawa, M. (2001) *Mutat. Res. DNA Repair*, **486**, 53.
- 43 Haltiwanger, B.M., Matsumoto, Y., Nicolas, E., Dianov, G.L., Bohr, V.A., and Taraschi, T.F. (2000) *Biochemistry*, **39**, 763.
- 44 Mol, C.D., Hosfield, D.J., and Tainer, J.A. (2000) *Mutat. Res.*, **460**, 211.
- 45 Ivanov, I., Tainer, J.A., and McCammon, J.A. (2007) *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 1465.
- 46 Takeuchi, M., Lillis, R., Demple, B., and Takeshita, M. (1994) *J. Biol. Chem.*, **269**, 21907.
- 47 Liua, C., Wangb, M., Zhanga, T., and Sunc, H. (2004) *Coord. Chem. Rev.*, **248**, 147.
- 48 Liu, X. and Liu, J. (2005) *Bioch. Biophys. Acta – Proteins & Proteomics*, **1753**, 217.
- 49 Garcin, E.D., Hosfield, D.J., Desai, S.A., Haas, B.J., Bjoras, M., Cunningham, R.P., and Tainer, J.A. (2008) *Nat. Struct. Mol. Biol.*, **15**, 515.
- 50 Coleman, J.E. (1992) *Annu. Rev. Biophys. Biomol. Struct.*, **21**, 441.
- 51 Iche-Tarrat, N., Barthelat, J.C., Rinaldi, D., and Vignoux, A. (2005) *J. Phys. Chem. B*, **205**, 109.
- 52 Florian, J. and Warshel, A. (1997) *J. Am. Chem. Soc.*, **119**, 5473.
- 53 Mordasini, T., Curioni, A., and Andreoni, W. (2002) *J. Chem. Phys.*, **116**, 6941.
- 54 Shibutani, S., Takeshita, M., and Grollman, A.P. (1991) *Nature*, **349**, 431.
- 55 Grollman, A.P. and Moriya, M. (1993) *Trends Genet.*, **9**, 246.
- 56 Michaels, M.L., Cruz, C., Grollman, A.P., and Miller, J.H. (1992) *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 7022.
- 57 Berti, P.J. and McCann, J.A.B. (2006) *Chem. Rev.*, **106**, 506.
- 58 Fromme, J.C., Banerjee, A., Huang, S.J., and Verdine, G.L. (2004) *Nature*, **427**, 652.
- 59 Lu, A.-L., Cuipa, M.S.I., and Shanabruch, W.G. (1990) *J. Bacteriol.*, **172**, 1232.
- 60 Li, X. and Lu, A.-L. (2001) *J. Bacteriol.*, **183**, 6151.
- 61 Hayashi, H., Tominaga, Y., Hirano, S., McKenna, A.E., Nakabeppu, Y., and Matsumoto, Y. (2002) *Curr. Biol.*, **12**, 335.
- 62 Ma, H., Lee, H.M., and Englander, E.W. (2004) *Nucleic Acids Res.*, **32**, 4332.
- 63 McGoldrick, J.P., Yeh, Y.-C., Solomon, M., Essigmann, J.M., and Lu, A.-L. (1995) *Mol. Cell Biol.*, **15**, 989.
- 64 Wooden, S.H., Bassett, H.M., Wood, T.G., and McCullough, A.K. (2004) *Cancer Lett.*, **205**, 89.
- 65 Gogos, A., Cillo, J., Clarke, N.D., and Lu, A.-L. (1996) *Biochemistry*, **35**, 16665.
- 66 Noll, D.M., Gogos, A., Granek, J.A., and Clarke, N.D. (1999) *Biochemistry*, **38**, 6374.
- 67 Li, X., Wright, P.M., and Lu, A.-L. (2000) *J. Biol. Chem.*, **275**, 8448.
- 68 Chmiel, N.H., Golinelli, M.P., Francis, A.W., and David, S.S. (2001) *Nucleic Acids Res.*, **29**, 553.
- 69 Manuel, R.C., Czerwinski, E.W., and Lloyd, R.S. (1996) *J. Biol. Chem.*, **271**, 16218.
- 70 Manuel, R.C. and Lloyd, R.S. (1997) *Biochemistry*, **36**, 11140.
- 71 Porello, S.L., Williams, S.D., Kuhn, H., Michaels, M.L., and David, S.S. (1996) *J. Am. Chem. Soc.*, **118**, 10684.
- 72 Guan, Y., Manuel, R.C., Arvai, A.S., Parikh, S.S., Mol, C.D., Miller, J.H., Lloyd, R.S., and Tainer, J.A. (1998) *Nat. Struct. Biol.*, **5**, 1058.
- 73 Wright, P.M., Yu, J.A., Cillo, J., and Lu, A.-L. (1999) *J. Biol. Chem.*, **274**, 29011.
- 74 Pope, M.A., Chmiel, N.H., and David, S.S. (2005) *DNA Rep.*, **4**, 315.
- 75 Manuel, R.C., Hitomi, K., Arvai, A.S., House, P.G., Kurtz, A.J., Dodson, M.L., McCollough, A.K., Tainer, J.A., and Lloyd, R.S. (2004) *J. Biol. Chem.*, **279**, 46930.
- 76 McCann, J.A.B. and Berti, P.J. (2008) *J. Am. Chem. Soc.*, **130**, 5789.
- 77 Lee, C.T., Yang, W.T., and Parr, R.G. (1988) *Phys. Rev. B*, **37**, 785–789.
- 78 Becke, A.D. (1998) *Phys. Rev. A*, **38**, 3098–3100.
- 79 Troullier, N.J. and Martins, L. (1991) *Phys. Rev. B*, **43**, 8861–8869.
- 80 Li, H., Robertson, A.D., and Jensen, J.H. (2005) *Proteins*, **61**, 704.

18

Computational Electronic Structure of Spin-Coupled Diiron-Oxo Proteins

Jorge H. Rodriguez

18.1

Introduction

Oxo- and hydroxo-bridged diiron centers are ubiquitous in biochemistry. The importance of these structural motifs is illustrated by the name given to an entire class of metalloproteins, namely the diiron-oxo proteins [1–3]. Figure 18.1 illustrates the structure of the metallic cores of these proteins, which display oxo- and hydroxo-bis(acetato)-bridged diiron centers.

Interest in diiron-oxo proteins has grown as some crystallographic structures have been solved and new members of the group have been identified [1–3]. Despite the structural similarities of their active sites, these proteins have various roles. For example, hemerythrin (Hr) [4, 5] is involved in dioxygen transport, uteroferrin (Uf) [6] catalyzes the hydrolysis of phosphate esters, and methane monooxygenase (MMO) [3, 7] catalyzes the hydroxylation of methane. Most diiron-oxo proteins react with O₂, binding it reversibly in the case of Hr [4, 5] or activating it in the case of MMO [7], ribonucleotide reductase [8] and stearoyl-CoA desaturase [9]. In addition, several binuclear transition metal complexes that model the active site structure and spectroscopic properties of diiron proteins have been synthesized [10, 11]. A general characteristic of these complexes is the presence of two or more paramagnetic spin centers that interact with each other and couple (anti)ferromagnetically.

Geometric structures, many physicochemical properties and biochemical reactivities of the protein diiron centers are intimately related to their valence-shell electron–electron interactions. In particular, the anti- or ferromagnetic order of the diiron centers is largely due to superexchange interactions between unpaired electrons of partially filled iron d-shells [12].

The accurate quantitative and qualitative description of open-shell electron–electron interactions requires the inclusion of subtle exchange and correlation effects that, to a good approximation, can be accounted for by spin density functional theory (SDFT) or wavefunction-based correlated *ab initio* methods. To gain microscopic insight about the specific physicochemical properties and biochemical functions

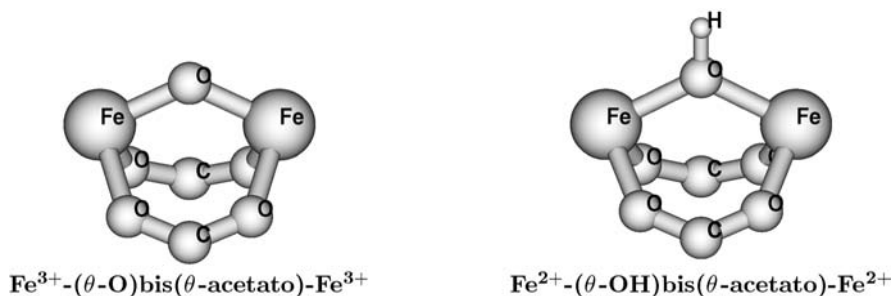


Figure 18.1 Structure of the metallic cores of diiron-oxo proteins.

of diiron proteins it is important to carry out electronic structure calculations on their metallic centers and also on model compounds that closely resemble their binuclear cores. In particular, one can apply spin density functional theory to investigate the electronic environment of diiron centers and also to model their interactions with molecular oxygen and substrates. In addition, spectroscopic probes such as ^{57}Fe Mössbauer, EPR and magnetic susceptibility provide a wealth of experimental information about the physicochemical properties of diiron-oxo proteins. Such experimental data is often analyzed in terms of phenomenological spin Hamiltonians, which complements the information obtained from electronic structure calculations based on SDFT [13]. In this brief chapter I describe how modern methods based on SDFT and phenomenological spin Hamiltonians have been used to investigate physicochemical properties of diiron centers in proteins and related model compounds.

18.2

(Anti)ferromagnetic Spin Coupling

One of the most interesting features of diiron-oxo proteins is the anti- or ferromagnetic spin coupling of their active centers. Such coupling is well established from Mössbauer [13, 14], susceptibility [15], EPR [16, 17], MCD [18] and other experiments. The theoretical investigation of exchange interactions responsible for spin coupling has been a subject of intense study [11, 12, 19] and the results obtained have greatly contributed to the understanding of diiron-oxo (anti)ferromagnetism. The possible roles and effects of spin coupling between the iron ions on biochemical reactivity are most likely significant but need to be further elucidated. In particular, the application of SDFT to study diiron-oxo enzymes shows great promise and can potentially provide significant insight about their catalytic cycles and nature of their reaction intermediates.

Ground state antiferromagnetism in binuclear complexes arises from interactions between pairs of unpaired electrons, each member of the pair nominally localized on a different metal ion. Although the phenomenon of antiparallel spin coupling resembles a magnetic interaction between two spins, its physical origin

is electrostatic in nature (Heisenberg [20]) and a consequence of the antisymmetry requirement of the Pauli exclusion principle [20]. Dirac [21, 22] showed that this interaction, although electrostatic in origin, could be written as a scalar product of the spin operators of two electrons. Such formulation was readily extended and applied by Van Vleck [23] to exchange interactions between atoms in molecules and solids. As such, the spin coupling between two metal ions with single-ion operators S_1 and S_2 can be represented and quantified by the Heisenberg–Dirac–Van Vleck Hamiltonian:

$$\mathcal{H}^{\text{HB}} = J \mathbf{S}_1 \cdot \mathbf{S}_2 \quad (18.1)$$

In general, the net spin coupling within a binuclear complex results from a complex admixture of antiferromagnetic and ferromagnetic interactions [24–26]. Often, the former are dominant and the spin ground state corresponds to the lowest (i.e., $S_{\text{min}} = S_1 - S_2$) eigenvalue of the total spin operator $\mathbf{S} = \mathbf{S}_1 + \mathbf{S}_2$. In this case, the J constant of Equation (18.1) is positive and the binuclear complex is said to have an antiferromagnetic ground state.

In oxo-bridged binuclear iron complexes the magnetic d orbitals (i.e., those hosting unpaired electrons) of the two metal sites can overlap via p orbitals of their nominally diamagnetic bridging ligands. As a result of such indirect overlap, several superexchange (i.e., bridge-mediated) pathways can be formed that favor an antiparallel alignment of the electron spins. Such antiferromagnetic pathways can be represented as idealized electronic configurations of the form $\text{Fe}(\text{d}_{\text{li}}^{\uparrow}) : \mu\text{-O}(\text{p}^{\uparrow}) : \text{Fe}2(\text{d}_{\text{2i}}^{\uparrow})$. Here, the atomic orbitals of the iron sites can have equal (i.e., $\text{d}_{\text{li}} = \text{d}_{\text{2i}}$) or mixed (i.e., $\text{d}_{\text{li}} \neq \text{d}_{\text{2i}}$) local symmetry [11, 27]. Depending on the oxidation and spin state of the cations involved [28], molecular symmetry and geometry [29, 30], there may be several pairs of interacting electrons or, equivalently, several antiferromagnetic pathways.

18.3

Spin Density Functional Theory of Antiferromagnetic Diiron Complexes

A useful step towards understanding the reactivity of diiron-oxo enzymes has been to elucidate the mechanisms leading to their strong or weak antiferromagnetism. As an illustration, Rodriguez and McCusker [12] used SDFT to study two model compounds originally synthesized by Armstrong, Lippard, *et al.* [31] and whose crystallographic structures and magnetic properties mimic those of diiron-oxo proteins (Figure 18.2).

Figure 18.2 displays the optimized geometries of the $(\mu\text{-O})\text{bis}(\mu\text{-acetato})$ -bridged and $(\mu\text{-OH})\text{bis}(\mu\text{-acetato})$ -bridged complexes whose ground state and magnetic properties have been studied in detail [12, 31]. Susceptibility measurements showed that the latter (protonated) complex displays a coupling constant ($J = 34 \text{ cm}^{-1}$) dramatically lower than that of the former ($J = 242 \text{ cm}^{-1}$). Therefore, it was of great interest to understand how two nearly identical complexes gave rise to dramatically different magnetic parameters. A series of meticulously planned SDFT calculations

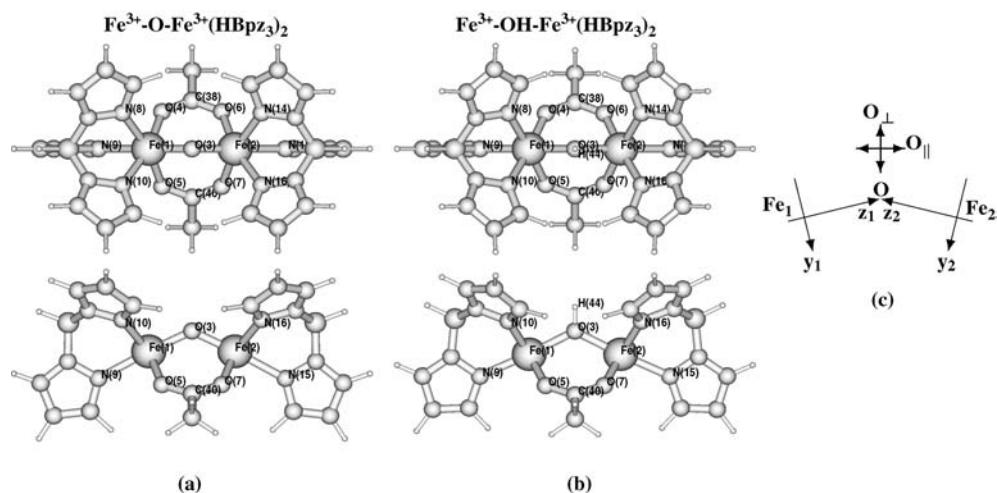


Figure 18.2 Binuclear antiferromagnets $\text{Fe}_2(\mu\text{-O})(\mu\text{-O}_2\text{CCH}_3)_2(\text{HBpz}_3)_2$ and $[\text{Fe}_2(\mu\text{-OH})(\mu\text{-O}_2\text{CCH}_3)_2(\text{HBpz}_3)_2]^+$: optimized structures at U-BPW91/6-31G* level in C_{2v} symmetry. Views are perpendicular (top) and parallel (bottom) to $\text{Fe}_1\text{-}\mu\text{O-Fe}_2$ or $\text{Fe}_1\text{-}\mu\text{O}^{\text{H}}\text{-Fe}_2$ planes. (a) Two $(\mu\text{-O})\text{bis}(\mu\text{-acetato})$ -bridged Fe^{3+} ($S = 5/2$) ions are spin-coupled via superexchange interactions (Figure 18.3) to a $S = S_1 - S_2 = 0$ ground state; the experimental coupling constant that parameterizes the Heisenberg Hamiltonian $\mathcal{H}^{\text{HB}} = JS_1 \cdot S_2$ is $J = 242 \text{ cm}^{-1}$ [31]; (b) the $(\mu\text{-OH})\text{bis}(\mu\text{-acetato})$ -bridged analog ($J = 34 \text{ cm}^{-1}$) [31]; (c) coordinate system used to describe symmetry of atomic and molecular orbitals.

provided a detailed semiquantitative understanding of their microscopic mechanisms of spin coupling. Figure 18.3 shows results inferred from extensive broken symmetry calculations on the complete and optimized molecular structures of the complexes [12]. The broken symmetry approximation [19] imposes convergence to a

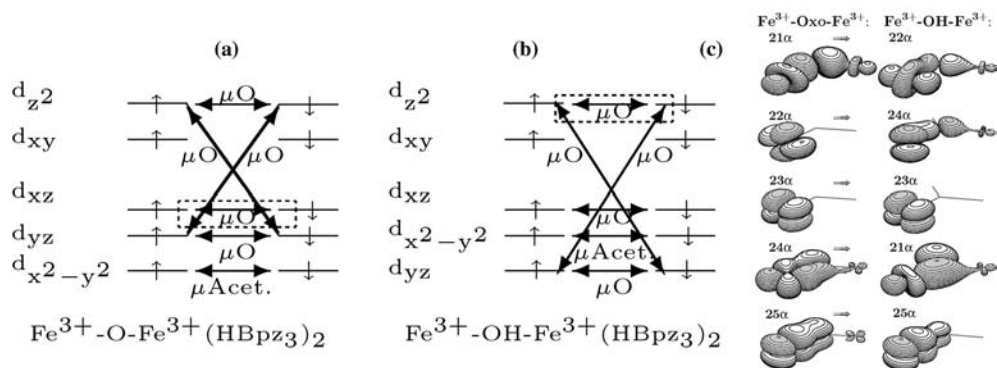


Figure 18.3 Results inferred from extensive broken symmetry calculations on the complete and optimized molecular structures of (a) $(\mu\text{-O})\text{bis}(\mu\text{-acetato})$ -bridged and (b) $(\mu\text{-OH})\text{bis}(\mu\text{-acetato})$ -bridged complexes shown in Figure 18.2; (c) the five one-electron broken symmetry orbitals of each complex.

set of molecular orbitals that are fairly localized on one or the other iron ion. Such localized orbitals mimic superexchange pathways that are responsible for magnetic ordering. Superexchange interactions can be thought of as bridge-mediated electron–electron interactions that ultimately result in magnetic ordering. As illustrated by Figures 18.3a and 18.3b it was possible to identify the main superexchange pathways. One main result of this work was the discovery that, in addition to the oxo- or hydroxo-bridges, the carboxylato bridges also play a role as pathways for superexchange. While the role of the carboxylatos in the oxo-bridged complex is meaningful but not dominant, their role as propagators of antiferromagnetic ordering in the protonated analog is more important. The study found six main superexchange pathways, of which the dominant ones are indicated by the dashed boxes [12]. These findings are not only relevant to these particular complexes but to other related molecular structures and help to understand the antiferromagnetism of, for example, the diiron-oxo proteins azidomet- and azidometmyo-hemerythrin. Figure 18.3c displays (in arbitrary ordering) the five one-electron broken symmetry orbitals (pathways), which are mainly of Fe1(3d) character and host α (spin up) electrons. Equivalent orbitals of main Fe2(3d) character (not shown) that host the spin down (β) electrons were also found. The bottom orbitals (25α) show how the main pathway of the oxo-bridged complex, of main Fe1(d_{xz}) composition, is fairly delocalized towards the other site (Fe2) via π interactions with the bridge. As shown in the figure, upon protonation the same pathway becomes more localized on Fe1, which lowers its efficiency as a mediator of antiferromagnetism. This largely explains the dramatic weakening upon protonation of the spin coupling as measured by the exchange constants.

At a more quantitative level, remarkable progress has been made on the theoretical prediction of coupling constants. The eigenstates of the Heisenberg exchange Hamiltonian (Equation 18.1) have energies given by Equation 18.2. It follows that the exchange interaction partially removes the degeneracy of the various spin manifolds giving rise to energy splittings proportional to J :

$$E_{2S+1}^{\text{HB}} = \frac{1}{2}J\{S(S+1) - S_1(S_1+1) - S_2(S_2+1)\} \quad (18.2)$$

To calculate the exchange constant from density functional calculations one can use Equation 18.2 in conjunction with spin projections [32–34] on the spin unrestricted wavefunctions. One can derive [12, 33] a general expression valid for symmetric or asymmetric binuclear complexes:

$$J = 2 \frac{E_{2S_{\text{max}}+1}^{\text{UKS}} - E_{2S_{\text{min}}+1}^{\text{UKS}}}{S_{\text{max}}(S_{\text{max}}+1) - \langle S^2 \rangle_{2S_{\text{min}}+1}^{\text{UKS}}} \quad (18.3)$$

where $E_{2S_{\text{max}}+1}^{\text{UKS}}$ and $E_{2S_{\text{min}}+1}^{\text{UKS}}$ correspond to the unrestricted Kohn–Sham energies of molecular spin states with highest and lowest spin multiplicity, respectively. A most appealing feature of Equation 18.3 is that it is based on expectation values of energy and spin operators that can be evaluated directly from converged SDFT

calculations. We notice that the formal computation of $\langle S_2 \rangle$ for Kohn–Sham wavefunctions is not straightforward [35]. In practice, however, the methods implemented by Schlegel *et al.* [36] for *ab initio* wavefunctions appear to give a reasonable approximation [12, 32]. In fact, by means of Equation 18.3, theoretical values of unprecedented accuracy have been obtained [12].

Equation 18.3 provides a simple yet powerful method for computing exchange constants. However, to obtain accurate values one still needs to overcome a most difficult problem. That is, one needs to accurately evaluate:

$$E_{2S_{\max}+1}^{\text{UKS}} - E_{2S_{\min}+1}^{\text{UKS}}$$

As it is also true for other physical quantities, the accuracy of this energy difference is somewhat dependent on the particular exchange-correlation functional used to carry out SDFT calculations. In practice, calculations based on the MPW1-PW91 [37] functional have produced fairly accurate results. The computed values [12] for $\text{Fe}^{3+}\text{-O-Fe}^{3+}(\text{HBpz}_3)_2$ ($J = +152.7 \text{ cm}^{-1}$) and $\text{Fe}^{3+}\text{-OH-Fe}^{3+}(\text{HBpz}_3)_2$ ($J = +23.3 \text{ cm}^{-1}$) compare well with experiment and indicate that the method outlined above works equally well for strongly and weakly coupled binuclear centers.

18.4

Phenomenological Simulation of Mössbauer Spectra of Diiron-oxo Proteins

Mössbauer spectra of diiron centers can be simulated by diagonalization of the spin Hamiltonian [13, 14]:

$$\mathcal{H} = J\mathbf{S}_1 \cdot \mathbf{S}_2 + \sum_{i=1}^2 \{ \mathbf{S}_i \cdot \tilde{\mathbf{D}}_i \cdot \mathbf{S}_i + \beta \mathbf{S}_i \cdot \tilde{\mathbf{g}}_i \cdot \mathbf{H} + \mathbf{S}_i \cdot \tilde{\mathbf{a}}_i \cdot \mathbf{I}_i + \mathbf{I}_i \cdot \tilde{\mathbf{P}}_i \cdot \mathbf{I}_i - \beta_n \mathbf{g}_n \cdot \mathbf{H} \cdot \mathbf{I}_i \} \quad (18.4)$$

where $\mathcal{H}$ includes Heisenberg exchange, zero-field splitting (ZFS), electronic Zeeman, magnetic hyperfine, electric quadrupole and nuclear Zeeman interactions. The first three terms of Equation 18.4, which correspond to electronic interactions, can also parameterize data from electron paramagnetic resonance (EPR) and magnetic susceptibility [17]. Accordingly, these can be grouped as an electronic ($\mathcal{H}_{\text{elec}}$) Hamiltonian. Similarly, the last three terms of Equation 18.4 include nuclear interactions and can be grouped as a nuclear ($\mathcal{H}_{\text{nuclear}}$) Hamiltonian.

Rodriguez, Debrunner, *et al.* have applied genetic algorithms for searching the multiparameter space of Equation 18.4 to simulate Mössbauer and EPR spectra. The spectacular success of this approach is illustrated by their analysis of the complex spectra of nitrosyl derivatives of hemerythrin [13] and reduced uteroferrin [14].

18.4.1

Antiferromagnetic Diiron Center of Hemerythrin

Hemerythrin (Hr), an important member of the family of diiron-oxo proteins, is found in marine invertebrates [1–3, 38]. As illustrated by Figure 18.4, the function

of hemerythrin is related to the important but non-enzymatic role of reversible dioxygen binding [39]. This is essential for the transport and storage of molecular oxygen. Combined crystallographic [4, 5, 40, 41] and spectroscopic [1, 18, 42, 43] data show that deoxyHr contains a weakly antiferromagnetically coupled [13, 39, 44–47] hydroxo-bis(carboxylato)-bridged $\text{Fe}^{2+}\text{-Fe}^{2+}$ center [5, 48]. By contrast, oxyHr and other derivatives, such as azidometHr, contain strongly antiferromagnetically coupled [44, 45] oxo-bis(carboxylato)-bridged $\text{Fe}^{3+}\text{-Fe}^{3+}$ centers. Thus, despite their structural similarity, there are meaningful differences in electronic structure among the various forms of hemerythrin and, in general, among the various members of the family of diiron-oxo proteins.

18.4.2

Nitric Oxide Derivative of Hr

Hemerythrin has the physiologically important forms deoxyHr and oxyHr, which contain pairs of hydroxo-bridged high-spin ferrous and oxo-bridged high-spin ferric ions, respectively (Figure 18.4). It has been reported that some single anions bind to the five-coordinate iron site of deoxyHr [18]. Similarly, nitric oxide binds reversibly to one iron of deoxyHr, forming the adduct deoxyHrNO [49]. Upon binding to one iron site of deoxyHr, nitric oxide forms a $\{\text{FeNO}\}^7 (S_1 = 3/2)$ group that in turn couples antiferromagnetically to the $\text{Fe}^{2+} (S_2 = 2)$ site, producing a molecular Kramers doublet ($S = 1/2$) ground state. The Mössbauer simulations confirmed such spin

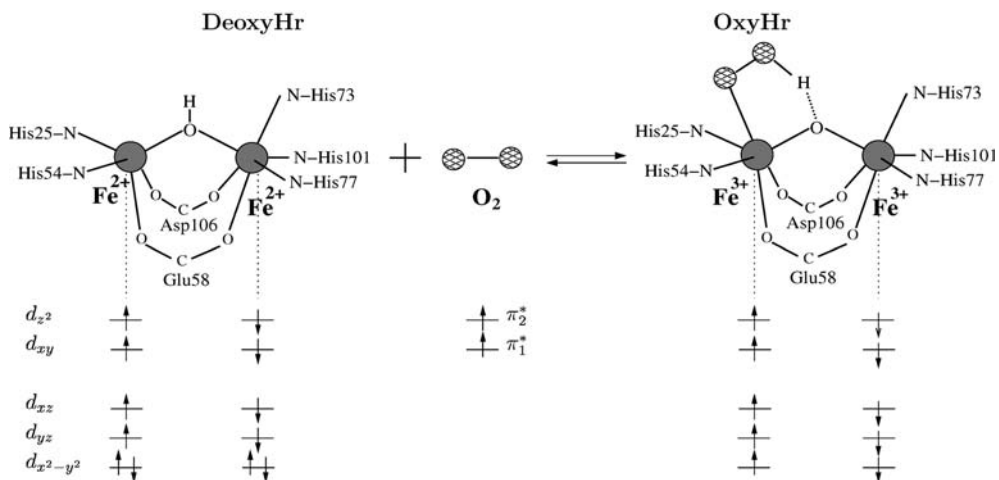


Figure 18.4 DeoxyHr has an antiferromagnetic binuclear center where the unpaired electrons of one ferrous ion are antiparallel to those of the other, giving rise to a total spin $S = S_1 - S_2 = 0$, where $S_1 = S_2 = 2$. Upon binding to the triplet ($S = 1$) ground state of O_2 the protein cycles to the OxyHr state,

where the unpaired electrons of one ferric ion are antiparallel to those of the other, giving rise to a total spin of $S = S_1 - S_2 = 0$, where $S_1 = S_2 = 5/2$. DeoxyHr has one five-coordinate and one six-coordinate ferrous site. OxyHr displays a terminal hydroperoxy ligand that hydrogen bonds to the μ -oxo group.

configurations for the two sites of deoxyHrNO since its 100 K spectra clearly showed two distinct doublets with isomer shifts and quadrupole splittings characteristic of $\{\text{FeNO}\}^7(S_1 = 3/2)$ and $\text{Fe}^{2+}(S_2 = 2)$ species. In contrast to its parent form, which is EPR silent, deoxyHrNO displays the unusual EPR tensor $g^{\text{eff}} = (1.84, 1.84, 2.77)$ [49]. These g values have been rationalized by considering the effect of the sizable ZFS on the energy levels of the, otherwise dominant, exchange interaction [18, 46, 50].

The Mössbauer parameters of $\{\text{FeNO}\}^7(S = 3/2)$ complexes are of particular interest because they cannot be adequately classified as arising from common idealized oxidation states of iron. In particular, their isomer shifts are reduced in magnitude with respect to the high-spin ferrous and enhanced with respect to the high-spin ferric configurations. Several models for the electronic configuration of $\{\text{FeNO}\}^7$ complexes have been proposed [51–53]. Some of these models have been described in terms of ionic electronic configurations of iron and NO. Ionic models, however, have some difficulty explaining Mössbauer, EPR, susceptibility and SDFT results, which suggest considerable anisotropic covalency and, in general, strong valence delocalization. An alternative to localized models has been presented by Rodriguez *et al.* [13] These authors interpreted the unusual Mössbauer isomer shifts of $\{\text{FeNO}\}^7(S = 3/2)$ complexes and other physicochemical parameters in terms of strong Fe-NO valence-electron delocalization. In what follows, some results are shown that illustrate how spin Hamiltonian simulations with complementary SDFT calculations were used to interpret spectroscopic data from this extraordinary example of strong bonding between the NO radical and iron.

Rodriguez, Debrunner, *et al.* simulated applied-field Mössbauer spectra and EPR data of the mixed valence center of the nitrosyl derivative of deoxyHr (deoxyHrNO) (Figure 18.5). Simulations of the highest quality were obtained by searching for optimum values within the multiparameter space of the spin Hamiltonian (18.4) with the aid of genetic algorithms [13, 14]. The simulations required diagonalization of the complete $(2S_1 + 1) \times (2S_2 + 1)$ matrix of the electronic Hamiltonian operator that includes exchange, zero field splitting and electronic Zeeman interactions. Magnetic hyperfine, electric quadrupole and nuclear Zeeman interactions were also included in the simulation of Mössbauer data. The results were consistent with a binuclear center of the form $\text{Fe}^{2+}(S_1 = 2) - \{\text{FeNO}\}^7(S_2 = 3/2)$ where two iron species are antiferromagnetically coupled. Figure 18.5 shows that simulations and experimental data are nearly identical. In addition, as shown in Table 18.1, the calculated and experimental EPR $\tilde{g}$ tensors are nearly identical. Thus, from diagonalization of (18.4) and its parameterization with the aid of genetic algorithms, a single set of spin Hamiltonian parameters that reproduce simultaneously Mössbauer spectra and EPR g values was obtained. This powerful computational methodology yielded not only the Heisenberg exchange constant ($J = 27.83 \text{ cm}^{-1}$) but also ZFS (D, E) and hyperfine parameters in the intrinsic spin representation. As shown in Figure 18.5, the relative orientations of the ZFS ($\tilde{D}_{1,2}$) and other tensors were obtained and, in conjunction with SDFT calculations, were mapped onto the actual molecular geometry: $\mathcal{R}(\tilde{a}_1 \rightarrow \tilde{P}_1) = (18, 12, 67)^\circ$, $\mathcal{R}(\tilde{a}_2 \rightarrow \tilde{P}_2) = (55, 53, 29)^\circ$ and $\mathcal{R}(\tilde{D}_1 \rightarrow \tilde{D}_2) = (0, 90, 28)^\circ$. Here, $\tilde{a}_{1,2}$ and $\tilde{P}_{1,2}$ are the single iron magnetic hyperfine and electric quadrupole tensors, respectively.

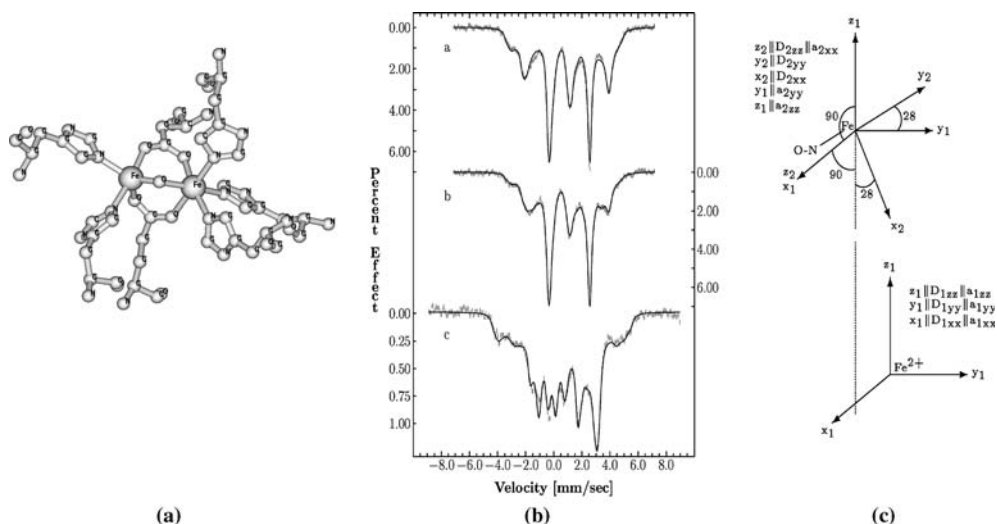


Figure 18.5 (a) DeoxyHr [4, 5] has an $\text{Fe}^{2+}\text{-OH-Fe}^{2+}$ center antiferromagnetically coupled to a $(S_1 = 2) - (S_2 = 2) = 0$ ground state; upon binding of NO ($S = 1/2$) to its five-coordinate Fe^{2+} site, deoxyHr becomes a mixed-valence species with an $\{\text{FeNO}\}^7 (S = 3/2)$ moiety antiferromagnetically coupled to a $\text{Fe}^{2+} (S = 2)$ site [13]; (b) variable-field Mössbauer spectra of deoxyHrNO and computer simulations (solid lines) obtained by parameterization of (18.4) – notice that simulations and experimental data are virtually identical;

(c) absolute orientations of the ZFS and hyperfine tensors with respect to the crystallographic structure of deoxyHrNO obtained from combined spin Hamiltonian simulations and SDFT calculations. Our complementary spin Hamiltonian and SDFT calculations show that the strong Fe–NO bond essentially defines the z (principal) axis of the ZFS and establishes an extremely important correlation between the spin Hamiltonian parameters and the actual geometric molecular structure.

18.4.3

Antiferromagnetic Diiron Center of Reduced Uteroferrin

The purple acid phosphatase uteroferrin (Uf) is a member of the family of diiron-oxo proteins. Uteroferrin is found in pig allantoic fluid and catalyzes the hydrolysis of phosphate esters under acidic pH conditions [54]. Uf can be found in two oxidation states. The oxidized diferric form (Ufo) is enzymatically inactive and EPR silent due to a singlet ($S = 0$) ground state that results from strong antiferromagnetic coupling [54]. By contrast, the reduced mixed valence form (Ufr) exhibits catalytic activity and is characterized by a highly anisotropic EPR tensor [15, 54]. Ufr has a binuclear active center with ferric and ferrous ions antiferromagnetically coupled by a weak exchange interaction [15]. Such a mixed valence diiron center produced complex but highly informative Mössbauer spectra in the presence of external magnetic fields. From analysis based on the spin Hamiltonian, a single set of Hamiltonian parameters that reproduce simultaneously Mössbauer spectra and EPR g values was obtained.

Table 18.1 Spin Hamiltonian parameters obtained from simulations (Figure 18.5) of Mössbauer and EPR data of deoxyHrNO [13].

$J = 27.83 \text{ cm}^{-1}$						
Site	S	$D \text{ (cm}^{-1}\text{)}$	$E \text{ (cm}^{-1}\text{)}$	$\tilde{a}/g_n\beta_n \text{ (Tesla)}$	$\mathcal{R}(\tilde{a} \rightarrow \tilde{D}, \tilde{g})$	$\tilde{g}$
Fe^{2+}	2	+ 5.96	+ 0.18	$-(18.5, 10.4, 13.8)$	$(0, 0, 0)^\circ$	$(2.23, 2.24, 2.14)$
$\{\text{FeNO}\}^7$	3/2	+ 18.93	+ 2.65	$-(29.7, 25.0, \approx 22)$	$(0, 90, 28)^\circ$	$(2.00, 2.00, 2.00)$
$\tilde{g}_{\text{EPR}}^{\text{eff}} = (1.84, 1.84, 2.77)$						
$\tilde{g}_{\text{Calc}}^{\text{eff}} = (1.83, 1.85, 2.78)$						

The corresponding simulations, which are nearly identical to the experimental data, are shown in Figure 18.6.

As shown in Table 18.2, all the spin Hamiltonian parameters corresponding to the two iron sites (i.e., Fe^{3+} and Fe^{2+}) have been obtained. By diagonalization of the electronic part of Equation 18.4 these parameters also yielded calculated EPR g values virtually identical to the experimental data. Most notably, this methodology also afforded the exchange coupling constant, J , by direct simulation of Mössbauer spectra.

18.5 Conclusion

The results presented above illustrate how diagonalization of the spin Hamiltonian, in conjunction with genetic algorithms, allows the simulation of complex resonance spectra in great detail. This procedure determines the spin Hamiltonian parameters of spin coupled transition metal complexes, including the Heisenberg exchange interaction. The phenomenological spin Hamiltonian parameters measure the strength of exchange, ZFS and hyperfine interactions but do not necessarily explain their microscopic physical origin. On the other hand, spin density functional theory

Table 18.2 Spin Hamiltonian parameters obtained from simulations (Figure 18.6) of Mössbauer and EPR data of reduced uteroferrin [14].

[illegible]

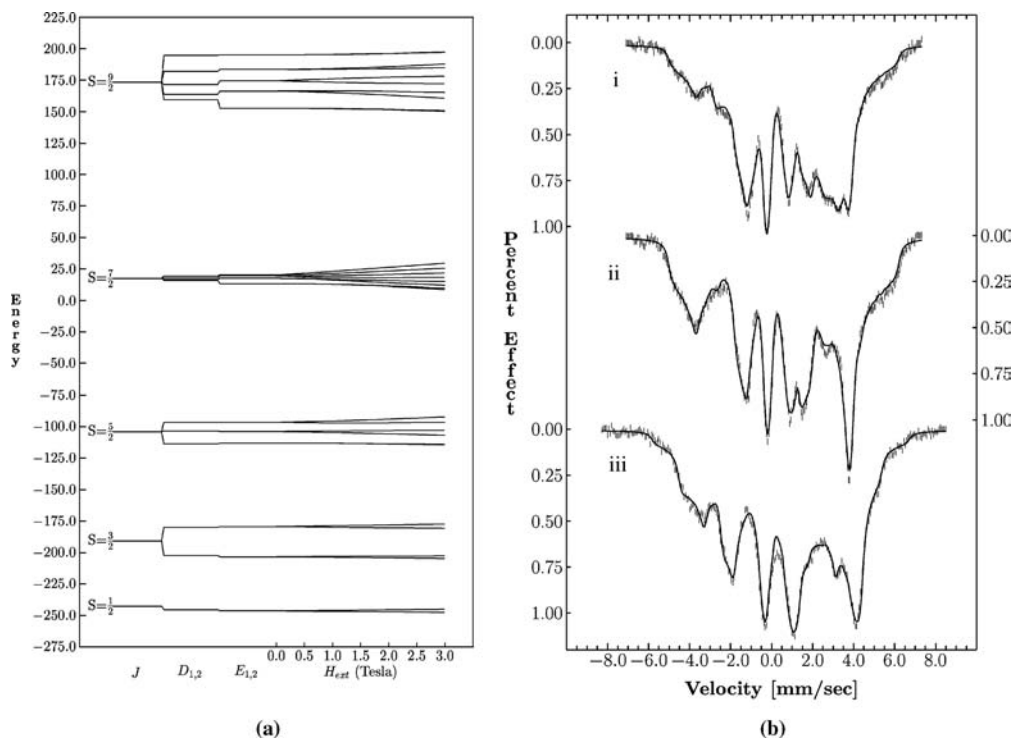


Figure 18.6 (a) Energy level diagram of the diiron-oxo protein reduced uteroferrin obtained from diagonalization of Hamiltonian (18.4). The ground doublet ($S = 1/2$) and higher spin multiplets are shown as the exchange, zero field splitting and Zeeman interactions are turned on. The axial and rhombic ZFS parameters are D_i and E_i . The energies in units of cm^{-1} correspond to one specific molecular orientation (along the x -axis), perpendicular to the quantization (z) axis determined by an applied magnetic field. (b) Computer simulations (solid lines) of Mössbauer

spectra of reduced uteroferrin on top of experimental data. Data were recorded under (i) 0.032 T perpendicular, (ii) 0.032 T parallel and (iii) 3.7 T parallel applied fields with respect to the incident excitation beam. The solid lines are simulations based on the spin Hamiltonian (18.4) and the application of genetic algorithms. Reduced uteroferrin has a mixed valence center where a high spin Fe^{3+} ($S = 5/2$) ion is antiferromagnetically coupled to a Fe^{2+} ($S = 2$) ion, giving rise to a Kramers doublet ($S = 1/2$) ground state. The simulations represent the combined contributions of both iron sites [14].

allows one to understand the detailed electronic structure that gives rise to electric and magnetic phenomena and provides greater insight about the physical origin of exchange and hyperfine interactions. SDFT calculations also permit the accurate prediction of many spectroscopic parameters, including Heisenberg exchange couplings. Results obtained by the parameterization of spin Hamiltonians and those from SDFT are, therefore, complementary. Indeed, this combined approach makes a direct connection between phenomenological parameters, *ab initio* electronic structure, molecular geometry and experimental data. This methodology

is applicable not only to binuclear metal centers in proteins but, quite generally, to other transition metal complexes of interest in (bio)inorganic chemistry. The present quantum mechanical understanding of the electronic structure of active sites of diiron-oxo proteins should contribute significantly to elucidating their mechanisms of catalytic activity and biochemical reactivity.

Acknowledgment

The author thanks the support from the National Science Foundation (NSF) via CAREER award CHE-0349189 (JHR).

References

- 1 Que, L. Jr and True, A.E. (1990) *Prog. Inorg. Chem.*, **38**, 97.
- 2 Kurtz, D.M. (1990) *Chem. Rev.*, **90**, 585.
- 3 Feig, A.L. and Lippard, S.J. (1994) *Chem. Rev.*, **94**, 759.
- 4 Stenkamp, R.E., Sieker, L.C., and Jensen, L.H. (1984) *J. Am. Chem. Soc.*, **106**, 618.
- 5 Stenkamp, R.E., Sieker, L.C., Jensen, L.H., McCallum, J.D., and Sanders-Loehr, J. (1985) *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 713.
- 6 Guddat, L.W., McAlpine, A.S., Hume, D., Hamilton, S., Jersey, J., and Martin, J.L. (1999) *Structure Fold. Des.*, **7**, 757.
- 7 Rosenzweig, A.C., Frederick, C.A., Lippard, S.J., and Nordlund, P. (1993) *Nature*, **366**, 537.
- 8 Nordlund, P., Sjöberg, B.M., and Eklund, H. (1990) *Nature*, **345**, 593.
- 9 Fox, B.G., Shanklin, J., Somerville, C., and Miinck, E. (1993) *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 2486.
- 10 Armstrong, W.H. and Lippard, S.J. (1983) *J. Am. Chem. Soc.*, **105**, 4837.
- 11 Hotzelmann, R., Wieghardt, K., Florke, U., Haupt, H.-J., Weatherburn, D., Bonvoisin, J., Blondin, G., and Girerd, J. (1992) *J. Am. Chem. Soc.*, **114**, 1683.
- 12 Rodriguez, J.H. and McCusker, J.K. (2002) *J. Chem. Phys.*, **116**, 6253.
- 13 Rodriguez, J.H., Xia, Y.M., and Debrunner, P.G. (1999) *J. Am. Chem. Soc.*, **121**, 7846.
- 14 Rodriguez, J.H., Ok, H.N., Xia, Y.M., Debrunner, P.G., Hinrichs, B.E., Meyer, T., and Packard, N. (1996) *J. Phys. Chem.*, **100**, 6849.
- 15 Day, E.P., David, S., Peterson, J., Dunham, W., Bonvoisin, J., Sands, R., and Que, L. Jr (1988) *J. Biol. Chem.*, **263**, 15561.
- 16 Averill, B.A., Davis, J.C., Burman, S., Zirino, T., Sanders-Loehr, J., Loehr, T.M., Sage, J.T., and Debrunner, P.G. (1987) *J. Am. Chem. Soc.*, **109**, 3760.
- 17 Bencini, A. and Gatteschi, D. (1990) *EPR of Exchange Coupled Systems*, Springer Verlag, Berlin.
- 18 Reem, R.C. and Solomon, E.I. (1987) *J. Am. Chem. Soc.*, **109**, 1216.
- 19 Noodleman, L. (1981) *J. Chem. Phys.*, **74**, 5737.
- 20 Heisenberg, W. (1928) *Z. Phys.*, **49**, 619.
- 21 Dirac, P. (1929) *Proc. R. Soc. London*, **A123**, 714.
- 22 Levy, P. (1969) *Phys. Rev.*, **177**, 509.
- 23 Van Vleck, J.H. (1934) *Phys. Rev.*, **45**, 405.
- 24 Anderson, P.W. (1959) *Phys. Rev.*, **115**, 2.
- 25 Anderson, P.W. (1950) *Phys. Rev.*, **79**, 350.
- 26 Hay, P.J., Thibault, J.C., and Hoffmann, R.J. (1975) *J. Am. Chem. Soc.*, **97**, 4884.
- 27 Brown, C.A., Remar, G.J., Musselman, R.L., and Solomon, E.I. (1995) *Inorg. Chem.*, **34**, 688.
- 28 Rodriguez, J.H., Xia, Y.M., Debrunner, P.G., Chaudhuri, P., and Wieghardt, K. (1996) *J. Am. Chem. Soc.*, **118**, 7542.
- 29 Goodenough, J.B. (1955) *Phys. Rev.*, **100**, 564.
- 30 Goodenough, J.B. (1963) *Magnetism and the Chemical Bond*, Interscience, New York.
- 31 Armstrong, W.H., Spool, A., Papaefthymiou, G.C., Frankel, R.B., and

- Lippard, S.J. (1984) *J. Am. Chem. Soc.*, **106**, 3653.
- 32 Rodriguez, J.H., Wheeler, D.E., and McCusker, J.K. (1998) *J. Am. Chem. Soc.*, **120**, 12051.
 - 33 Yamaguchi, K., Jensen, F., Dorigo, A., and Houk, K. (1988) *Chem. Phys. Lett.*, **149**, 537.
 - 34 Yamaguchi, K., Takahara, Y., and Fueno, T. (1986) Applied quantum chemistry, in *Applied Quantum Chemistry* (eds V.H. Smith Jr, H.F. Schaefer III, and K. Morokuma), D. Reidel Publishing, Dordrecht.
 - 35 Wang, J. and Becke, A.D. (1995) *J. Chem. Phys.*, **102**, 3477.
 - 36 Chen, W. and Schlegel, H.B. (1994) *J. Chem. Phys.*, **101**, 5957.
 - 37 Adamo, C. and Barone, V. (1998) *J. Chem. Phys.*, **108**, 664.
 - 38 Stenkamp, R.E. (1994) *Chem. Rev.*, **94**, 715.
 - 39 Lippard, S.J. and Berg, J.M. (1994) *Principles of Bioinorganic Chemistry*, University Science Books, Mill Valley, CA.
 - 40 Sheriff, S., Hendrickson, W.A., and Smith, J.L. (1983) *Life Chem. Rep. Suppl. Ser.*, **1**, 305.
 - 41 Holmes, M.A., Trong, I.L., Turley, S., Sieker, L.C., and Stenkamp, R.E. (1991) *J. Mol. Biol.*, **218**, 583.
 - 42 Okamura, M.Y., Klotz, I.M., Johnson, C.E., Winter, M., and Williams, R.J.P. (1969) *Biochemistry*, **8**, 1951.
 - 43 Garbett, K., Johnson, C.E., Klotz, I.M., Okamura, M.Y., and Williams, R.J.P. (1971) *Arch. Biochem. Biophys.*, **142**, 574.
 - 44 Moss, T.H., Moleski, C., and York, J.L. (1971) *Biochemistry*, **10**, 840.
 - 45 Dawson, J., Gray, H., Hoenig, H.E., Rossman, G., Shredder, J., and Wang, R.-H. (1972) *Biochemistry*, **11**, 461.
 - 46 Reem, R.C. and Solomon, E.I. (1984) *J. Am. Chem. Soc.*, **106**, 8323.
 - 47 Maroney, M.J., Kurtz, D.M., Nocek, J.M., Pearce, L., and Que, L. Jr (1986) *J. Am. Chem. Soc.*, **108**, 6871.
 - 48 Holmes, M.A. and Stenkamp, R.E. (1991) *J. Mol. Biol.*, **220**, 723.
 - 49 Nocek, J.K., Kurtz, D.M. Jr, Sage, J.T., Debrunner, P.G., Maroney, M.J., and Que, L. Jr (1985) *J. Am. Chem. Soc.*, **107**, 3382.
 - 50 Sage, J.T. and Debrunner, P.G. (1986) *Hyperfine Interact.*, **29**, 1399.
 - 51 Brown, C.A., Pavlosky, M.A., Westre, T.E., Zhang, Y., Hedman, B., Hodgson, K.O., and Solomon, E.I. (1995) *J. Am. Chem. Soc.*, **117**, 715.
 - 52 Bill, E., Bernhardt, F.H., Trautwein, A.X., and Winkler, H. (1985) *Eur. J. Biochem.*, **147**, 177.
 - 53 Hauser, C., Glaser, T., Bill, E., Weyhermuller, T., and Wieghardt, K. (2000) *J. Am. Chem. Soc.*, **122**, 4352–4365.
 - 54 Antanaitis, B.C., Aisen, P., Lilienthal, H.R., Roberts, R.M., and Bazer, F.W. (1980) *J. Biol. Chem.*, **255**, 11204.

19

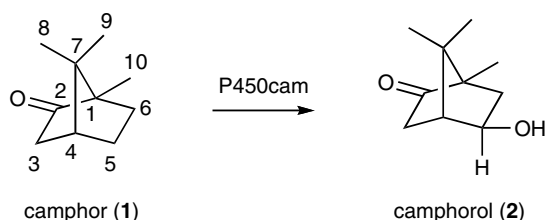
Accurate Description of Spin States and its Implications for Catalysis

Marcel Swart, Mireia Güell, and Miquel Solà

19.1

Introduction

Reactivity patterns in organometallic and bioinorganic chemistry often depend critically on the spin state, for example, on the spin-state preferences of reactants, products, intermediates and transition states [1, 2]. One of the best examples of this is presented by the catalytic cycle of cytochrome P450cam that catalyzes the hydroxylation of (*R*)-camphor (**1**) to 5-*exo*-camphorol (**2**) (Scheme 19.1) [3].



Scheme 19.1 Hydroxylation reaction catalyzed by cytochrome P450cam enzyme.

In the catalytic cycle [4], a low-spin doublet is observed for the resting state, which goes to a high-spin sextet after substrate binding [5–10]. In subsequent steps other spin-state dependent features are present (see below). This spin-flip in the first step seems to be determined primarily by the presence or absence of water molecules in the active site, and is vital for the specificity of the reaction taking place. Key factors for this specificity are the presence of a potassium ion close to the active site [11] and the presence of a tyrosine residue (Tyr96) in the active site. The latter is easily comprehended as the tyrosine residue serves as an anchor for the natural substrate, (*R*)-camphor, and prepares it maximally for the subsequent hydroxylation [12–14]. For instance, a previous study [12] has shown that upon binding to Tyr96 the substrate

is distorted, especially at C5, where the reaction takes place. Replacement of the Tyr96 residue by phenylalanine reduces the regio- and stereospecificity dramatically [11]. The role of the potassium is not related directly to the hydroxylation reaction but more with an entrance channel of water to the active site [11].

Another intriguing aspect of the catalytic cycle of P450cam is related to the step after the substrate has entered the active site [3]. The cycle continues with an electron transfer from a reducing agent, in the natural system putidaredoxin [15, 16], to give a quintet state after the transfer. This electron transfer takes place *only* when substrate is present, that is, only for the high spin state. Moreover, it seems that dioxygen can only enter the active site when the electron transfer has taken place. In this process of dioxygen binding, there is a second spin-flip present. The paramagnetic dioxygen molecule (with a triplet ground-state) on its own already induces a change of spin-state, which would have lead to a triplet. Instead, it goes to a singlet state.

To help understand these subtleties and explain the missing steps in the reaction mechanism that occur too fast for experiments to follow, theoretical chemistry can play an important role. However, theory is not without its own problems. The most popular method for studying (bio)inorganic catalysis is presented by density functional theory (DFT) [17–19], owing to its efficiency, which enables one to treat large systems of up to hundreds of atoms in a reasonable time. Almost all studies in the literature so far have used either the B3LYP [20, 21] or BP86 [22, 23] functional – for most simple cases they both give good results. This is no longer true when spin-states with energy levels in close proximity are involved [24, 25], that is, when a transition metal like chrome, manganese, iron, cobalt or nickel is present. Previous studies [26–28] have shown that both of these functionals are unable to correctly predict the spin ground-state of these transition-metal complexes. Early generalized gradient approximations (GGA) functionals like BP86 tend to over-stabilize the low-spin state, while hybrid functionals like B3LYP tend to over-stabilize the high-spin state (due to the inclusion of a portion of Hartree–Fock exchange) [25, 29]. Several remedies have been proposed, such as lowering the amount of Hartree–Fock exchange in B3LYP to 15% (B3LYP*) [26], mixing the Becke88 and PW91x [30, 31] exchange functionals (XLYP, X3LYP) [32] or introducing a Hubbard U parameter [33], but none of these was really satisfactory for all situations. In fact, it was suggested that the spin-state energies should be calculated with several functionals. This unfortunate situation was radically changed by the application of Handy and Cohen’s optimized exchange (OPTX) functional (which is abbreviated as O in combination with other functionals, see below) [34].

Previous validation studies have shown the validity of the OPBE [24, 35] (and OLYP) [36–42] functional for the spin-state splittings of iron complexes. While for the vertical spin-state splittings (see Figure 19.1 for the difference between vertical and relaxed splittings) several DFT functionals could be trusted to give the correct spin ground-state [24], this picture changed for relaxed splittings [25]. In the latter case, only OPBE could be trusted completely. All the other functionals failed, including B3LYP* and the recent M05 and M06 functionals [43, 44]. Perhaps the only exception could be the TPSSH functional [45, 46], which for most complexes gives good results, but still fails in some cases.

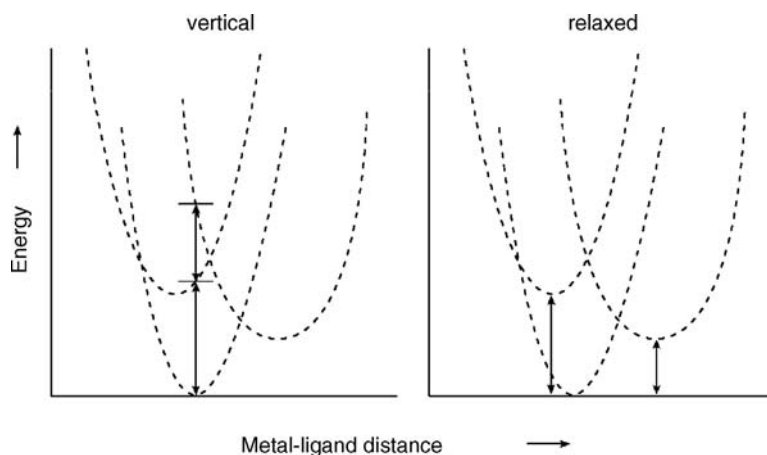


Figure 19.1 Vertical versus relaxed spin-states.

19.2

Influence of the Basis Set

Despite the excellent results obtained in most other studies, some recent studies have criticized the OPBE functional [47–50], claiming that it gives inferior results for spin-state splittings. These criticisms were mainly based on small energy differences, that is, spin state splittings of less than 1 kcal mol^{-1} ($1 \text{ kcal} = 4.184 \text{ kJ}$). Moreover, one of the major differences was the use of small basis sets or basis sets containing effective core potentials (ECPs) in these papers. The ECPs are in fact not basis sets but model Hamiltonians [51], and when combined with a valence basis set are here called ECPB basis sets. Since Yoshizawa and coworker [52] found that the influence of the basis set was many times larger (up to 8 kcal mol^{-1}) than the energy differences on which the OPBE functional was criticized, we recently performed a systematic investigation of the influence of the basis set. In that study [53], we showed that the use of Slater-type orbital (STO) and Gaussian-type orbital (GTO) basis sets give the same answer in the case of large basis sets. However, STO basis sets converge very rapidly with basis set size, while GTO basis sets converge much slower. This is in particular true for relaxed spin-state splittings, where for the GTO basis sets we needed to go to the very demanding cc-pVTZ and cc-pVQZ basis sets to achieve good results [53].

The use of ECPBs for spin-state splittings was shown to be unreliable. The replacement of the core electrons by a model Hamiltonian turned out to give splittings that are fundamentally different from the results obtained with the converged GTO and STO basis sets. This is true in particular for high spin-states. Because of the unreliability of the ECPBs for spin-state splittings, they will not be discussed further in this chapter.

Although the use of large GTO basis sets (cc-pVTZ, cc-pVQZ, Roos-ANO-aug-dz) gave good results it is not in general true that by making the basis set larger one converges to the correct result [53]. For instance, a series of Pople-type basis sets

showed oscillating behavior, with the largest basis (6-311+G**) even giving the worst results of the series. Similarly, for the Ahlrichs basis sets it is in fact not true that adding more valence and/or polarization functions increases the accuracy. Nevertheless, for vertical splittings the Ahlrichs basis sets gave far superior results than the Pople ones, and converged much faster.

The worst performing bases in the GTO series were, not surprisingly, the small 3-21G and 6-31G, and the often-used 6-31G* basis set. Pulay and coworkers, however, noted that the 6-31G* basis set for transition metals is flawed [54], and proposed a modification of it (m6-31G*) that came to our attention due to their very recent paper about the “danger of using quantum chemistry programs as black boxes” [55]. Independently, we had already found that by making a small adjustment the 3-21G, 6-31G basis sets (with and without polarization functions) could be significantly improved. This is best shown by going back to one of the compounds [FeFHOH] in our recent paper [53], which was used as prototypical small example of relaxed spin states. Moreover, the compound did not show any problems with SCF convergence with any of the different quantum-chemistry programs used, which enables a straightforward comparison.

Table 19.1 gives the relaxed spin-state splittings for several basis sets. The reference values are obtained with a large STO basis (TZ2P) and a large GTO basis (cc-pVQZ)

Table 19.1 Relaxed spin-state splittings (kcal mol⁻¹) of Fe(FH)OH and Ni(edt)₂²⁻ ^{a)} with the OPBE functional and several STO/GTO basis sets.

Basis set	FeFHOH $E_{\text{doub}} - E_{\text{quar}}$	FeFHOH $E_{\text{sext}} - E_{\text{quar}}$	Ni(edt) ₂ ²⁻ $E_{\text{trip}} - E_{\text{sing}}$
Reference data ^{b)}			
TZ2P ^{c)}	26.8	8.9	10.3
cc-pVQZ ^{d)}	26.1	9.1	11.4
Small GTOs			
3-21G ^{d),e)}	28.2	-0.7	-24.2
3-21G* ^{d),e)}	28.1	-0.1	-15.8
6-31G ^{d),e)}	26.0	0.8	-8.0
6-31G* ^{b),d)}	27.3	0.8	-1.8
Modified GTOs			
m6-31G* ^{d),e)}	26.8	7.3	5.3
s3-21G ^{d),e)}	25.1	18.2	10.4
s3-21G* ^{d),e)}	24.8	18.0	15.7
s6-31G ^{d),e)}	27.8	11.8	6.2
s6-31G* ^{d),e)}	27.7	11.4	10.7

a) edt = ethane-1,2-dithiolate.

b) See Reference [53].

c) STO basis set.

d) GTO basis set.

e) This work.

and show similar values, especially for the quartet–sextet splitting. The doublet energy is placed at somewhat different energies by the two large basis sets, but this is most likely influenced by the small differences in geometry and the spin-contamination, for which no correction was done.

The small Pople-type GTO basis sets 3-21G, 6-31G and 6-31G* wrongly place the sextet state at almost the same energy as the quartet ground-state (Table 19.1). The modification of the 6-31G* basis by Pulay and coworkers (m6-31G*) [54] indeed improves the results significantly, and leads to a quartet–sextet splitting of about 7 kcal mol^{−1}. Although this still underestimates the splitting by about 2 kcal mol^{−1}, it is a significant improvement and in fact after the huge basis sets (cc-pVTZ, cc-pVQZ, Roos-ANO-aug-dz) it is the best GTO result so far.

The only difference between the 6-31G* and m6-31G* basis sets is observed in the 3d shell of the first-row transition metals, where the former lacks a sufficiently diffuse outer d orbital. For instance, for iron the most diffuse d-function has an exponent of ~ 0.50 in 6-31G* that is reduced to ~ 0.36 in m6-31G*. At the same time, the most compact d-function also changes considerably, with an exponent of ~ 23.15 (6-31G*) versus ~ 19.30 (m6-31G*) [54].

We also investigated the poor performance of the small GTO basis sets, but from a different perspective. GTOs have two major disadvantages: in the core region and at long radial distance; that is, they do not have a cusp at the nucleus as they should have, and they fall to zero too rapidly. Since either one of these could be responsible for the poor performance we investigated where the deficiency of the small GTOs came from. To do so, we added d-functions to the iron basis set using an even-tempered approach for choosing the value of the exponent. We started by adding d-functions with large exponents for a better description of the region near the nucleus. Although this led to significant changes of the total energies, the quartet–sextet splitting was hardly affected. Instead, when we added a diffuse d-function we immediately observed the desired result; that is, by doing so, the quartet–sextet splitting increased significantly and the correct spin ground-state was obtained. This is in line with the observations by Pulay and coworkers [54] who also found that the 3d shell of the transition metals was not diffuse enough. Hereafter, we refer to our corrected basis sets, in which a diffuse d-function was added for transition metals, as the spin-state-corrected basis sets (s3-21G, s6-31G*, etc.).

Table 19.1 gives the spin-state splittings for the spin-state-corrected GTO basis sets for our test molecule FeFHOH (**3**), as well as the nickel-complex Ni(edt)₂^{2−} (**4**, edt = ethane-1,2-dithiolate, see Figure 19.2). The latter complex is of relevance for the NiFe hydrogenase enzyme (Section 19.6.3) [28]. Note that the 3-21G* basis was obtained by adding a polarization f-function with exponent 0.8 to 3-21G iron, in similar fashion as 6-31G* versus 6-31G. The spin-state splittings are indeed significantly improved by these spin-state-corrected GTOs. All of them now correctly and clearly identify the quartet state of FeFHOH as ground-state, with the sextet higher in energy by 11–18 kcal mol^{−1}. Moreover, the trend of increasing basis set size is now to lower the quartet–sextet splitting, which was also observed for the STO basis sets.

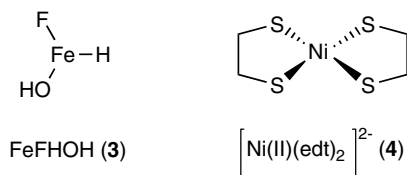


Figure 19.2 Simple model compounds FeFHOH (3) and Ni(edt)_2^{2-} (4).

The iron–ligand distances of FeFHOH are also significantly improved by the spin-state-corrected GTOs and m6-31G*. For example, the mean absolute deviation with respect to the reference TZ2P data [53] are 0.036 Å (3-21G) versus 0.023 Å (s3-21G), 0.018 Å (6-31G) versus 0.008 Å (s6-31G) and, finally, 0.024 Å (6-31G*) versus 0.017 Å (m6-31G*) and 0.011 Å (s6-31G*). Therefore, in all cases the modifications do lead to significantly more accurate results.

The spin-state splitting of Ni(edt)_2^{2-} is also drastically improved by the modified small GTOs. Without exception, the original small GTO basis sets all wrongly predict a triplet ground-state for the complex with a substantial deviation from the reference TZ2P data that indicated a singlet–triplet splitting of +10 kcal mol^{−1} (Table 19.1). The worst result is obtained with 3-21G, similar to the situation of FeFHOH, with a splitting of −24 kcal mol^{−1}; note that this amounts to a deviation of about 34 kcal mol^{−1}! The widely used 6-31G* basis does not fare any better with a splitting of −2 kcal mol^{−1}. It still does not give the correct spin ground-state, and the deviation is still substantial (12 kcal mol^{−1}).

Clearly better results are obtained with the m6-31G* basis and our spin-state-corrected basis sets. All of them correctly predict the spin ground-state for this nickel-complex, with a singlet–triplet splitting that is reasonably close to the reference data with the TZ2P basis. Similar to FeFHOH, good results are obtained with the s6-31G* basis set.

Based on these results, we do not recommend the use of one of the original small GTO basis sets but instead the use of one of the modified ones if one needs to use a small basis set. In particular, our spin-state corrected s6-31G* basis seems to be a reasonable choice in terms of accuracy versus computational demand.

19.3

Spin-Contamination Corrections

When one uses the unrestricted Hartree–Fock (UHF) method, it often happens that the obtained wavefunction does not correspond to a pure spin-state. This arises because the α - and β -orbitals are not necessarily the same [56], with concomitant contamination of the pure spin-state. An easy and straightforward descriptor for the amount of spin-contamination is obtained by inspecting the expectation value of $\langle S^2 \rangle$. For pure spin states, this expectation value is given by $S \cdot (S + 1)$, where S is the total spin polarization ($S = 0.5$ for a doublet, 1.0 for a triplet, etc.). This leads to expectation values for pure spin states of 0 (singlet), 0.75 (doublet), 2.0 (triplet), 3.75 (quartet), 6.0

Table 19.2 Orbital occupations of contaminated-singlet and pure-triplet Ti-porphyrin.

Irrep	Contaminated singlet		Pure triplet	
	N_α	N_β	N_α	N_β
A_1	20	20	20	20
A_2	11	11	11	11
B_1	16	15	16	15
B_2	15	16	16	15

(quintet), 8.75 (sextet), and so on. However, in many cases one obtains a larger value for $\langle S^2 \rangle$ than expected, which results from mixing in of a state with higher multiplicity. The easiest way of demonstrating this is given by a system that has an overall singlet, but which is made up of an α -doublet in one irreducible representation (irrep) coupled with a β -doublet in a different irrep. This is for instance the case for the singlet state of Ti-porphyrin [57], where within C_{2v} symmetry an α -doublet in B_1 is coupled to a β -doublet in B_2 (Table 19.2).

However, this $\alpha\beta$ combination is not a pure spin-state but rather a 1 : 1 mixture of singlet and triplet; the pure singlet would actually be obtained with $(\alpha\beta - \beta\alpha)/\sqrt{2}$. The admixture of singlet and triplet states is reflected in the expectation value of $\langle S^2 \rangle$, which has a value of 1.0. To obtain the energy of the pure spin state (E_{pure}), one can calculate the energy of a pure triplet state (E_{S+1} , e.g., $\alpha\alpha$), and correct the contaminated energy (E_{cont}) for it as follows:

$$E_{\text{pure}} = \frac{E_{\text{cont}} - a \cdot E_{S+1}}{1 - a}; \quad a = \frac{\langle S^2 \rangle_{\text{cont}} - S \cdot (S + 1)}{\langle S^2 \rangle_{S+1} - S \cdot (S + 1)} \quad (19.1)$$

Equation (19.1) results directly from considering the energy as the sum of the pure spin state plus a portion of a state with higher multiplicity ($S + 1$). Note that this holds both for UHF [56] and for UDFT [58].

Corrections for spin-state contamination are typically performed only for the energy but we have in the past couple of years [29, 57, 59] advocated its use also for the gradient, and hence geometry optimizations, and the Hessian matrix, for vibrational frequencies. This is easily achieved within the QUILD program [59], where the application of Equation 19.1 is fully automated. As a result, the obtained spin-state splittings correspond really to the pure spin-state at the pure spin-state geometry, with corresponding frequencies.

As an example of the relevance of this spin-contamination correction for spin-state splittings, consider again our FeFHOH system. The doublet state was found to be severely spin-contaminated, as shown by the expectation values of $\langle S^2 \rangle$, which typically had values of around 1.7 instead of the required 0.75 [53]. If we now correct the energy for spin contamination the quartet-doublet splitting increases by some 11 kcal mol⁻¹. Furthermore, the correction leads to changes in the doublet iron-ligands of 0.004 Å (Fe–F), 0.006 Å (Fe–H) and 0.010 Å (Fe–O). Moreover, we also noticed in our follow-up paper on the influence of the basis set that for some DFT

functionals it depended on the basis set as to whether a spin-contaminated or a pure spin-state was obtained. Comfortingly, the energy of the spin-contaminated calculation after correction was found to be similar to that of the pure spin state.

19.4

Influence of Self-Consistency

In the past we have often exploited the METAGGA scheme within the ADF program [24, 25, 60, 61], in which the energies of a large number of DFT functionals can be computed in a post-SCF fashion. Namely, we first carry out the SCF procedure with one functional, for example, OPBE with the TZ2P basis. The program then uses the resulting (OPBE/TZ2P) Kohn–Sham orbitals, densities, kinetic energy densities, and so on to compute the energies of several other functionals. The energy of the latter functionals is therefore not obtained in a self-consistent manner. Indeed, it may differ by some kcal mol^{-1} from the value that would have been obtained when that particular functional was used in the SCF procedure. However, and more importantly, the relative energies of, for example, a reactant complex versus a transition state, or the relative energy of a doublet versus a quartet state, are much less affected (on the order of $0.1\text{--}0.3 \text{ kcal mol}^{-1}$) [24, 62].

This situation might be different for functionals involving a portion of Hartree–Fock exchange (the hybrid functionals), since this term may depend more directly and/or differently on the functional. Moreover, because of the different implementation of HF exchange in the STO-based ADF program [63] versus GTO-based quantum-chemistry programs, it is *a priori* unclear whether this influences the spin-state splittings. Therefore, we re-examined the FeFHOH and Ni(edt)_2^{2-} systems that were already used in Section 19.2, but now by looking at different functionals. Table 19.3 gives the results for several DFT functionals, including the mostly used B3LYP, BP86, B3LYP* functionals, the recently developed X3LYP [32] and M06 [44], and also the semiempirical PM6 method [64]. The latter is seen to be useless for spin-state splittings, as it predicts a doublet ground-state for FeFHOH with the quartet state higher in energy by 22 kcal mol^{-1} and the sextet state 66 kcal mol^{-1} above the quartet.

In the first and fourth columns of Table 19.3 we give the energy of a functional when it is used itself within the SCF procedure within the ADF program, in the second and fifth columns when the energy is determined post-SCF on OPBE/TZ2P orbitals and densities, and in the third and last columns when the functional used is SCF with the s6-31G* basis set. When we compare the results of the STO and GTO basis set, we see that there is a more or less constant shift of about $1\text{--}3 \text{ kcal mol}^{-1}$ for both systems. This in itself is not that surprising since we compare here a triple- ζ valence plus double-polarization basis with a double- ζ valence with single-polarization on the heavy atoms. However, these differences do not influence the assignment of the spin ground-state (unlike the original small GTO basis sets, see Section 19.2). The spin-state splittings of the functionals with respect to each other are even more constant (Table 19.3). For instance, the quartet–sextet splitting of

Table 19.3 Relaxed spin-state splittings (kcal mol⁻¹) of FeFHOH and Ni(edt)₂²⁻ with several DFT functionals, SCF and post-SCF, and with both a STO and GTO basis set.

Functional	$E_{\text{sext}} - E_{\text{quar}}$ FeFHOH			$E_{\text{trip}} - E_{\text{sing}}$ Ni(edt) ₂ ²⁻		
	SCF(STO) ^{a)}	Post-SCF ^{a),b)}	SCF(GTO) ^{c)}	SCF(STO) ^{a)}	Post-SCF ^{a),b)}	SCF(GTO) ^{c)}
OPBE	8.9	8.8 ^{d)}	11.4	10.2	10.2 ^{d)}	10.7
BP86	17.5	17.6	19.5	10.8	11.1	10.7
B3LYP	12.5	13.1	15.5	0.7	2.4	1.2
B3LYP*	14.3	14.6	17.1	4.1	5.0	4.5
X3LYP	12.0	12.7	15.0	-0.1	2.1	0.5
TPSS	N/a ^{e)}	17.6	20.2	N/a ^{e)}	9.3	8.9
M05	N/a ^{e)}	2.0	3.4	N/a ^{e)}	5.1	4.7
M06	N/a ^{e)}	5.3	5.1	N/a ^{e)}	8.5	7.1
PM6	66.3 ^{f)}	—	—	2.0	—	—

a) Using all-electron TZ2P STO-basis for ADF calculations, SZ STO-basis for PM6.

b) Post-SCF on OPBE orbitals densities.

c) Using s6-31G* GTO-basis.

d) The difference between the SCF and post-SCF energy of OPBE results from the fact that the fitted density is used in the former and the “exact” density in the latter; for more details about the differences between these two densities see Reference [63].

e) Not available as SCF functional in ADF program.

f) The semiempirical PM6 predicts a doublet-quartet splitting of -22.1 kcal mol⁻¹, which is wrongly a doublet ground-state.

FeFHOH shows the trend OPBE $\ll$ X3LYP < B3LYP < B3LYP* $\ll$ BP86, for both the STO and the GTO basis. The same pattern emerges for the singlet–triplet splitting of Ni(edt)₂²⁻, even though the order of the functionals is different: X3LYP < B3LYP $\ll$ B3LYP* $\ll$ OPBE $\sim$ BP86. Finally, the influence of self-consistency is a little larger for hybrid functionals (0.4–0.6 kcal mol⁻¹) than for pure functionals (0.1–0.3 kcal mol⁻¹), but still sufficiently low to enable the post-SCF scheme.

19.5

Spin-States of Model Complexes

Given the problems of theoretical methodologies in correctly describing the spin ground-state of transition-metal complexes, it is, *a priori*, unclear how this will affect studies on reaction mechanisms in metalloenzymes. In these systems, the situation is even more complicated as not only the spin-state may be important but also the coupling with the protein environment. Furthermore, the way in which the protein environment is modeled starts to play a role as well; that is, if it is described as if it were a dielectric continuum or included explicitly within QM/MM (quantum mechanics/molecular mechanics) calculations [65–67]. [QM/MM methods are reviewed in Chapters 2–4, continuum solvation models are discussed in Chapter 4.] In the latter case, the actual QM/MM setup also may influence the results severely. For instance, how are the electrostatic interactions between the QM and MM regions

dealt with, where the electrostatic field of the MM region is felt by the QM electrons or not, and, if so, how to avoid double-counting? The latter might occur when one uses for instance a standard force field like Amber [68] or Charmm [69], where the electrostatic interactions are taken into account at the MM level. Simply applying the same MM charges for coupling to the QM electrons is very likely going to distort the QM/MM interaction terms significantly. A further complication of QM/MM calculations is posed by the question of where to cut off the QM region. In Section 19.6.1 we will return to this latter issue by discussing an example where this latter question is of the utmost importance.

Because of all these problems and doubts that arise when dealing with reaction mechanisms in enzymes, many studies have focused first on smaller model systems where the separation of the QM and MM region is more natural. For instance, metalloporphyrins are widely used as models for reactivity patterns of the heme-group in iron proteins. Since the smaller model systems allow for a rapid and more detailed study, a comparison can be made between different methodologies that in many cases can be directly coupled to experimental data on the same model system. However, one should be careful as the experimental and theoretical systems are not always one and the same. The best example is again given by iron-porphyrin, which has been theoretically studied very often without any axial ligands or peripheral substituents [50, 70–74]. Experimentally however, this system is unattainable and it is usually present *with* the substituents and/or ligands. This can have a marked effect on the properties of the macrocycle, which might actually result from a synergistic effect of having both a metal atom and meso-substituents present [75].

Even when the system studied is the same in both the experimental and the computational study, a difference may persist that hampers a straightforward comparison. For instance, one often uses X-ray crystallography for determining structural features of transition-metal complexes. However, it might be that the crystallization procedure influences the structure of the complex to such an extent that it is no longer compatible with the structure of an isolated complex (whether in the gas-phase or in solution). Nevertheless, these crystal packing effects are often ignored by experimentalists, but at the same time used by theoreticians as explanation for why their optimized structure(s) differs considerably from the experimental crystal structure. A further complication may arise for certain complexes that show temperature-dependencies of the population of different spin-states. This is for instance observed in spin-crossover compounds, where at low temperatures a low-spin state prevails while at elevated temperatures a higher spin-state is observed. Typically, these compounds contain Fe(II) ions, which facilitates the use of Mössbauer spectroscopy to probe the temperature dependence of the population of the different spin-states.

Apart from comparing with experimental data, it is now often also possible to compare DFT results with *ab initio* data such as CCSD(T) or CASPT2. Given that these methodologies are much more computationally demanding, further simplifications are necessary. In Section 19.6.2, we discuss several small models for iron-porphyrin with an axial histidine ligand, where the smallest models have been described also by CCSD(T).

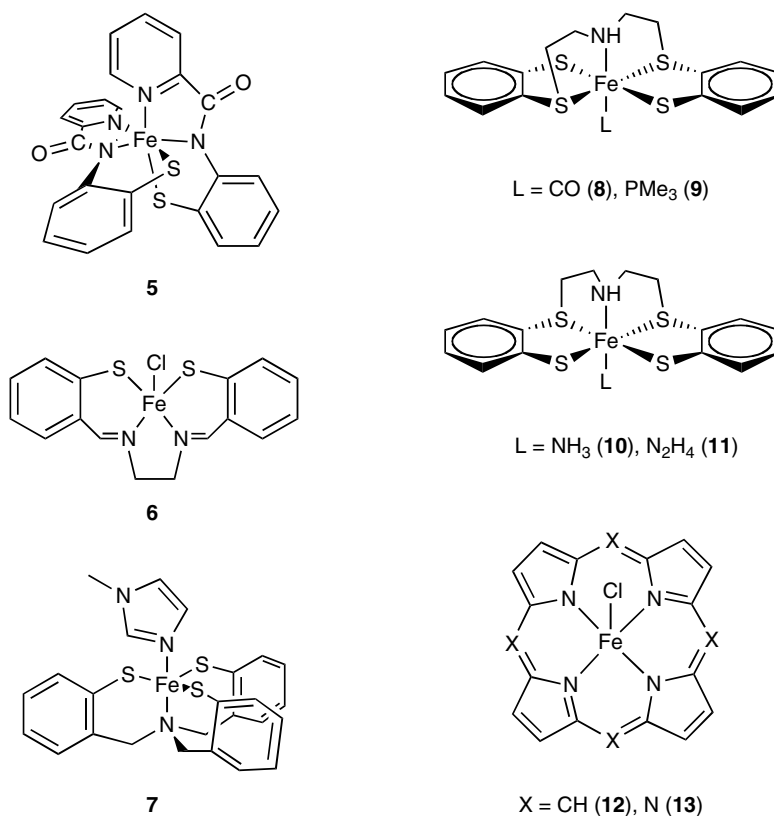


Figure 19.3 Iron complexes used in validation studies of spin-states.

In the past we have focused on several transition-metal complexes where the spin ground-state is known experimentally, and determined several other molecular properties as well. Initially [24], the vertical spin-state splittings were investigated for several Fe(II) and Fe(III) complexes (5–11, see Figure 19.3) and several interesting features observed. First, early GGA functionals and LDA tend to overestimate the stability of low-spin states and disfavor higher spin states [29]. More recent GGA functionals that include the OPTX functional provided the correct spin-ground states, which we attributed [24] to the inclusion of s^4 -terms [s is a reduced density gradient: $s = |\nabla \rho| / (2 \cdot \rho \cdot k_F)$, $k_F = (3\pi^2 \cdot \rho)^{1/3}$] in the parameterization of this functional. In addition, several MGGA and hybrid functionals gave the correct vertical spin-state splittings. The success of these *vertical* splittings does not, however, necessarily mean that all of these functionals correctly provide the *relaxed* (adiabatic) splittings. This became evident in subsequent studies [25], where the energy of each spin-state was determined at its own optimized geometry (right-hand side of Figure 19.1). Because of the population of the anti-bonding $d_{x^2-y^2}$ orbital in high spin states, which leads to larger metal–ligand distances, the structure relaxation resulted in the largest changes for the high spin state. Remarkably, out of the set of about ten functionals

that could be trusted for vertical splittings only one remained (OPBE) for relaxed splittings. The others, including the recent X3LYP and M06 functionals, failed for at least one of the complexes. Moreover, a clear separation could be noticed between early GGA and LDA functionals, which disfavor high spin-states, and hybrid functionals that disfavor low spin-states. The failure of B3LYP for low spin states had been noticed before – not just for iron complexes [25, 26, 76] but also for systems including other transition metals like manganese and nickel [27, 28]. Interestingly, despite its simplicity the OPBE functional turned out to be able to provide the spin-ground state for all iron complexes studied.

The over-stabilization of high spin-states by hybrid functionals like B3LYP is easily understood as coming from the inclusion of a portion of Hartree–Fock exchange [29]. The tendency of Hartree–Fock itself to favor high spin-states results directly from the absence in this method of (favorable) electron-correlation between unlike spins. This leaves only the (favorable) electron-correlation (exchange) between like spins. Since the number of exchange interactions is much larger in high spin-states than in low spin-states (10 versus 4 for a d^5 system) [29], the high-spin is favored. Reiher and coworkers noticed this as well and varied the amount of HF exchange in the B3LYP functional [76]. By lowering it from 20% to 15% (to give B3LYP*), they found for several iron compounds a significant improvement over B3LYP [26]. However, the amount of HF exchange needed was later found to be not universal; for some complexes it should be lowered even further to 12% [76]. This is an undesirable situation since one does not know *a priori*, or in fact *a posteriori*, if the amount of HF exchange is suitable for the particular transition-metal complex under study.

The reliability and robustness of the OPBE functional for spin-state splittings has been confirmed in a recent study [25]. In that study, several benchmark systems and challenging iron-complexes (14–19, see Figure 19.4) were investigated with the OPBE functional. For all these systems, the method works excellently. The singlet–quintet splittings from OPBE were very similar to reference benchmark data, which had previously been obtained at the CASPT2 level by Pierloot and coworker [77]. The mean absolute deviation (MAD) between the two methods was only 1–2 kcal mol^{−1} [25], which is an order of magnitude smaller than the deviation of other functionals such as BP86 (MAD = 15 kcal mol^{−1}), B3LYP (MAD = 11 kcal mol^{−1}) or PBE0 (MAD = 9 kcal mol^{−1}) [77]. However, even though the MAD of B3LYP and PBE0 is smaller than that of BP86 they failed to predict the correct spin ground-state of the bipyridyl complex 16. In contrast, the BP86 functional did predict the correct spin ground-state of the three benchmark complexes 14–16, albeit with an elevated MAD.

The benchmark systems also clearly identified the problem of Hartree–Fock with spin-states, as it predicted for all three systems a high-spin ground-state with the low-spin state at very high energy (70–80 kcal mol^{−1}).

The reliability of OPBE was further shown for the challenging iron complexes 17–19 [25], which included the spin-crossover compound 17, for which Reiher showed that B3LYP and B3LYP* failed. Also included were two complexes with pyridylmethylaniline (pma) ligands, which are structurally very similar, yet display different spin ground-states experimentally [78]. The only difference between

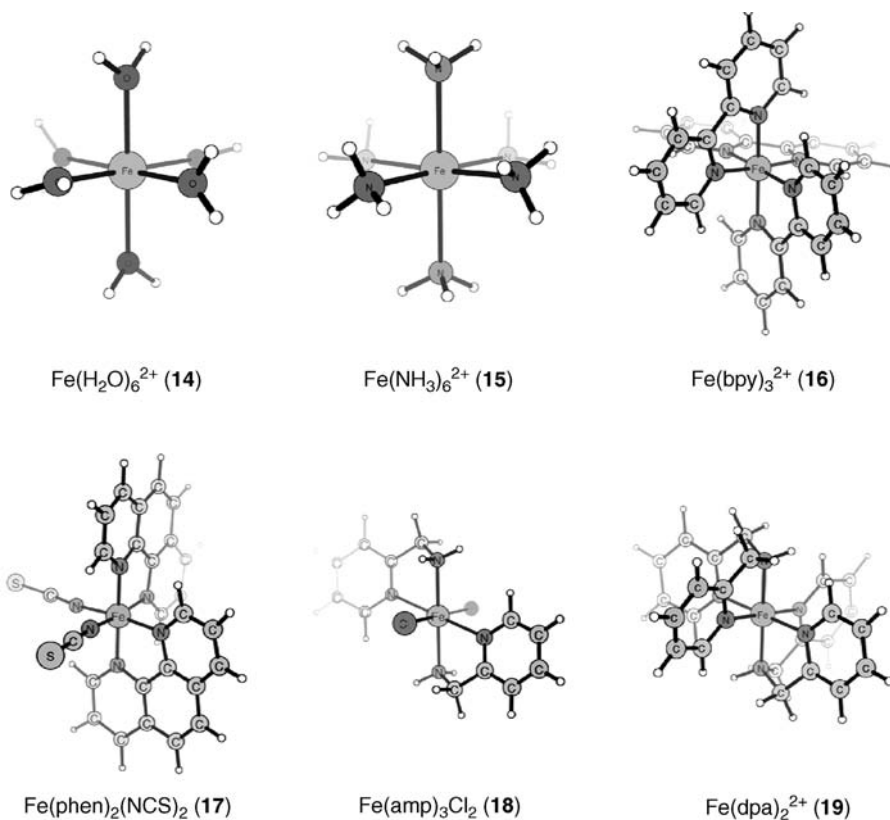


Figure 19.4 Benchmark and challenging iron complexes.

these mono-pma [18, $\text{Fe}(\text{amp})_2\text{Cl}_2$] and di-pma [19, $\text{Fe}(\text{dpa})_2^{2+}$] complexes is that in the former two chlorides occupy the axial positions, while in the latter these positions are taken by pyridines (Figure 19.4). The OPBE functional correctly predicted the spin ground-state for both complexes, and was again the only functional able to do so. By now unsurprisingly, the early GGA and LDA functionals failed for the high-spin complex while the hybrid functionals failed for the low-spin complex.

In subsequent studies, we have applied the OPBE functional to several typical systems that consist of transition-metal complexes with pyrazolylborate/pyrazolylmethane [79] and triazacyclononane ligands. In the first study, we examined the spin ground-state of the isolated complexes of these spin-crossover compounds and how they are influenced by substituent patterns. Since the spin-crossover phenomena is regarded as a cooperative effect of intermolecular interactions between a large number of (replicated) complexes, our study enables the separation of the spin-crossover phenomena from the intrinsic spin-state properties of the isolated transition-metal complex. Interestingly, the neutral pyrazolylmethane ligands provided a ligand environment that was not really different from that of the negatively charged pyrazolylborate ligand. Moreover, the substituent effect on the spin ground-state at

low temperatures was correctly reproduced by our OPBE data. In the second study, we investigated transition-metal complexes with the cyclononane ligand and how these are affected by different redox states of the metal ions. In all cases, OPBE predicted the correct spin ground-state and spectroscopic properties, such as infrared frequencies, that are in accordance with experimental data.

19.6

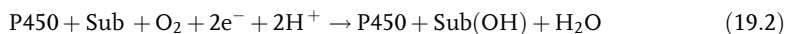
Spin-States Involved in Catalytic Cycles

Often when a transition metal is present in the active site of a metalloenzyme, different spin-states may be involved in the reaction mechanism. This has, for example, been shown by Siegbahn and coworker for extradiol dioxygenases [80], by Shaik and coworkers for two-state reactivity in P450 enzymes [81], and by Hillier and coworkers [82] and De Gioia and coworkers [28, 83] for NiFe hydrogenase. In the following sections, several catalytic mechanisms are discussed where spin-states and the methodology used play a role.

19.6.1

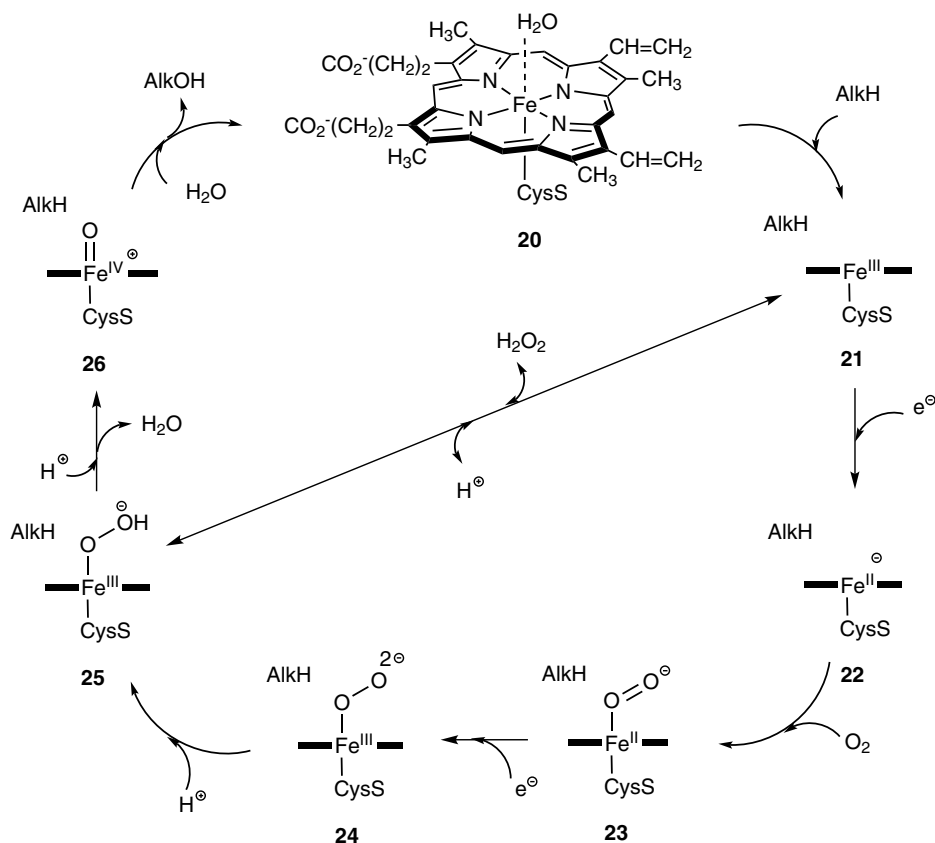
Cytochrome P450cam

The P450cam enzyme catalyzes regio- and stereospecifically the hydroxylation of camphor [10, 84, 85]:



Scheme 19.2 gives the consensus mechanism for this hydroxylation reaction [4, 86–91]. Starting from the resting state **20**, the substrate enters the active site (**21**) during which the first spin-flip takes place from doublet (**20**) to sextet (**21**). Through both model systems [70] and complete enzyme systems [92], it has been shown that this spin-flip results directly from the change of axial ligand at iron [93]; that is, in the resting state **20** a water molecule (or perhaps a hydroxide [70]) is present while after substrate binding (**21**) the axial position is empty.

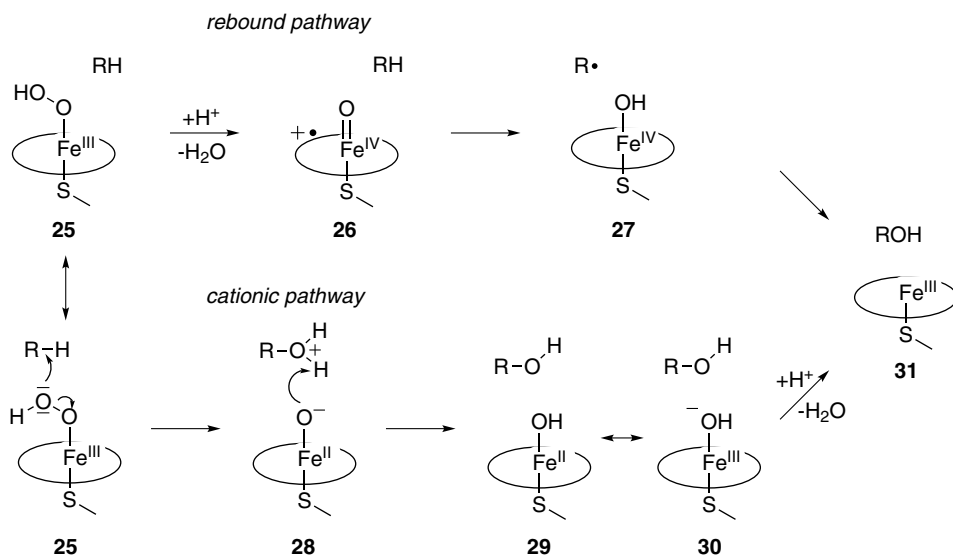
The substrate does not bind directly to the heme group but instead is anchored to a vital Tyr96 residue, which prepares the substrate for the subsequent steps in the reaction [12, 14, 94]. After the first electron reduction by putidaredoxin [16] (**22**), dioxygen enters (**23**) and a second spin-flip is observed. The second electron reduction and protonation of the distal oxygen leads to what is called “compound 0” (**25**, cpd0), which corresponds to the last experimentally observed step in the mechanism. Starting from this intermediate, there are two pathways to product formation, involving either the “rebound” pathway or a cationic pathway (Scheme 19.3) [86]. In the former pathway, another protonation of the distal oxygen and subsequent leaving of the formed water molecule leads to “compound I” (**26**, cpdI), the presumed active oxidant of the catalytic cycle that has been the subject of many theoretical studies.



Scheme 19.2 Catalytic cycle of cytochrome P450cam.

Much less attention has been given to the cationic pathway, or other alternatives such as the “somersault” pathway [95, 96]. The latter pathway involves a hydroxyl radical ($-\text{FeO} \cdots \text{HO}^\bullet$, cpdII) that may convert into water on abstraction of a hydrogen atom from the substrate. Furthermore, most studies start from cpdI without taking into account whether it would be formed at all from cpd0. Ehlers and coworkers recently [96] performed a study using both model systems and QM/MM calculations, using OPBE with STO basis sets. Their study indicated that the formation of cpdI is actually endothermic for cytochrome P450 [96]. This was in stark contrast to the formation of cpdI in peroxidase and catalase, which was found to be exothermic. Interestingly, this concurs with the fact that cpdI is observed in the latter two enzymes, but not in P450s. Moreover, the $\text{cpd0} \rightarrow \text{cpdI}$ process in P450 had a barrier of formation that was much larger than the $\text{cpdI} \rightarrow \text{cpd0}$ back reaction.

Apart from the complications due the competition between different pathways, the situation is even more complex because most studies on P450 enzymes used the B3LYP functional, often with an ECPB or non-modified Pople basis set. Both factors are not particularly favorable for consistent and reliable results when spin-states are



Scheme 19.3 Rebound versus cationic pathway.

involved. Yoshizawa and coworker reported in 2003 a study [52] on the rebound mechanism and reported transition states for both the hydrogen-abstraction step ($26 \rightarrow 27$ in Scheme 19.3) and the subsequent rebound step ($27 \rightarrow 31$). They studied both the doublet and quartet pathways at the B3LYP level and found that the hydrogen abstraction step is rate determining. However, they predicted that the quartet pathway was substantially favored over the doublet pathway even though the latter is the more logical spin-state; for instance, intermediate **23** has a low-spin singlet ground state and cpd0 (**25**) a doublet ground state. Subsequent single-point calculations at the OPBE/TZP level indeed gave a different picture, with the doublet pathway substantially more favorable.

The influence of the enzyme environment is not limited to just providing the anchoring Tyr96 residue, but seems to be much more subtle [5]. Representation of the axial cysteine ligand as either thiolate (SH^-), methylthiolate (CH_3S^-) or ethylthiolate ($\text{C}_2\text{H}_5\text{S}^-$) reportedly influences the spin distribution in the iron–sulfur bond, which is coupled to the spin distribution in the porphyrin ring [10, 97]. Moreover, the stability of any radical character on the sulfur seems to be regulated by three hydrogen-bonds from backbone amide NH hydrogens towards the sulfur [81, 98]. In the absence of these H-bonds, substantial radical character may be observed on the sulfur [9] that disappears with the introduction of the hydrogen bonds [92, 99].

The role of the heme propionate groups is another controversial subject where the protein environment may play a significant role. Guallar and coworkers have reported that these peripheral heme-substituents may control the hydrogen-abstraction chemistry in P450 enzymes [99–101], which is in line with the experimental finding that substrate binding perturbs the hydrogen bonding between one propionate and

the protein matrix [102]. Shielding of the propionates by an aspartate residue (Asp297) was implicated in regulation of the electron delivery to the iron center and the charge delocalization in the active site. This seems to be supported by crystal structures obtained at cryogenic temperatures [103], where only one “active” chain carries an oxygen ligand to iron. The other “inactive” chain does not show an oxygen, neither in the ferrous dioxy intermediate nor in the putative cpdI structure. Interestingly, the Asp297–propionate distance is substantially different in the two chains, with values of 2.36 Å in the “active” chain and 2.73 Å in the “inactive” chain. In their modeling studies [100], Guallar and coworkers observed a short aspartate–propionate distance (2.32 Å) *only* with the aspartate deprotonated ($-\text{COO}^-$) and when included within the QM region. With the aspartate residue protonated ($-\text{COOH}$) substantially longer distances were observed, both when it is included as MM (2.78 Å) or QM (2.80 Å). When the deprotonated aspartate ($-\text{COO}^-$) is placed in the MM region, the residue moves far from the propionates (3.33 Å) due to electrostatic repulsion, and rotates its orientation. So the question arises as to how it can be that such a short O–O distance is observed in the crystal structure and within the modeling studies of deprotonated Asp297 in the QM region. It is caused by the Asp297 and propionates carrying a substantial amount of radical character [99, 100]. This spin delocalization occurs where both the propionate and Asp297 donate electron density to the porphyrin π -cation, leaving an unpaired electron in the carboxylate–carboxylate contact. The importance of radical character localized on the propionates and aspartate was subsequently dismissed by Harvey, Mulholland and coworkers [104] as being an artifact from using a non-equilibrated crystal structure. However, in the latter study the hydrogen bonds towards sulfur were not included in the QM region and neither was the Asp297, which may have influenced their findings. Furthermore, the conclusions were based on structures obtained in molecular dynamics simulations where Asp297 was treated at the MM level. As Guallar and coworkers showed, MM is unable to correctly describe the interaction of the deprotonated Asp297 with the propionate groups of a heme. For the deprotonated aspartate (at the MM level), Harvey and Mulholland [104] observed spin-density on the propionates of about 0.2e, of similar magnitude as Guallar observed. Nevertheless, this radical character did not seem to influence the catalytic activity, something that Guallar and Thiel also observed [105]. Therefore, the presence and importance of radical character on the propionates and Asp297 remains unclear, but might be implicated in the electron transfer pathway from the putidaredoxin to the iron center [100]. Nevertheless, especially in view of the different competitive (rebound, cationic, agostic, somersault) pathways that are proposed for the catalytic activity of P450 enzymes, more detailed explorations with appropriate descriptions of these subtle effects are needed before a final conclusion can be reached.

19.6.2

His-Porphyrin Models

In contrast to the cytochrome P450 enzymes, where iron has cysteine as fifth axial ligand, there are also numerous iron proteins with histidine as fifth ligand. These

proteins include hemoglobin, myoglobin, cytochrome *c* and peroxidases. Computational studies have focused mainly on model systems with a porphyrin ring instead of the full heme group and an imidazole group instead of histidine. This reduction of the system is probably of no influence for the spin ground-state since crystal structures [106] of similar model compounds (albeit with peripheral substituents on porphyrin) indicated a high-spin quintet ground-state, similar to the iron proteins. Surprisingly, DFT methods predicted instead a triplet ground-state – even the B3LYP method that normally over-stabilizes high-spin states. Rovira and coworkers [107] have shown an energy difference of about $6.5 \text{ kcal mol}^{-1}$ between the triplet and quintet. Moreover, they showed that this energy splitting strongly depends on how far the iron is displaced out of the plane of its four coordinating nitrogen atoms. This out-of-plane distance might be dictated by the peripheral substituents on the experimental model compounds, as well as on the heme group. Therefore, we start by comparing with *ab initio* data obtained by Harvey and coworker [108] and Ghosh and coworkers [109, 110], and then discuss some other model compounds.

19.6.2.1 Reference Data (Harvey)

Strickland and Harvey [108] investigated an iron-porphyrin-imidazole (FePorIm) system with several DFT functionals, and to gain insight in the spin-state splittings they also investigated small model systems for FePorIm. In these latter Fe(II) model systems (**32** and **34**, Figure 19.5) the porphyrin ring was replaced by two amidines, and imidazole by water. As such, C_{2v} symmetry could be used and single-point coupled cluster calculations on the B3LYP optimized geometries were still feasible.

For the coupled cluster calculations, several correlation-consistent basis sets were used while, instead, in the DFT calculations an ECPB basis on iron (LACV3P) was combined with a Pople basis (6-311G*) on the rest (Table 19.4). We took the same B3LYP geometries and performed a series of DFT single-point calculations with the same correlation-consistent basis sets as were used in the coupled cluster calculations. For the DFT functionals that were used in the original paper, we see several striking differences. First, at first sight it seems that the singlet state of the second model (**34**) is placed at significantly higher energy with the hybrid functionals when using the better basis sets. However, it turns out that this results only from them

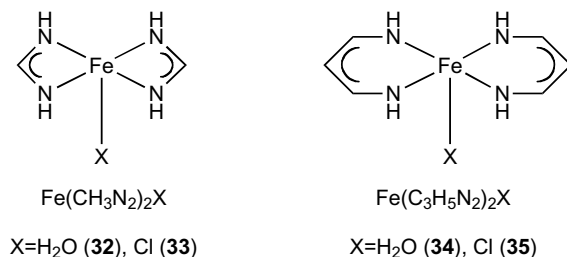


Figure 19.5 Model iron compounds by Harvey (**32** and **34**) [108] and Ghosh (**33** and **35**) [109, 110].

Table 19.4 Spin-state splittings (kcal mol⁻¹) for FePorIm model systems **32** and **34**.

Method	Model 32			Model 34			MAD ^{a)}
	Singlet	Triplet	Quintet	Singlet	Triplet	Quintet	
LACV3P(Fe), 6-311G*(amid.)							
BP86 ^{b)}	23.6	0.9	0	2.8	−20.5	0	
BLYP ^{b)}	21.2	1.1	0	3.3	−17.9	0	
B3PW91 ^{b)}	37.8	13.6	0	20.8	−4.2	0	
B3LYP ^{b)}	31.3	10.7	0	16.1	−5.6	0	
cc-pVTZ(Fe):cc-pVDZ(amid.)							
CCSD(T) ^{b)}	42.7	20.7	0	25.9	3.7	0	
BP86 ^{c)}	22.4	−3.6	0	3.3	−23.5	0	23.6
BP ^{d)}	25.0	−2.0	0	5.9	−21.9	0	21.5
B3LYP	32.2	7.3	0	46.2 _{19.0} ^{e)}	−7.4	0	13.8
B3LYP*	29.1	4.4	0	38.9 _{14.5} ^{e)}	−11.5	0	14.5
B3PW91	32.2	6.2	0	44.4 _{16.6} ^{e)}	−10.4	0	14.4
M05	53.3	23.7	0	70.0 _{41.0} ^{e)}	7.6	0	15.4
M06-2X	51.3	23.9	0	74.6 _{44.0} ^{e)}	15.4	0	18.1
M06-L	33.9	11.8	0	23.0	−2.7	0	6.8
OPBE	41.1	8.0	0	21.7	−12.4	0	8.7
TPSS	21.8	−4.2	0	4.3	−22.0	0	23.3
X3LYP	33.2	8.1	0	47.8 _{19.9} ^{e)}	−6.4	0	13.5
s6-31G*							
BP ^{d)}	22.4	−3.5	0	3.5	−23.7	0	
B3LYP	29.3	5.4	0	45.0 _{16.2} ^{e)}	−9.6	0	
B3LYP*	26.3	2.5	0	37.9 _{11.8} ^{e)}	−13.7	0	
B3PW91	29.1	4.2	0	42.9 _{13.6} ^{e)}	−12.7	0	
M06-L	29.5	9.3	0	18.5	−5.6	0	
OPBE	38.0	6.3	0	18.5	−14.4	0	
TZP post-SCF							
BP	23.1	−2.6	0	5.3	−21.9	0	
B3LYP	35.2	10.0	0	23.7	−4.2	0	
B3LYP*	30.6	6.1	0	17.6	−9.3	0	
M06-L	33.2	12.6	0	25.5	−0.5	0	
OPBE	41.4	8.7	0	22.9	−11.4	0	

a) Mean absolute deviation from CCSD(T) data.

b) From Reference [108].

c) Perdew86 GGA-correlation with PZ81 LDA-correlation.

d) Perdew86 GGA-correlation with VWN LDA-correlation.

e) Subscript values obtained after restarting from OPBE.

having different orbital occupations than the pure functionals or indeed the ones obtained by Strickland and Harvey. Note that this occurs *only* for the hybrid functionals, not for the pure ones, and only for the larger of the two model systems. Performing the SCF first with the OPBE functional and then restarting from this with

the hybrid functionals does indeed give the lower energy, with the correct orbital occupations. Since the computational setup is the same for all functionals, and for both model systems, and the SCF converged rapidly and smoothly, this odd behavior cannot be attributed to the quantum-chemistry program used (NWChem). Instead, because with proper initial orbital occupations it does converge to the lower energy, it must somehow result from the initial guess that is poorly obtained from the hybrid functionals. Second, all DFT functionals predict a substantially lower energy for the triplet state than coupled cluster does. The same trend was shown by Harvey and Strickland, but with a different basis set used for the DFT functionals. Improving the basis set brings the triplet energies down, and the singlet energies up. Both model systems show that the OPBE functional significantly overestimates the stability of the triplet state, by 13–16 kcal mol⁻¹, compared to the CCSD(T) data with the same basis set. In contrast, for the singlet–quintet splittings OPBE gives the expected good results, with a deviation of only 1–4 kcal mol⁻¹. Although this over-stabilization of the triplet state is observed for most DFT functionals, the poor performance of OPBE is troubling and warrants further examination. One possible explanation might come from the similarity with the two-center three-electron (2c–3e) bond, for which OPBE also over-stabilizes, by approximately 10 kcal mol⁻¹. Nevertheless, the smallest deviation is still observed for the pure OPBE and M06-L functionals, followed by hybrid functionals and finally the early GGA functionals.

We also repeated the same calculations, but now with our spin-state-corrected s6-31G* basis set. The observed spin-state splittings differ by only up to 2 kcal mol⁻¹ with respect to the much more demanding correlation-consistent basis sets (Table 19.4). For instance, a typical single-point at cc-pVTZ(Fe):cc-pVDZ(amidines) takes about ten times as long as the same single-point at s6-31G*.

19.6.2.2 Reference Data (Ghosh)

Ghosh, Taylor and coworkers studied almost the same model system as Harvey, but with a chloride as axial ligand [109, 111]. They performed single-point CCSD(T) calculations on PW91 optimized geometries for the Fe(III) compound **33** [109], and single-point CASPT2 and CCSD(T) calculations on B3LYP optimized geometries for the Fe(II) compound **35** [111]. We followed the same procedure but instead calculated the energy for several DFT functionals. We used the s6-31G* basis both for obtaining the B3LYP geometry of **35**, and then used ADF with the TZP basis for the subsequent single-points; the geometry of **33** was obtained with ADF at PW91/TZP. Table 19.5 shows the results.

The geometry obtained with the s6-31G* is similar to that obtained by Ghosh and coworkers: we find Fe–N distances of 1.88/1.91 Å (Ghosh 1.90/1.93 Å) and Fe–Cl distances of 2.18/2.35 Å (Ghosh 2.20/2.35 Å); the PW91/TZP geometry was identical. The concordance of the geometry of **35** also shows up in the B3LYP triplet–quintet splittings, which are 4.1 kcal mol⁻¹, compared to 4.8 kcal mol⁻¹ by Ghosh and coworkers.

The early GGA functionals fail to indicate the spin ground-state, as already mentioned by Ghosh, while OPBE and several MGGA and hybrid functionals do show a high-spin ground-state for both **33** and **35**. It is unclear which of the reference

Table 19.5 Spin-state splittings for model compounds **33** and **35**.

	33			35		
	Doublet	Quartet	Sextet	Singlet	Triplet	Quintet
RCCSD(T) ^{a)}	—	17.1	0	—	6.0	0
CASPT2 ^{b)}	—	—	—	—	13.4 (15.9 ^{c)})	0
PW91 ^{a),b)}	—	−3.0	0	—	~0	0
B3LYP ^{b)}	—	—	—	—	4.8	0
TZP						
B3LYP	27.1	7.5	0	31.4	4.1	0
B3LYP*	22.3	4.7	0	27.6	1.8	0
BLYP	10.9	−1.4	0	17.5	−3.4	0
BP	11.2	−1.9	0	19.0	−3.5	0
M06-L	35.8	13.5	0	36.3	10.5	0
M06	46.3	14.7	0	45.3	13.5	0
OPBE	26.2	6.8	0	28.3	2.6	0
TPSS	7.7	−3.0	0	17.1	−4.3	0
X3LYP	28.3	8.1	0	32.6	4.6	0

a) From Reference [109].

b) From Reference [110].

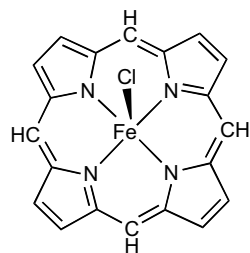
c) 3s3p correlation included in CASPT2.

values is correct, as CCSD(T) and CASPT2 report energy differences of about 10 kcal mol^{−1}. Therefore, nothing can be said about the appropriateness of the different density functionals.

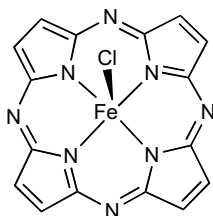
19.6.2.3 Other Model Systems

Another model system that has been used previously is iron-porphyrin with a chloride ligand, as it has a high-spin ground-state just like FePorIm. Moreover, several studies focused on the difference between iron-porphyrin-chloride (**36**) and iron-porphyrazine-chloride (**37**) [109, 112]. Even though the only difference between the two is the replacement of the meso carbon with a nitrogen (Figure 19.6), the spin ground-state changes from a sextet for **36** to a quartet for **37**. Studies using early GGA functionals confirmed the spin ground-state for **37**, but the sextet ground-state for **36** could not be reproduced [112]. It was argued that the presence of peripheral substituents might influence the spin-state splitting in favor of the sextet state. Nevertheless, the octa-ethyl analog still showed a quartet ground-state [112]. Moreover, a more recent study at the OPBE/TZP level did provide the correct spin ground-state for both compounds without any peripheral substituents.

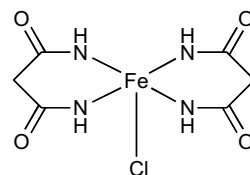
We have also studied several other model systems such as the Collins complex (**38**, Figure 19.6) and the model compounds **32–35**, but with different axial ligands. Moreover, we explored how the spin-state splittings change if we go from the small model systems for FePorIm to FePorIm and then to the experimentally observed system Fe[(TpOMePP)(1,2-diMe-Im)] (Figure 19.7).



Fe-porphyrin (36)



Fe-porphyrizin (37)



Collins' Fe(IV)-complex (38)

Figure 19.6 Iron-porphyrin (36) and iron-porphyrizin (37) with an axial chloride ligand and Collins' Fe(IV)-complex (38).

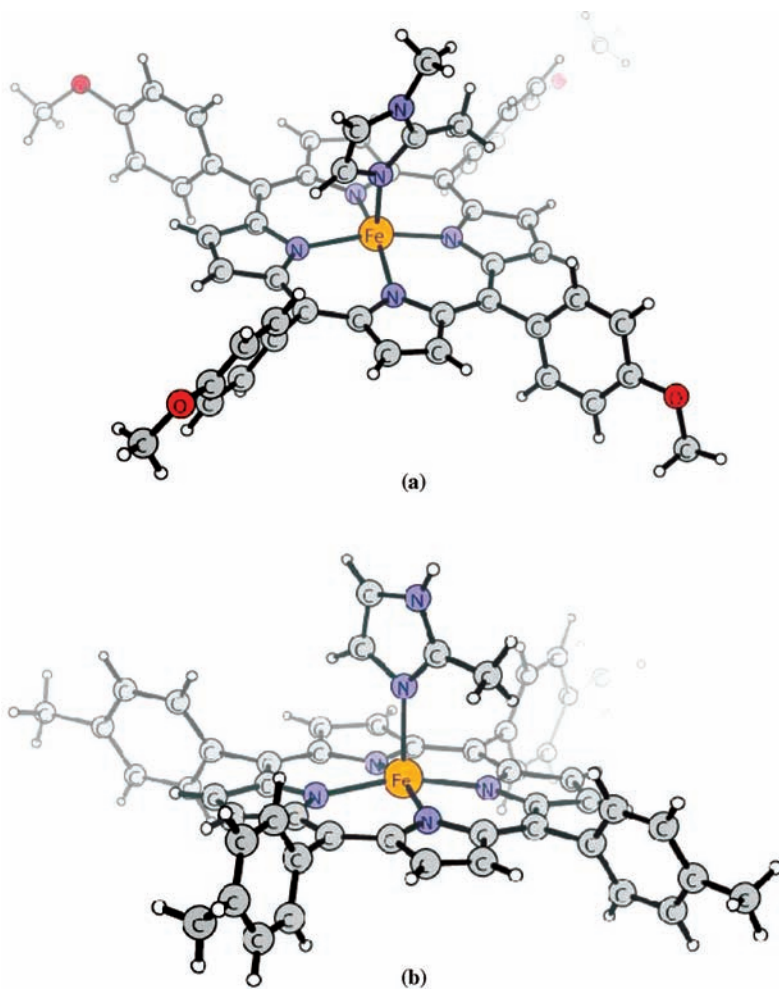


Figure 19.7 Experimentally-observed [106] FePorIm complexes: (a) Fe[tetra(*para*-OMe-phenyl)porphyrin(1,2-diMe-Im)] (TALLEY) and (b) Fe[(tetraphenyl-porphyrin)(2-MeHIm)] (TALLOI).

These results will be described in detail elsewhere, and we will in focus here only on the series of FePorIm models, from the small amidine complexes (**33**, **35**) to Fe [(TpOMePP)(1,2-diMe-Im)]. Table 19.6 gives in-plane iron–nitrogen and the iron–axial ligand distances as obtained with OPBE/TZP calculations. All three possible spin-states (low, intermediate and high) were considered.

The iron–ligand distances showed the usual trend of about 2.0 Å for the in-plane Fe–N in the low and intermediate spin-state, and about 2.1 Å in the high-spin; the axial Fe–N distances range from about 1.9 Å in the low-spin to about 2.3 Å in intermediate and about 2.2 Å in high-spin. For the large systems, the Fe–imidazole distance is slightly larger in the triplet state for the tetraphenylporphyrin (TALLOI), where it is found at 2.5 Å. For the tetra(*para*-OMe-phenyl)porphyrin complex in the triplet state, the imidazole moves away from iron and remains near the periphery of the porphyrin ring. This also happens for the triplet state of **33** with imidazole as axial ligand. It does not influence the stability, as it is seen that the triplet states remain lowest in energy with the quintets (slightly) higher. Note that this might be an artifact of DFT methods, which for the small model compounds of Strickland and Harvey (Section 19.6.2.1) overestimated the stability of the triplet state. Nevertheless, along

Table 19.6 Fe–ligand distances (Å) and spin-state splittings (ΔE , kcal mol^{−1}) for FePorIm models, obtained at OPBE/TZP.

	Fe–L ^{a)}			ΔE		
	Low	Intm.	High	Low	Intm.	High
cpd 33						
X = Cl ^{−b)}	1.94/2.19	1.98/2.25	2.09/2.19	29.0	7.0	0
X = Im ^{c)}	1.98/1.84	1.97/2.27	2.13/2.17	27.6	12.5	0
cpd 35						
X = Cl ^{− b)}	1.88/2.19	1.94/2.29	2.09/2.22	5.7	0	3.3
X = Im ^{c)}	1.90/1.87	1.89/6.47 ^{d)}	2.09/2.25	22.3	0	14.6
cpd 36						
X = Cl ^{− b),e)}	1.99/2.13	2.02/2.24	2.11/2.20	18.3	3.9	0
X = Im ^{c)}	1.98/1.89	2.00/2.33	2.08/2.20	6.0	0	1.0
Exptl system						
TALLEY ^{c),f)}	1.95/1.93	1.96/7.46 ^{g)}	2.09/2.27	6.9	−0.4	0
TALLOI ^{c),h)}	1.97/1.93	1.99/2.46	2.09/2.23	8.6	−1.2	0

a) Indicated are average in-plane/axial iron–ligand distances.

b) Fe(III)-complex.

c) Fe(II)-complex.

d) Imidazole has moved away from iron and hydrogen-bonds to two NH groups.

e) From Reference.

f) Fe(Tp-OMePP)(1,2-Me₂Im).

g) Imidazole has moved away to the periphery of the porphyrin ring.

h) Fe(TPP)(2-MeHIm).

the series of FePorIm systems, we see a gradual convergence of the spin-state splittings, which are for the largest systems observed in a narrow range of less than 10 kcal mol^{-1} .

19.6.3

NiFe Hydrogenase

Nickel is, like iron, an essential trace element for sustaining life and is found in many enzymes [113]. Among such enzymes, NiFe hydrogenase stands out because it has a very peculiar active site where both an iron and a nickel are present [114]. The two metals present in the active site are bridged by two cysteine residues; nickel has two more cysteine residues as ligand while the coordination sphere around iron is completed by cyanide and carbon-monoxide groups. These latter groups are unusual for biologic systems, and may be needed to keep the Fe(II) in its low-spin singlet state. The redox state of nickel undergoes a Ni(II)/Ni(III) change during catalysis. The enzyme catalyzes the (reversible) cleavage of a hydrogen molecule:



The enzyme catalyzes the reversible oxidation of a hydrogen molecule (H_2), but exactly how is a matter of dispute. There seems to be a general consensus that the Ni(III) species has a low-spin doublet state, but the ground-state of the Ni(II) species may either be a singlet or triplet. From model compounds it is well known that four-coordinated Ni(II)-complexes can have both spin states, depending on the properties of the ligands. De Gioia and coworkers therefore focused first on two smaller nickel complexes before turning to the complete enzyme [28, 83]. They investigated a low-spin complex $\text{Ni}(\text{edt})_2^{2-}$ (**4**) and an high-spin complex $\text{Ni}(\text{SPh})_4^{2-}$ (**39**) (Figure 19.8). Note that “high-spin” for nickel complexes corresponds to the triplet state, and not a quintet state as for iron.

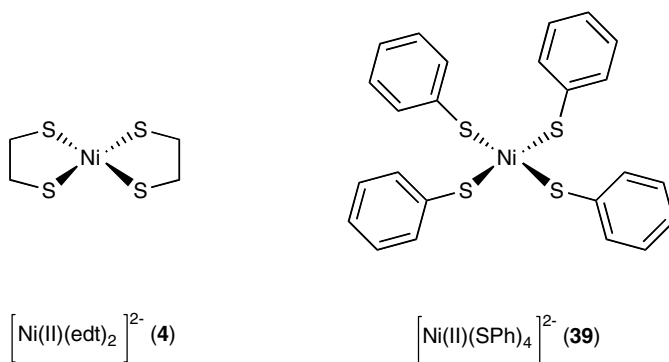


Figure 19.8 NiFe hydrogenase model compounds [28] $\text{Ni}(\text{edt})_2^{2-}$ (**4**) and $\text{Ni}(\text{SPh})_4^{2-}$ (**39**).

Table 19.7 Nickel–ligand distances (Å) and spin-state splittings (kcal mol^{−1}) for NiFe hydrogenase model compounds **4** and **39**.

	Average Ni–ligand 4		Energy 4		Average Ni–ligand 39		Energy 39	
	Singlet	Triplet	Singlet	Triplet	Singlet	Triplet	Singlet	Triplet
Exptl	2.195	—	l.s.	—	—	2.292	—	h.s.
TZVP basis								
B3LYP ^{a)}	2.288	2.386	0	−1.0	2.300	2.379	15.3	0
B3LYP ^{*a)}	2.280	2.373	0	2.7	2.290	2.363	11.1	0
BP86 ^{a)}	2.250	2.320	0	9.8	2.258	2.308	3.5	0
s6-31G*								
B3LYP	2.276	2.373	0	0.9	— ^{b)}	— ^{b)}	— ^{b)}	— ^{b)}
B3LYP*	2.269	2.360	0	4.4	— ^{b)}	— ^{b)}	— ^{b)}	— ^{b)}
BP86	2.245	2.313	0	11.3	— ^{b)}	— ^{b)}	— ^{b)}	— ^{b)}
BP	2.243	2.312	0	10.9	— ^{b)}	— ^{b)}	— ^{b)}	— ^{b)}
M06-L	2.259	2.353	0	2.2	— ^{b)}	— ^{b)}	— ^{b)}	— ^{b)}
OPBE	2.211	2.282	0	10.6	— ^{b)}	— ^{b)}	— ^{b)}	— ^{b)}
TZP								
BP	2.251	2.326	0	9.8	2.254	2.307	2.5	0
OPBE	2.215	2.298	0	9.0	2.229	2.296	4.3	0

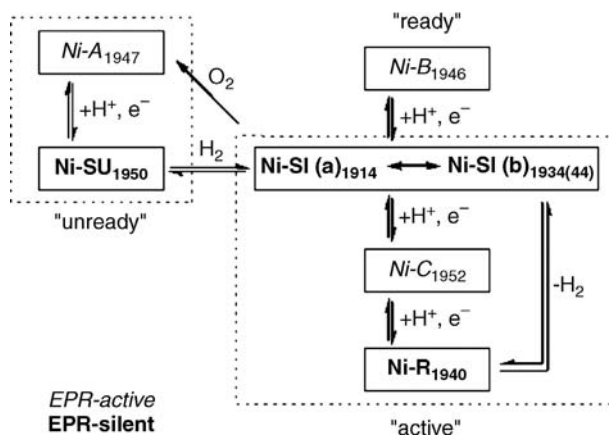
a) From Reference [28].

b) Not attempted due to system size.

In Table 19.7 we report the nickel–ligand distances and spin-state splittings for Ni(edt)₂^{2−} (**4**) and Ni(SPh)₄^{2−} (**39**) as obtained with several DFT functionals and basis sets. The best agreement with the experimental structure was observed for the OPBE functional, followed by BP86. These pure DFT functionals also provided the correct spin ground-state, with a clear singlet–triplet (S–T) splitting of some 10 kcal mol^{−1}. In contrast, the B3LYP functional wrongly predicts it to have a triplet ground-state [28]. The spin ground-state of the triplet complex Ni(SPh)₄^{2−} (**39**) was correctly predicted by all four functionals with a S–T splitting of about 3.5 kcal mol^{−1} for the pure functionals and 11–15 kcal mol^{−1} for the hybrid functionals.

For both nickel-complexes, the hybrid functionals B3LYP and B3LYP* show nickel–ligand distances that are too long by about 0.09–0.10 Å. The pure functionals perform much better, with deviations of 0.01–0.02 Å (OPBE) and 0.02–0.05 Å (BP86). Interestingly, when carrying out single-point calculations on the OPBE optimized geometries, the B3LYP functional gives the correct spin ground-state (albeit with a very small S–T splitting of about 1 kcal mol^{−1}).

Several intermediates of the catalytic cycle of NiFe hydrogenase have been observed; some of these are EPR-active, others EPR-silent (Scheme 19.4) [114, 115]. The former are thought to result from Ni(III), the latter from Ni(II), although



Scheme 19.4 Proposed catalytic cycle for NiFe hydrogenase (subscript indicates the CO frequency).

the involvement of iron and/or thiyl radicals may not be ruled out completely [115]. It was shown that the alternating Ni-B, Ni-SI (silent), Ni-C and Ni-R (ready) species can be smoothly interconverted in redox titrations and that each species has a characteristic CO and CN stretching frequency. An attempt was made to quantify the number of electrons involved in the steps of the catalytic cycle, which suggested that (i) Ni-R is one electron more reduced than Ni-C, (ii) Ni-C is more reduced than Ni-B and (iii) Ni-C and Ni-B states are separated by at least one EPR silent state. Several proposals for the oxidation states of Ni in the “ready” and “active” species have been made, such as III(B)-II(SI)-I(C)-0(R), III(B)-II(SI)-I(C)-I(R) and III(B)-II(SI)-III(C)-II(R) [115]. Other proposals for the redox chemistry involve the ligands and/or the iron.

Because of the ambiguity of the redox states of the metals in the different intermediates, which are structurally not unambiguously identified, various structures have been proposed for the different “active” intermediates. Many of them are based on DFT calculations using (mainly) the BP86 and B3LYP functionals, and often with poor basis sets (such as ECPBs). The small NiFe hydrogenase model compounds showed (see above) that B3LYP cannot be trusted for either the geometry or the spin ground-state. Yet, it was used (with ECPBs) by Hall and coworkers to support the claim of the existence of high-spin Ni(II) in the catalytic cycle [116]. We revisited the original data, performed single-point calculations on their optimized geometries and found a completely different picture. Table 19.8 gives the spin-state splittings with several functionals.

The more accurate OPBE functional gives a clear separation of the singlet–triplet states (Table 19.8), indicating that these model systems cannot be observed in the high-spin state.

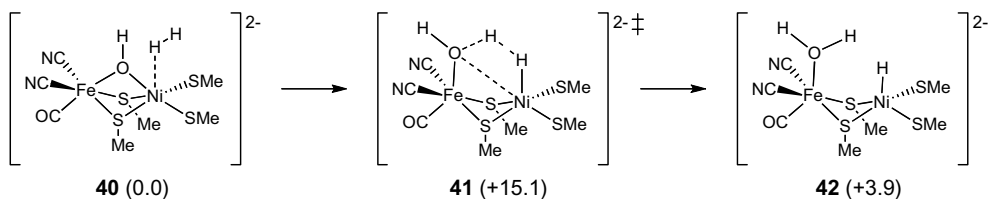
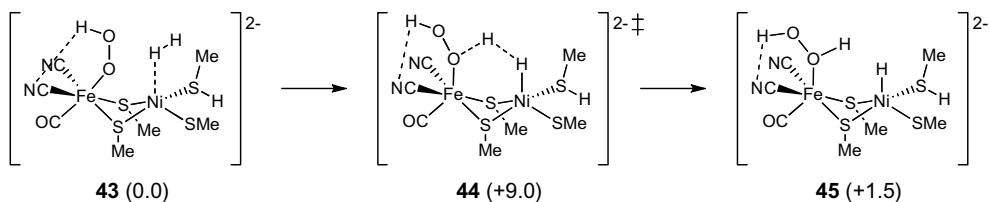
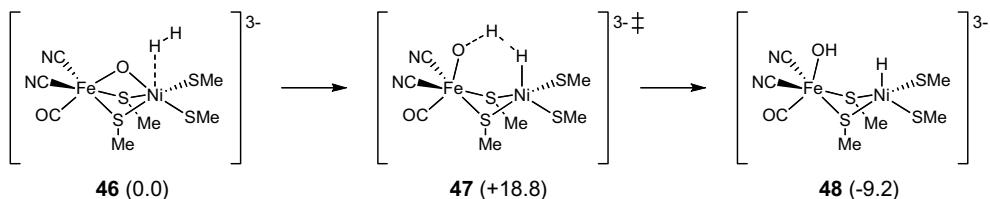
Hillier and coworkers have investigated several pathways for the formation of Ni-SI species from Ni-B and Ni-A; for the latter pathway two routes have been studied [82]. These pathways involved many steps; therefore, we focus here on the rate-determining steps of each of the three routes. Scheme 19.5 shows these three steps

Table 19.8 Spin-state splittings (kcal mol⁻¹) for model systems^{a)} of Ni-SI(a) and Ni-R.

	Ni-SI(a) Singlet	Ni-SI(a) Triplet	Ni-R Singlet	Ni-R Triplet
s6-31G* basis				
B3LYP	0	-0.9	0	1.8
OPBE	0	9.6	0	10.5
DZP ^{b)}				
B3LYP	0	-0.4	0	5.4
B3LYP*	0	3.0	0	7.3
BP	0	11.6	0	11.7
OPBE	0	8.8	0	10.8
M06	0	5.0	0	12.7
M06-L	0	4.3	0	5.1
X3LYP	0	-0.9	0	5.3

a) Geometries taken from Reference [116].

b) Post-SCF on OPBE/DZP orbitals/densities.

route a (Ni³⁺)*route b* (Ni²⁺)*route c* (Ni³⁺)**Scheme 19.5** Three routes studied by Hillier and coworkers [82].

together with the original energies obtained at the BP86/DGDZVP level. These three steps correspond to (see also Scheme 19.4) the formation of the Ni-SI species from Ni-B (route a) and the formation of Ni-SI from Ni-A (routes b and c). The rate-determining step in each route corresponds to the cleavage of the H-H bond of the hydrogen molecule.

Because of the possible involvement of high-spin Ni(II) in the catalytic cycle, we performed single-point calculations for both the low- and high-spin state for all three rate-determining steps. As shown in Table 19.9, the involvement of high-spin states in these three routes is unlikely. All DFT functionals predict lower energies for the low-spin state than for the high-spin. This is even the case for the hybrid functionals B3LYP, B3LYP*, and so on that tend to overestimate the stability of higher spin-states. In addition, note that the transition state of route b in fact is not a transition state in the high-spin state, but a minimum, that is, the corresponding “barrier” has a negative value (Table 19.9).

Of course, this does not mean that high-spin pathways are not involved in the other “active” species. Furthermore, the situation might be altered greatly by the presence of the protein matrix. However, structures 40–48 have a nickel that is mostly present with a tetrahedral coordination environment, and not square-planar. Note that this may be relevant for the enzyme-catalyzed reaction, where the nickel also shows a tetrahedral geometry in some (or perhaps all) steps.

Table 19.9 Spin-state splittings (kcal mol^{−1}) for three pathways studied by Hillier and coworkers.^{a)}

	40	41 [‡]	42	43	44 [‡]	45	46	47 [‡]	48
Ref. (ls)	0	15.1	3.9	0	9.0	1.5	0	18.8	−9.2
s6-31G*									
BP86 (ls)	0	16.8	2.4	0	7.2	−3.9	0	20.9	−10.7
BP (ls)	0	17.0	2.4	0	7.2	−3.9	0	21.0	−10.7
DZP ^{b)}									
BP (ls)	0	12.0	−2.5	0	5.7	−5.7	0	18.2	−15.4
BP (hs)	26.5	40.4	31.6	26.4	17.0	29.2	15.8	37.4	26.5
DZP ^{c)}									
B3LYP (ls)	0	21.6	−1.6	0	11.2	−5.6	0	25.2	−16.6
B3LYP (hs)	29.2	39.6	35.1	29.1	16.2	30.1	12.9	40.5	31.8
B3LYP* (ls)	0	18.9	−1.9	0	9.8	−5.7	0	23.2	−16.4
B3LYP* (hs)	28.7	39.9	34.2	28.8	16.8	30.0	13.8	39.8	30.5
M06-L (ls)	0	16.1	5.0	0	10.8	−2.2	0	24.0	−9.4
M06-L (hs)	25.4	38.4	39.4	24.3	15.4	30.8	12.3	38.4	33.7
OPBE (ls)	0	17.8	−0.9	0	8.3	−3.9	0	23.5	−13.9
OPBE (hs)	23.8	44.2	32.2	23.3	16.9	30.3	13.0	40.0	26.9

a) Geometries taken from Reference [82].

b) STO-DZP basis, self-consistent.

c) STO-DZP basis, post-SCF on BP/DZP orbitals/densities.

19.7

Concluding Remarks

Spin-state splittings of transition-metal complexes form an intriguing aspect of computational chemistry studies. Several aspects of how to obtain these splittings have been discussed that include basic tools like DFT functionals, basis sets and spin-contamination corrections. The second half of this chapter has focused on practical applications observed in recent studies. Based on the recommendations that resulted from comparisons with benchmark data, the results of these studies and the conclusions drawn from them have been discussed. On several occasions it was seen that the claims made in these studies were not substantiated by the results when reexamined with improved methods and/or basis sets.

Because of systematic failures of popular DFT functionals for a proper description of spin-state splittings, the straightforward application to challenging problems like mechanistic insights of metalloenzymes is difficult. This may be circumvented by using the combination of an appropriate functional (OPBE, M06-L) and basis set (STOs, s6-31G*).

19.8

Computational Details

The results presented in this chapter have been obtained by various quantum-chemistry programs. All calculations involving Slater-type orbitals [117] have been obtained using the Amsterdam Density Functional (ADF, version 2007.01) program [118], while those involving Gaussian-type orbitals were obtained by either Gaussian03 (revision B.02) [119], NWChem (version 5.1) [120] or Orca (2.6.4) [121]. Whenever possible, the calculations with the OPBE functional [35] were obtained using the combination of VWN [122] for LDA-correlation and PBEc [123] for GGA-correlation, as it is done in ADF. Recently [53], we showed that using either PW92 [124] or VWN for LDA-correlation is of no influence for spin-state splittings. Furthermore, a distinction is made between BP86 and BP, where again the difference is in the treatment of LDA correlation. Gaussian and NWChem by default use the Perdew-Zunger (PZ81) [125] LDA-correlation with Perdew86 GGA-correlation (BP86), while ADF uses the superior VWN-functional for LDA-correlation (BP). The latter combination gives better results in general. Apart from these subtle differences, the results with the GTO basis sets could have been obtained with any of the three programs, except for those concerning recent functionals like the M06 suite [44].

Many geometry optimizations have been carried out with the QUILD program [59], which functions as a wrapper around the ADF, Orca or MOPAC2009 programs. It uses adapted delocalized coordinates [126] for smoother optimizations, which significantly improve the optimization convergence [59]. Experience with the QUILD performance has been translated into modifications for the model startup-Hessian in NWChem, for which different force constants have been employed (bonds 0.40,

angles 0.18, torsion 0.01). The use of spin-contamination corrections [70, 127] is automated within QUILD for both the energy and the gradient, which has been used for some of the systems as mentioned in the text.

References

- Harvey, J.N., Poli, R., and Smith, K.M. (2003) *Coord. Chem. Rev.*, **238**–239, 347–361.
- Harvey, J.N. (2004) *Struct. Bond.*, **112**, 151–183.
- Ortiz de Montellano, P.R. (1995) *Cytochrome P450 Structure, Mechanism and Biochemistry*, Plenum, New York.
- Tyson, C.A., Lipscomb, J.D., and Gunsalus, I.C. (1972) *J. Biol. Chem.*, **247**, 5777–5784.
- Loew, G.H. and Harris, D.L. (2000) *Chem. Rev.*, **100**, 407–419.
- Harris, D. and Loew, G. (1993) *J. Am. Chem. Soc.*, **115**, 8775–8779.
- Poulos, T.L., Finzel, B.C., and Howard, A.J. (1986) *Biochem.*, **25**, 5314–5322.
- Poulos, T.L., Finzel, B.C., and Howard, A.J. (1987) *J. Mol. Biol.*, **195**, 687–700.
- Green, M.T. (2000) *J. Am. Chem. Soc.*, **120**, 10772–10773.
- Dawson, J.H. and Sono, M. (1987) *Chem. Rev.*, **87**, 1255–1276.
- Di Primo, C., Hui Bon Hoa, G., Douzou, P., and Sligar, S.G. (1990) *J. Biol. Chem.*, **265**, 5361–5363.
- Swart, M., Groenhof, A.R., Ehlers, A.W., and Lammertsma, K. (2005) *Chem. Phys. Lett.*, **403**, 35–41.
- Schlichting, I., Jung, C., and Schulz, H. (1997) *FEBS Lett.*, **415**, 253–257.
- Raag, R. and Poulos, T.L. (1991) *Biochemistry*, **30**, 2674–2684.
- Hintz, M.J., Mock, D.M., Peterson, L.L., Tuttle, K., and Peterson, J.A. (1982) *J. Biol. Chem.*, **257**, 14324–14332.
- Shimada, H., Nagano, S., Ariga, Y., Unno, M., Egawa, T., Hishiki, T., Ishimura, Y., Masuya, F., Obata, T., and Hori, H. (1999) *J. Biol. Chem.*, **274**, 9363–9369.
- Parr, R.G. and Yang, W. (1989) *Density Functional Theory of Atoms and Molecules*, Oxford University Press, New York.
- Koch, W. and Holthausen, M.C. (2000) *A Chemist's Guide to Density Functional Theory*, Wiley-VCH Verlag GmbH, Weinheim.
- Dreizler, R. and Gross, E. (1995) *Density Functional Theory*, Plenum Press, New York.
- Becke, A.D. (1993) *J. Chem. Phys.*, **98**, 5648–5652.
- Stephens, P.J., Devlin, F.J., Chabalowski, C.F., and Frisch, M.J. (1994) *J. Phys. Chem.*, **45**, 11623–11627.
- Becke, A.D. (1988) *Phys. Rev. A*, **38**, 3098–3100.
- Perdew, J.P. (1986) *Phys. Rev. B*, **33**, 8822–8824; Erratum: (1986) *Phys. Rev. B*, **34**, 7406.
- Swart, M., Groenhof, A.R., Ehlers, A.W., and Lammertsma, K. (2004) *J. Phys. Chem. A*, **108**, 5479–5483.
- Swart, M. (2008) *J. Chem. Theory Comput.*, **4**, 2057–2066.
- Reiher, M., Salomon, O., and Hess, B.A. (2001) *Theor. Chem. Acc.*, **107**, 48–55.
- Sprovier, E.M., Gascon, J.A., McEvoy, J.P., Brudvig, G.W., and Batista, V.S. (2006) *J. Inorg. Biochem.*, **100**, 786–800.
- Bruschi, M., De Gioia, L., Zampella, G., Reiher, M., Fantucci, P., and Stein, M. (2004) *J. Biol. Inorg. Chem.*, **9**, 873–884.
- Swart, M. (2007) *Inorg. Chim. Acta*, **360**, 179–189.
- Perdew, J.P. (1991), in *Electronic Structure of Solids* (eds P. Ziesche and H. Eschrig), Akademie, Berlin, p. 11.
- Perdew, J.P., Chevary, J.A., Vosko, S.H., Jackson, K.A., Pederson, M.R., Singh, D.J., and Fiolhais, C. (1992) *Phys. Rev. B*, **46**, 6671–6687; Erratum (1993) *Phys. Rev. B*, **48**, 4978.
- Xu, X. and Goddard, W.A. III (2004) *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 2673–2677.
- Scherlis, D.A., Cococcioni, M., Sit, P., and Marzari, N. (2007) *J. Phys. Chem. B*, **111**, 7384–7391.

- 34 Handy, N.C. and Cohen, A.J. (2001) *Mol. Phys.*, **99**, 403–412.
- 35 Swart, M., Ehlers, A.W., and Lammertsma, K. (2004) *Mol. Phys.*, **102**, 2467–2474.
- 36 Lee, C., Yang, W., and Parr, R.G. (1988) *Phys. Rev. B*, **37**, 785–789.
- 37 Baker, J. and Pulay, P. (2002) *J. Chem. Phys.*, **117**, 1441–1449.
- 38 Baker, J. and Pulay, P. (2003) *J. Comput. Chem.*, **24**, 1184–1191.
- 39 Hoe, W.M., Cohen, A.J., and Handy, N.C. (2001) *Chem. Phys. Lett.*, **341**, 319–328.
- 40 Bento, A.P., Solà, M., and Bickelhaupt, F.M. (2005) *J. Comput. Chem.*, **26**, 1497–1504.
- 41 Xu, X. and Goddard, W.A. III (2004) *J. Phys. Chem. A*, **108**, 8495–8504.
- 42 Conradie, J. and Ghosh, A. (2007) *J. Phys. Chem. B*, **111**, 12621–12624.
- 43 Zhao, Y., Schultz, N.E., and Truhlar, D.G. (2006) *J. Chem. Theory Comput.*, **2**, 364–382.
- 44 Zhao, Y. and Truhlar, D.G. (2008) *Theory Chem. Acc.*, **120**, 215–241.
- 45 Tao, J.M., Perdew, J.P., Staroverov, V.N., and Scuseria, G.E. (2003) *Phys. Rev. Lett.*, **91**, 146401.
- 46 Staroverov, V.N., Scuseria, G.E., Tao, J., and Perdew, J.P. (2003) *J. Chem. Phys.*, **119**, 12129–12137.
- 47 Hirao, H., Kumar, D., Que, L. Jr., and Shaik, S. (2006) *J. Am. Chem. Soc.*, **128**, 8590–8606.
- 48 Lepetit, C., Chermette, H., Gicquel, M., Heully, J.-L., and Chauvin, R. (2007) *J. Phys. Chem. A*, **111**, 136–149.
- 49 Zein, S., Borshch, S.A., Fleurat-Lessard, P., Casida, M.E., and Chermette, H. (2007) *J. Chem. Phys.*, **126**, 014105.
- 50 Liao, M.-S., Watts, J.D., and Huang, M.-J. (2006) *J. Comput. Chem.*, **27**, 1577–1592.
- 51 Jensen, F. (1998) *Introduction to Computational Chemistry*, John Wiley & Sons, Inc., New York.
- 52 Kamachi, T. and Yoshizawa, K. (2003) *J. Am. Chem. Soc.*, **125**, 4652–4661.
- 53 Güell, M., Luis, J.M., Solà, M., and Swart, M. (2008) *J. Phys. Chem. A*, **112**, 6384–6391.
- 54 Mitin, A.V., Baker, J., and Pulay, P. (2003) *J. Chem. Phys.*, **118**, 7775–7782.
- 55 Martin, J., Baker, J., and Pulay, P. (2009) *J. Comput. Chem.*, **30**, 881–883.
- 56 Szabo, A. and Ostlund, N.S. (1982) *Modern Quantum Chemistry—Introduction to Advanced Electronic Structure Theory*, Macmillan Publishing Co.
- 57 Feixas, F., Solà, M., and Swart, M. (2009) *Can. J. Chem.*, **87**, 1063–1073.
- 58 Cohen, A.J., Tozer, D.J., and Handy, N.C. (2007) *J. Chem. Phys.*, **126**, 214104.
- 59 Swart, M. and Bickelhaupt, F.M. (2008) *J. Comput. Chem.*, **29**, 724–734.
- 60 Swart, M., Solà, M., and Bickelhaupt, F.M. (2007) *J. Comput. Chem.*, **28**, 1551–1560.
- 61 Swart, M. and Bickelhaupt, F.M. (2006) *J. Chem. Theory Comp.*, **2**, 281–287.
- 62 de Jong, G.T., Geerke, D.P., Diefenbach, A., and Bickelhaupt, F.M. (2005) *Chem. Phys.*, **313**, 261–270.
- 63 te Velde, G., Bickelhaupt, F.M., Baerends, E.J., Fonseca Guerra, C., van Gisbergen, S.J.A., Snijders, J.G., and Ziegler, T. (2001) *J. Comput. Chem.*, **22**, 931–967.
- 64 Stewart, J.J. (2007) *J. Mol. Model.*, **13**, 1173–1213.
- 65 Lovell, T., Himo, F., Han, W.-G., and Noodleman, L. (2003) *Coord. Chem. Rev.*, **238–239**, 211–232.
- 66 Himo, F. and Siegbahn, P.E.M. (2003) *Chem. Rev.*, **103**, 2421–2456.
- 67 Siegbahn, P.E.M. and Blomberg, M.R.A. (2000) *Chem. Rev.*, **100**, 421–437.
- 68 Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M. Jr., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., and Kollman, P.A. (1995) *J. Am. Chem. Soc.*, **117**, 5179–5197.
- 69 Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S.J., and Karplus, M. (1983) *J. Comput. Chem.*, **4**, 187–217.
- 70 Groenhof, A.R., Swart, M., Ehlers, A.W., and Lammertsma, K. (2005) *J. Phys. Chem. A*, **109**, 3411–3417.
- 71 Jones, D.H., Hinman, A.S., and Ziegler, T. (1993) *Inorg. Chem.*, **32**, 2092–2095.
- 72 Rydberg, P., Sigfridsson, E., and Ryde, U. (2004) *J. Biol. Inorg. Chem.*, **9**, 203–223.
- 73 Liao, M.-S., Watts, J.D., and Huang, M.-J. (2007) *J. Phys. Chem. A*, **111**, 5927–5935.
- 74 Liao, M.-S. and Scheiner, S. (2002) *J. Chem. Phys.*, **117**, 205–219.

- 75 Rosa, A., Ricciardi, G., and Baerends, E.J. (2006) *J. Phys. Chem. A*, **110**, 5180–5190.
- 76 Reiher, M. (2002) *Inorg. Chem.*, **41**, 6928–6935.
- 77 Pierloot, K. and Vancoillie, S. (2006) *J. Chem. Phys.*, **125**, 124303.
- 78 Malassa, A., Görls, H., Buchholz, A., Plass, W., and Westerhausen, M. (2006) *Z. Anorg. Allg. Chem.*, **632**, 2355–2362.
- 79 Güell, M., Solà, M., and Swart, M. (2009) *Polyhedron*, in press, Doi: 10.1016/j.poly.2009.06.006.
- 80 Siegbahn, P.E.M. and Haeflner, F. (2004) *J. Am. Chem. Soc.*, **126**, 8919–8932.
- 81 Schoneboom, J.C., Cohen, S., Lin, H., Shaik, S., and Thiel, W. (2004) *J. Am. Chem. Soc.*, **126**, 4017–4034.
- 82 Jayapal, P., Sundararajan, M., Hillier, I.H., and Burton, N.A. (2006) *Phys. Chem. Chem. Phys.*, **8**, 4086–4094.
- 83 Bruschi, M., Zampella, G., Fantucci, P., and De Gioia, L. (2005) *Coord. Chem. Rev.*, **249**, 1620–1640.
- 84 Ortiz de Montellano, P.R. and De Voss, J.J. (2002) *Nat. Prod. Rep.*, **19**, 477–493.
- 85 Denisov, I.G., Makris, T.M., Sligar, S.G., and Schlichting, I. (2005) *Chem. Rev.*, **105**, 2253–2278.
- 86 Limberg, C. (2003) *Angew. Chem. Int. Ed.*, **42**, 5932–5954.
- 87 Harris, D.L. and Loew, G.H. (1998) *J. Am. Chem. Soc.*, **120**, 8941–8948.
- 88 Meunier, B., de Visser, S.P., and Shaik, S. (2004) *Chem. Rev.*, **104**, 3947–3980.
- 89 Guallar, V., Gherman, B.F., Lippard, S.J., and Friesner, R.A. (2002) *Curr. Opin. Chem. Biol.*, **6**, 236–242.
- 90 Shaik, S., Cohen, S., de Visser, S.P., Sharma, P.K., Kumar, D., Kozuch, S., Ogliaro, F., and Danovich, D. (2004) *Eur. J. Inorg. Chem.*, 207–226.
- 91 Shaik, S., Kumar, D., de Visser, S.P., Altun, A., and Thiel, W. (2005) *Chem. Rev.*, **105**, 2279–2328.
- 92 Schoneboom, J.C. and Thiel, W. (2004) *J. Phys. Chem. B*, **108**, 7468–7478.
- 93 Fisher, M.T. and Sligar, S.G. (1987) *Biochemistry*, **26**, 4797–4803.
- 94 Schulze, H., Hui Bon Hoa, G., and Jung, C. (1997) *Biochim. Biophys. Acta*, **1338**, 77–92.
- 95 Bach, R.D. and Dmitrenko, O. (2006) *J. Am. Chem. Soc.*, **128**, 1474–1488.
- 96 Groenhof, A.R., Ehlers, A.W., and Lammertsma, K. (2007) *J. Am. Chem. Soc.*, **129**, 6204–6209.
- 97 Harris, D.L. (2001) *Curr. Opin. Chem. Biol.*, **5**, 724–735.
- 98 Yoshioka, S., Tosha, T., Takahashi, S., Ishimori, K., Hori, H., and Morishima, I. (2002) *J. Am. Chem. Soc.*, **124**, 14571–14579.
- 99 Guallar, V., Baik, M.-H., Lippard, S.J., and Friesner, R.A. (2003) *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 6998–7002.
- 100 Guallar, V. and Olsen, B. (2006) *J. Inorg. Biochem.*, **100**, 755–760.
- 101 Guallar, V. (2008) *J. Phys. Chem. B*, **112**, 13460–13464.
- 102 Chen, Z., Ost, T.W.B., and Schelvis, J.P.M. (2004) *Biochemistry*, **43**, 1798–1808.
- 103 Schlichting, I., Berendzen, J., Chu, K., Stock, A.M., Maves, S.A., Benson, D.E., Sweet, B.M., Ringe, D., Petsko, G.A., and Sligar, S.G. (2000) *Science*, **287**, 1615–1622.
- 104 Zurek, J., Foloppe, N., Harvey, J.N., and Mulholland, A.J. (2006) *Org. Biomol. Chem.*, **4**, 3931–3937.
- 105 Altun, A., Guallar, V., Friesner, R.A., Shaik, S., and Thiel, W. (2006) *J. Am. Chem. Soc.*, **128**, 3924–3925.
- 106 Hu, C., Roth, A., Ellison, M.K., An, J., Ellis, C.M., Schulz, C.E., and Scheidt, W.R. (2005) *J. Am. Chem. Soc.*, **127**, 5675–5688.
- 107 Rovira, C., Kunc, K., Hutter, J., Ballone, P., and Parrinello, M. (1997) *J. Phys. Chem. A*, **101**, 8914–8925.
- 108 Strickland, N. and Harvey, J.N. (2007) *J. Phys. Chem. B*, **111**, 841–852.
- 109 Ghosh, A., Vangberg, T., Gonzalez, E., and Taylor, P. (2001) *J. Porphyrins Phthalocyanines*, **5**, 345–356.
- 110 Ghosh, A., Tangen, E., Ryeng, H., and Taylor, P.R. (2004) *Eur. J. Inorg. Chem.*, 4555–4560.
- 111 Ghosh, A., Persson, B.J., and Taylor, P.R. (2003) *J. Biol. Inorg. Chem.*, **8**, 507–511.
- 112 Liao, M.-S. and Scheiner, S. (2002) *J. Comput. Chem.*, **23**, 1391–1403.
- 113 Fraústo do Silva, J.J.R. and Williams, R.J.P. (1991) *The Biological Chemistry of the Elements. The Inorganic Chemistry of Life*, Oxford University Press, Oxford.

- 114 Siegbahn, P.E.M., Tye, J.W., and Hall, M.B. (2007) *Chem. Rev.*, **107**, 4414–4435.
- 115 Wang, H., Ralston, C.Y., Patil, D.S., Jones, R.M. *et al.* (2000) *J. Am. Chem. Soc.*, **122**, 10544–10552.
- 116 Fan, H.-J. and Hall, M.B. (2002) *J. Am. Chem. Soc.*, **124**, 394–395.
- 117 van Lenthe, E. and Baerends, E.J. (2003) *J. Comput. Chem.*, **24**, 1142–1156.
- 118 Baerends, E.J., Autschbach, J., Bérces, A., Berger, J.A. *et al.* (2007) ADF version 2007.01, SCM, Amsterdam.
- 119 Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E. *et al.* (2004) Gaussian 03, Revision B.02, Gaussian Inc., Pittsburgh PA, USA.
- 120 Bylaska, E.J., de Jong, W.A., Kowalski, K., Straatsma, T.P. *et al.* (2006) NWChem version 5.1, Pacific Northwest National Laboratory, Richland, Washington, USA.
- 121 Neese, F. (2007) ORCA, Bonn.
- 122 Vosko, S.H., Wilk, L., and Nusair, M. (1980) *Can. J. Phys.*, **58**, 1200–1211.
- 123 Perdew, J.P., Burke, K., and Ernzerhof, M. (1996) *Phys. Rev. Lett.*, **77**, 3865–3868; Erratum (1997) *Phys. Rev. Lett.*, **78**, 1396.
- 124 Perdew, J.P. and Wang, Y. (1992) *Phys. Rev. B*, **45**, 13244–13249.
- 125 Perdew, J.P. and Zunger, A. (1981) *Phys. Rev. B*, **23**, 5048–5079.
- 126 Swart, M. and Bickelhaupt, F.M. (2006) *Int. J. Quantum Chem.*, **106**, 2536–2544.
- 127 Wittbrodt, J.M. and Schlegel, H.B. (1996) *J. Chem. Phys.*, **105**, 6574–6577.

20

Quantum Mechanical Approaches to Selenium Biochemistry

Jason K. Pearson and Russell J. Boyd

20.1

Introduction

Understanding the intricate details of enzyme function at the molecular level is of fundamental importance to biological science. An impressive array of experimental techniques have led to many advances in biochemistry, but often critical information about the energetics and structural features of intermediates and transition states in enzymatic reactions is inaccessible. A major strength of quantum chemical calculations is their ability to elucidate many details of a given molecular system, which serves as a model for the real system. Rapid advances in computer technology as well as in computational algorithms and methods have made computational chemistry a viable and complementary partner to experiment in both chemistry and biochemistry.

Selenium was first discovered by the Swedish chemist Berzelius in 1818 [1] and was named after the Greek goddess of the moon, Selene. Eighteen years later, Lowig prepared diethyl selenide (for information on the nomenclature of selenium compounds, the reader is referred to the work by Guenther [2]), and in doing so became the first person to synthesize a selenium-containing compound [3]. Selenium has a wide variety of applications, particularly in medicine, despite being considered a deadly poison until 1957 when it was found that it was a micronutrient for such organisms as bacteria, birds and even mammals [4]. In addition to medicine, modern advances in selenium chemistry are driven by applications in organic synthesis [5–7], biochemistry [8], ligand chemistry [9] and as precursors for metal organic chemical vapor deposition (MOCVD) of semiconducting materials [10].

The study of the biochemical properties of selenium began in 1973 when it was discovered that it played a vital role as the main component in the active site of the antioxidant enzyme glutathione peroxidase (GPx) [11, 12]. Selenium has since been identified as the main component of the active site in several other enzymatic systems such as the iodothyronine deiodinases [13–16], thioredoxin reductases [17–22], selenophosphate synthetase [21] and selenoprotein P [23]. The selenium-containing residue in GPx was shown to be selenocysteine (Sec) [24], which is the

selenol (Se-H) analog of the thiol (S-H) amino acid cysteine (Cys). Sec is often referred to as the 21st naturally occurring amino acid because it is produced naturally in the body but, surprisingly, it is not typically included among the 20 amino acids in standard biochemistry texts.

Selenium and sulfur show marked differences in their biological behavior and indeed the substitution of sulfur for selenium in GPx is imperative as it provides the enzyme with its antioxidant ability. Although selenium shares many chemical properties with its group 16 neighbor, the incorporation of a chemically more active selenol moiety into the active site as a substitution for a thiol allows for a dramatic catalytic advantage. Direct comparison of the activity of natural GPx and the mutant murine GPx in which sulfur replaces selenium shows that the sulfur analog is 1000-fold less active than the parent enzyme [25]. Moreover, small molecule organo-selenium antioxidants have inactive sulfur-containing counterparts [26], reinforcing the central role played by selenium and the potential for harnessing its reductive ability.

To this end there have been many experimental and theoretical studies on the potential of synthetic organo-selenium compounds as antioxidants, enzyme inhibitors, antitumor and anti-infective agents, and immunomodulators [27]. The remainder of this chapter will focus on reviewing computational studies of selenium-containing systems, including GPx and several small-molecule GPx mimics with various potential biologic applications. A comprehensive survey of the relevant literature on the quantum biochemistry of organo-selenium compounds is beyond the scope of this chapter. We begin, instead, by briefly reviewing several benchmark studies designed to test quantum mechanical methods for predicting properties of selenium-containing compounds and subsequently summarize several applications of such methods to selenium biochemistry.

20.2

Quantum Mechanical Methods for the Treatment of Selenium

The earliest theoretical work on systems involving selenium focused mainly on the properties of small inorganic compounds, while investigations of the biologic aspects of selenium chemistry appeared later. One of the first papers highlighting calculations on selenium appeared in 1981 with Hinchliffe calculating the electronic structure and properties of OSeCl [28]. Subsequent work included investigation of the selenium–sulfur bond by Laitinen *et al.* [29], the structures of SeH₂ and SeH₂⁺ using non-relativistic effective core potentials (ECPs) by Muller *et al.* [30], supplementary d-orbital exponents for hypervalent selenium compounds by Angyan *et al.* [31], ionization energies of SeH and SeH₂ using a multiconfigurational approach with relativistic ECPs by Balasubramanian *et al.* [32, 33], and the ionization energies and bond dissociation energies of the SeH_n series by Binning and Curtiss [34, 35].

The first systematic investigation of quantum mechanical methods appropriate for biologically relevant organo-selenium compounds was reported by the current

authors in 2005 [36]. A series of eleven organo-selenium structures were constructed to mimic the bonding environment of the selenium atom in a wide range of biologic contexts and the ability of density functional methods to reproduce accurate geometries and bond dissociation energies was tested. Specifically, Becke's three-parameter exchange functional (B3) [37] in combination with the correlation functionals of Lee, Yang and Parr (LYP) [38] and Perdew and Wang (PW91) [39] were chosen on the basis of being widely used functionals and they were tested in conjunction with a series of Pople basis sets to probe the necessity of high angular momentum and diffuse functions.

The most reliable data was obtained with the B3PW91/6-311G(2df,p) method, although both the B3LYP and B3PW91 functionals performed well and the double-zeta 6-31G(d,p) basis set was shown to be a reasonable compromise [36]. In light of the performance of standard Gaussian basis sets, it was concluded that ECPs are not necessary to obtain good results for the geometries and bond dissociation energies of selenium-containing molecules. Moreover, it is also apparent that relativistic effects need not be taken into account, even though relativistic ECPs are frequently used to treat organo-selenium systems (see below). Bayse has evaluated MP2 and DFT methods for the reliable calculation of ^{77}Se chemical shifts on a wide range of organo-selenium compounds [40] and has shown that theoretical chemical shifts may be used as a probe for predicting the selenium oxidation state in selenoproteins and their mimics [41]. In general, the mPW1PW91 [42] functional was shown to predict the most reliable chemical shifts with either a triple-zeta Gaussian basis set or a limited set of relativistic ECPs. Alternatively, Keal and Tozer have investigated the use of DFT for ^{77}Se NMR shielding constants and have cited their own KT3 functional [43] as being superior although the mPW1PW91 functional was not tested.

20.3

Applications to Selenium Biochemistry

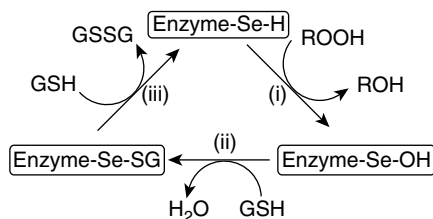
20.3.1

Computational Studies of GPx

Aerobic organisms derive their energy from the reduction of O_2 and are therefore susceptible to the damaging effects of small amounts of $\text{O}_2^{\cdot-}$, $\cdot\text{OH}$ and H_2O_2 that inevitably form during the metabolic consumption of oxygen. These three species, together with unstable intermediates in the peroxidation of lipids, are referred to as reactive oxygen species (ROS) [44] and their adverse effects include, but are not limited to, the destruction of key biologic components and the damage of cell membranes. This condition is referred to as "oxidative stress" [45, 46] and is particularly prevalent in the electron-transfer system of mitochondria [27]. Many conditions such as Alzheimer's disease, myocardial infarction, atherosclerosis, Parkinson's disease, autoimmune diseases, radiation injury, emphysema and sunburn are linked to damage from ROS as a result of an imbalance between

radical-generating and radical-scavenging systems and thus a major research effort is focused on combating oxidative stress.

The defense mechanism of the human body against oxidative stress is elaborate but the key steps involve the dismutation of superoxide ($\text{O}_2^{\bullet -}$) to H_2O_2 and O_2 by superoxide dismutase (SOD) [47] and the reduction of H_2O_2 by GPx [44, 48]. GPx reduces hydroperoxides to water (or the corresponding alcohol in the case of organic hydroperoxides) at the expense of a glutathione (GSH) cofactor [49] via the mechanism shown in Scheme 20.1.



Scheme 20.1 Catalytic cycle of GPx.

The selenium moiety in the Sec residue of the enzyme (enzyme-Se-H) is oxidized to the selenenic acid derivative (enzyme-Se-OH) by reduction of peroxide to the corresponding alcohol in step (i). The thiol GSH will then attack the selenenic acid and displace a water molecule to form a seleno-sulfide (enzyme-Se-S-G) intermediate in step (ii). To complete the cycle, a second GSH converts the seleno-sulfide intermediate back into the original selenol by liberating the oxidized glutathione (GSSG) in step (iii). Thus, two GSH molecules are consumed in the process, generating GSSG.

There are four families of GPx enzymes in mammalian organisms [25, 48, 50, 51], including the classical cytosolic GPx (cGPx), which are found in the cytosolic and mitochondrial compartments; phospholipid hydroperoxide GPx (PHGPx), which are intracellular and partially membrane-bound; plasma GPx (pGPx), which are plasma-specific enzymes; and gastrointestinal GPx (giGPx), which resides in the gastrointestinal tract. Wherever possible, the specific acronyms will be used in the text though collectively they shall be referred to as simply GPx.

The crystal structure of human pGPx was solved in 1997 and at that time it was discovered that the enzyme is tetrameric, consisting of two asymmetric units, each made up of a dimer with half-site reactivity [52]. Comparison of the crystal structure of human pGPx with that of bovine erythrocyte cGPx [24] reveals that in the latter there are two water molecules present in the active site that cannot be observed in the former due to the resolution of the crystal data. A key question is whether water is present in the active site of human pGPx and what role it plays in the antioxidant activity of the enzyme.

Morokuma *et al.* have pursued this question theoretically using both density functional theory and the hybrid quantum mechanics/molecular mechanics (QM/MM) method [53]. The structure of the active site of human pGPx was modeled using DFT at the B3LYP/6-31G(d) level with and without water molecules present. In

addition, the entire enzyme (monomer) was modeled using several variations of the QM/MM scheme where the active site was modeled using quantum mechanics [in this case HF/STO-3G or B3LYP/6-31G(d)] and the remainder of the structure is treated with molecular mechanics (in this case Amber). The results show that the root-mean-square deviation between the calculated structure and the crystal structure is minimized when two water molecules are included in the active site, thus providing strong evidence for the presence of two bound water molecules in the active site of human pGPx, especially in light of the fact that two water molecules are observed in bovine cGPx.

The peroxidase activity of pGPx has also been investigated theoretically using the same methods [54, 55]. The full catalytic cycle (as illustrated in Scheme 20.1) of pGPx has been modeled using an “active site only” system with the B3LYP/6-31G(d) method and using the full enzyme (monomer) with the QM/MM method employing B3LYP/6-31G(d)/Amber. Some important conclusions drawn from the study include the effect of residues neighboring the Sec throughout the course of the cycle. In particular Gln83, Gly50 and the two water molecules provide stabilizing interactions and the predicted overall barrier for the formation of the selenenic acid (enzyme-Se-OH) is $18.0 \text{ kcal mol}^{-1}$ ($1 \text{ kcal} = 4.184 \text{ kJ}$), which is in good agreement with the experimental value of $14.9 \text{ kcal mol}^{-1}$ [56]. Experimentally, the Gln83 residue has been suspected of participating in the antioxidant activity of GPx [24] and these calculations reveal that it plays a major role by facilitating proton transfer and providing an H-bond acceptor throughout the catalytic cycle. Proton transfer is also shown to be facilitated by the active site’s water molecules, elucidating the role water plays in the catalytic activity of pGPx. The effect of the rest of the surrounding protein is predicted to be minimal, generating an increase in the overall barrier of the reaction of only $0.70 \text{ kcal mol}^{-1}$; however, this is probably because the active site is located on the surface of the enzyme.

20.3.2

Computational Studies on GPx Mimics

20.3.2.1 GPx-like Activity of Ebselen

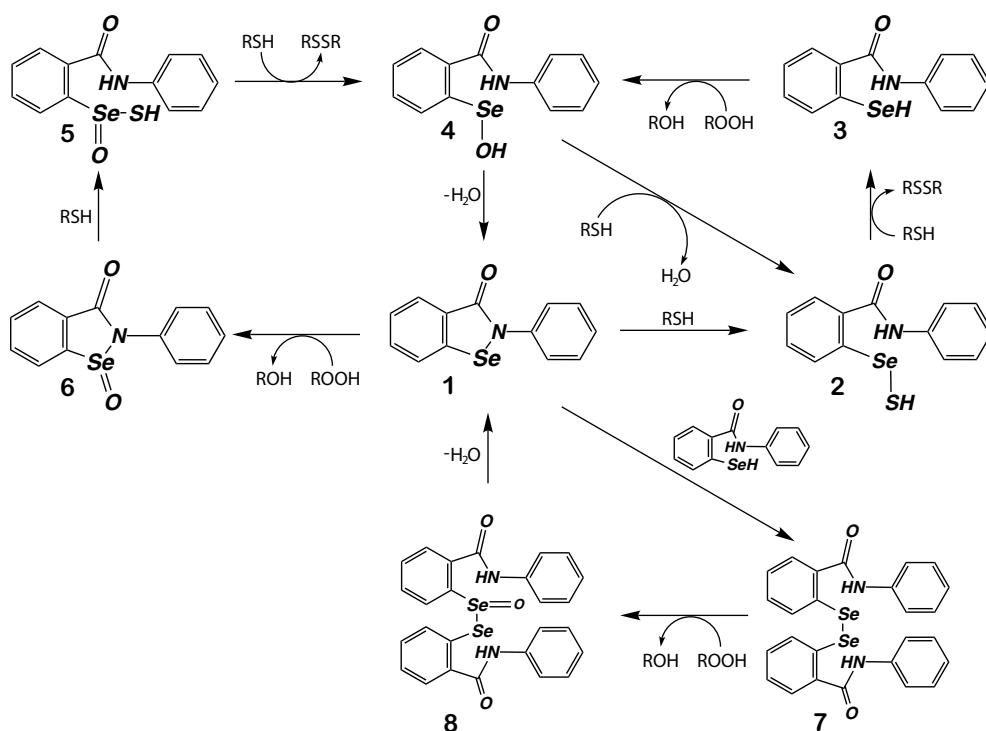
An increase in GPx activity results in an increased ability to cope with oxidative stress. This has been observed in endothelial cells injected with purified GPx, which show a marked increase in survival on exposure to hyperoxia and redox cyclers [57–59]. At first glance, this provides a promising defensive agent against the many detriments of oxidative stress and one might conclude that a wide variety of conditions may be treated with purified GPx. However, natural GPx proteins are not likely to be utilized as pharmacologic agents to increase intracellular GPx activity *in vivo* because they cannot be expressed in prokaryotes and they are not compatible with oral administration and cellular targeting [60]. For this reason it is desirable to develop synthetic GPx mimics, and indeed many organo-selenium compounds have been studied as bio-models that simulate catalytic functions of various enzymes, including GPx [27].

These model systems have a wide range of basic structures; however, none has received more attention than ebselen [2-phenyl-1,2-benzisoselenazol-3(2*H*)-one] (**1**).

Ebselen is a cyclic selenamide that has been extensively studied as an antioxidant and GPx mimic [61–66]. Although there have been many proposed GPx mimics, ebselen was the first compound to be used in clinical trials [61–63] and has attracted interest because of its anti-inflammatory, antiatherosclerotic and cytoprotective properties in both *in vitro* and *in vivo* models [67–71].

Ebselen was first prepared in 1924 by Lesser and Weiss [72]; however, it took six decades to realize its potential as a GPx mimic [61] by reacting slowly with H_2O_2 and other peroxides to afford a stable selenoxide product (6). Although ebselen lacks a selenol moiety (Se-H) – and, therefore, also lacks structural similarity to the active site Sec residue in any selenoenzyme – the selenamide bond is readily cleaved by thiol (Scheme 20.2) to afford a selenosulfide (2), which can then be converted into the selenol derivative of ebselen (3) by an additional thiol. In fact, a diselenide (7) can also be formed via the combination of the original ebselen molecule and the selenol derivative [73]. Scheme 20.2 illustrates the available pathways for interconversion of ebselen (1) to its diselenide (7) and selenol (3) derivatives and for the reduction of peroxides by each [73–75].

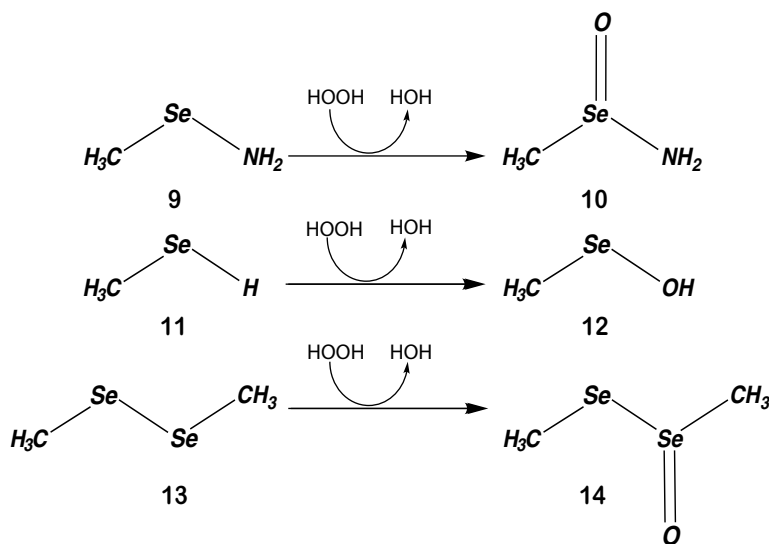
The selenol form of ebselen, or the selenolate anion (Se^-) to be more precise, is the active form in terms of the reduction of hydrogen peroxide and indeed it would seem



Scheme 20.2 Summary of catalytic cycles involving the reduction of peroxides by ebselen, ebselen selenol and ebselen diselenide.

that the existence of a selenol functionality would be required for a GPx-like mechanism (i.e., $3 \rightarrow 4 \rightarrow 2$). Nevertheless, GPx mimics in the literature have a wide range of structures, including selenamides [76–79], diselenides [74, 80–84], allyl selenides [74], aryl selenides [85] and seleninate esters [86]. Although there are key differences, many of these compounds are related to one of the three derivatives of ebselen in the redox map in Scheme 20.2; that is, these mimics are usually selenols, diselenides or have a divalent selenium atom bound within a heterocycle. Consequently, the reductive pathway of each becomes an important piece of the puzzle in understanding selenium bioactivity, not just that of the selenol.

The catalytic cycle of ebselen has been controversial as a consequence of major differences in working conditions such as solvents, pH and the choice of peroxide. For this reason, several mechanisms have been proposed; however, the most reliable data under physiologically relevant conditions is that shown in Scheme 20.2. To better understand the role played by ebselen, ebselen diselenide, ebselen selenol and the selenolate anion, the current authors carried out two concurrent DFT studies of the reduction of hydrogen peroxide, the first using model compounds [87] and the second using the full structures [88]. The model compounds were constructed to be the simplest structures that maintained the immediate bonding environment about the selenium atom (Scheme 20.3).



Scheme 20.3 Model reactions for the reduction of hydrogen peroxide by ebselen (9), ebselen selenol (11) and ebselen diselenide (13).

For this investigation geometry optimizations were performed with Becke's three-parameter exchange functional (B3) in conjunction with the correlation functional proposed by Perdew and Wang (PW91) using a 6-311G(2df,p) basis set as suggested previously for the reliable prediction of organo-selenium geometries

Table 20.1 Summary of Gibbs energy barriers for the reduction of hydrogen peroxide by models of ebselen, ebselen selenol and ebselen diselenide (Scheme 20.3).

Reaction	Gibbs energy barrier (kcal mol ⁻¹)
Ebselen model oxidation (step 1; 9 → 10)	56.7
Ebselen model oxidation (step 2; 9 → 10)	6.4
Solvated ebselen model oxidation (step 1; 9 → 10)	60.2
Solvated ebselen model oxidation (step 2; 9 → 10)	8.5
Ebselen selenol model oxidation (13 → 14)	53.4
Solvated ebselen selenol oxidation (13 → 14)	57.7
Ebselen selenol model anion oxidation	49.4
Ebselen diselenide oxidation (13 → 14)	35.3
Solvated ebselen diselenide oxidation (13 → 14)	29.6

and energetics [36]. Transition states were located using Schlegel's synchronous transit-guided quasi-Newton (STQN) method [89, 90] and were linked to reactant and product complex structures by the use of an intrinsic reaction coordinate calculation [91, 92]. Frequency calculations were performed on all optimized structures using the B3PW91/6-311G(2df,p) method to obtain accurate thermochemical data and to confirm whether a structure was a minimum or first-order saddle point; accurate energies were obtained for all structures via single-point calculations using the 6-311++G(3df,3pd) Pople basis set with the above DFT method. Solvent effects were incorporated implicitly with single-point calculations using the conductor-like polarizable continuum model [93] (CPCM with a dielectric constant of 78.39 for water) at the B3PW91/6-311++G(3df,3pd) level and explicitly (for the case of the deprotonated selenol anion reaction) with the inclusion of three water molecules. For the selenol anion reaction, diffuse functions were also included on heavy atoms for the geometry optimizations and frequency calculations as well as in the transition state searches.

Table 20.1 summarizes the reaction Gibbs energy barriers. The barriers are largely overestimated as Moregenstern *et al.* [94] have determined the experimental Gibbs energy barriers for ebselen, ebselen diselenide and ebselen selenol to be 18.6 ± 0.2 , 18.1 ± 0.1 and 16.5 ± 0.1 kcal mol⁻¹, respectively. This overestimation is predominantly due to the model compounds themselves, as can be demonstrated from the complexes predicted in the oxidation of the ebselen model (Figure 20.1).

The reaction coordinate of this species illustrates a two-step conversion into products via a proton abstraction from the selenamide. Such a process is not possible for ebselen as the nitrogen atom does not have any available protons and thus it was apparent that the chemistry of the full molecules could not be captured using truncated models of their structure. Therefore, investigation of the full molecules was necessary.

Owing to the greater computational cost of modeling the full structures, in particular ebselen diselenide, geometry optimizations were performed with Becke's three-parameter exchange functional (B3) in conjunction with the correlation functional proposed by Lee, Yang and Parr (LYP) using the 6-31G(d,p) Pople basis

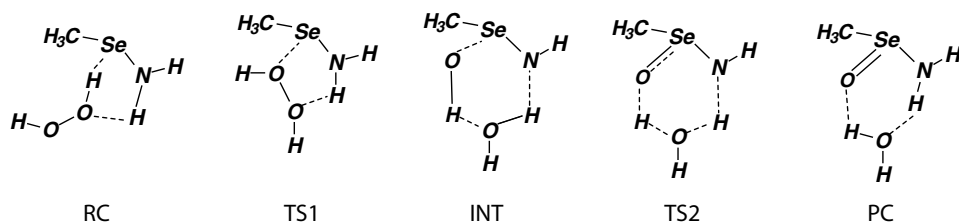


Figure 20.1 Schematic structures for the reduction of hydrogen peroxide by the ebselen model. RC indicates a reactant complex, TS a transition state, INT an intermediate and PC a product complex.

set. This level of theory is computationally less expensive than that used for the model structures but was shown to be reliable for the prediction of organo-selenium geometries and energetics [36]. Accurate energies were obtained for all structures via single-point calculations using the 6-311++G(3df,3pd) Pople basis set with the B3LYP method. Aqueous solvent effects were again incorporated implicitly with single-point calculations using the CPCM model at the B3LYP/6-311++G(3df,3pd) level. In addition, for the selenol anion reaction, diffuse functions were included on heavy atoms for the geometry optimizations and frequency calculations as well as in the transition state searches.

In the model compound study, the selenolate anion was shown to be more active towards the reduction of hydrogen peroxide than the neutral selenol. In addition, aromatic selenols have an experimental pK_a of approximately 6, again suggesting that at physiological pH (≈ 7) the anion is the most likely species to be present; however, it was also prudent to probe the likelihood of the selenol zwitterion as a contributor to the GPx activity of ebselen selenol. Therefore, the conversion of the neutral selenol into the zwitterion was investigated (Figure 20.2).

Calculations on this system show that the zwitterion is unlikely to be the most active reducing agent as it lies on an unstable potential energy surface. The zwitterion is nearly equal in energy to the transition state between it and the selenol and thus is expected to be short-lived if produced at all. Sarma and Mugesch have also experimentally determined that the ebselen selenol zwitterion is an unstable isomer of the neutral selenol [95]; therefore, the neutral selenol was not included in the full molecule investigation, only the selenolate anion.

The predicted solution-phase Gibbs energy barriers for the reduction of hydrogen peroxide by ebselen, the selenolate anion and ebselen diselenide are 36.8, 32.5 and 38.4 kcal mol⁻¹, respectively. Both the gaseous and solution-phase barriers are

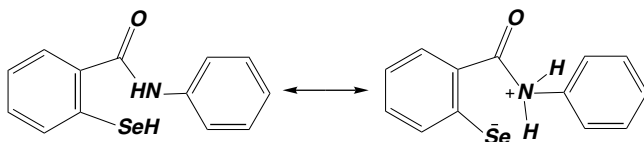


Figure 20.2 Interconversion between the neutral selenol and its zwitterion.

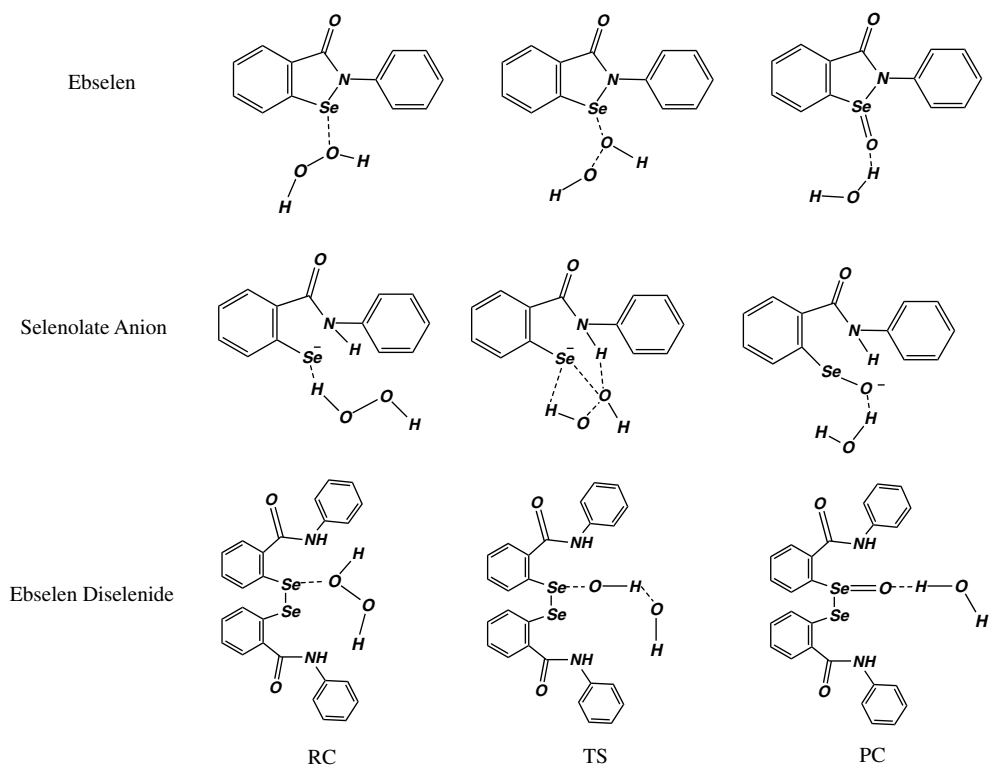


Figure 20.3 Schematic reaction mechanisms for the reduction of hydrogen peroxide by ebselen, the selenolate anion and ebselen diselenide.

indeed lower than for the case of the model compounds with the exception of the diselenide, which exhibits similar barriers. Although the barriers are still overestimated with respect to the experimental values, there is a qualitative agreement between the results and experiment in that the selenolate anion is predicted to be the most active of the three.

Mechanistically, all three reactions proceed via a similar proton shift mechanism (Figure 20.3). The incoming hydrogen peroxide molecule is oriented with one end facing the substrate selenium atom. The proton on this end of the peroxide is transferred to the neighboring oxygen atom while the selenoxide bond is formed, resulting in a simultaneous proton transfer/selenoxide bond formation.

To probe in more detail the reasons for the higher activity of the selenolate anion, the topology of the electron density was analyzed using the quantum theory of atoms in molecules (QTAIM) [96]. QTAIM provides a tool with which one can unambiguously partition any molecular system based on the topology of its electron density into regions of space that define each atom in the molecule. A consequence of this partitioning is that the properties of the atoms can be summed quantitatively to yield

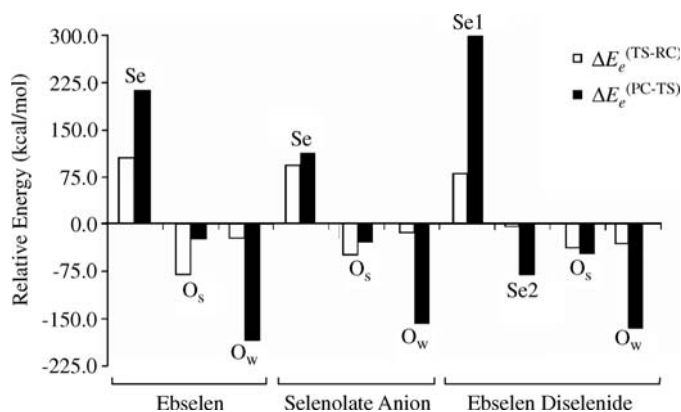


Figure 20.4 Atomic electronic energy changes for the reduction of hydrogen peroxide by ebselen, ebselen selenol and ebselen diselenide. The difference in atomic energy between the TS and RC, $\Delta E_e^{(TS-RC)}$, is shown as white bars while the difference in atomic energy between the PC and TS, $\Delta E_e^{(PC-TS)}$, is shown in black. A positive

value indicates a destabilization and a negative value indicates a stabilization; O_s indicates the peroxide oxygen that becomes the selenoxide oxygen and O_w indicates the peroxide oxygen atom of the resultant water molecule. For the diselenide, Se1 is the oxidized selenium atom.

the total value of that property for the full molecular system. Therefore, one can decompose any such molecular property into the individual atomic contributions. The implications of this have been observed empirically for many decades as the transferability of the properties attributed to atoms and functional groups between different molecules.

Decomposition of the molecular energy into its atomic components for the reactant complex (RC), transition state (TS) and product complex (PC) along a particular reaction coordinate affords the construction of an atomic reaction energy profile as shown in Figure 20.4. These profiles outline the atomic energy changes throughout the course of the reaction and in this particular case indicate a significant destabilization of the selenium atom and corresponding stabilization of the peroxide oxygen atoms in proceeding both from the reactant complex to the transition state and from the transition state to the product complex. The differences between the energy profiles of the three species are very subtle; however, the selenium atom in the case of the selenolate anion is clearly far less destabilized than in the case of the other two species despite the peroxide oxygen atoms having a very similar profile in all three species.

An explanation for this behavior is found in the decomposition of the electronic charge of each system. Analysis of the electronic population data for each case shows that the selenium atom loses 0.94–1.02 *e* of electronic charge by being oxidized, which is entirely transferred to the two peroxide oxygen atoms. In all cases, the oxygen atom of the resultant water molecule (O_w) recovers the majority of electronic charge (~60%) while the selenoxide oxygen (O_s) recovers the rest (~40%). The charge transfer for each case is slightly different though in terms of when it takes place. For the ebselen and ebselen diselenide systems, the selenium atom loses 0.36 and 0.30 *e*

of electronic charge at the TS, respectively, and loses an additional 0.58 and 0.72 e of electronic charge at the PC, respectively.

The selenium atom of the selenolate anion, however, has a greater electron population to begin with, which facilitates the loss of 0.47 e of electronic charge at both the TS and the PC. This reflects the higher propensity of an anion to be oxidized and elucidates the energy profile of the selenolate and the lower barrier to oxidation.

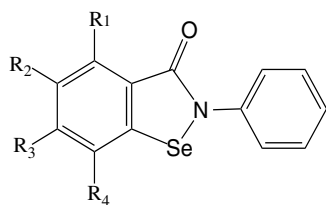
Sarma and Mugesh [95] have also used DFT to provide valuable supplemental data for their experimental investigation into the role of thiols in the catalytic cycle of ebselen and some ebselen analogues [97]. The structures of ebselen and several analogues were optimized with the B3LYP/6-31G(d) method with the effect of solvation in an aqueous medium incorporated using the isodensity polarizable continuum model (IPCM) [98]. Subsequent calculations were used to investigate the ^{77}Se chemical shifts using the GIAO method [99] and atomic interactions using the natural bond orbital (NBO) method [100].

Reduction of peroxides by the selenol form of ebselen is accompanied by a series of additional steps analogous to the catalytic cycle of GPx itself (Scheme 20.2). Sarma and Mugesh have shown that this process is further complicated by thiol exchange reactions whereby an incoming thiol that could potentially react with the seleno-sulfide (2) intermediate will generally prefer to attack the more nucleophilic selenium atom, yielding an exchange of the thiol substituents and no net reaction [95]. Thiol exchange has therefore been implicated in the relatively low GPx-like activity of ebselen, depending upon the thiol. These results have been supported by the theoretical work of Bachrach *et al.* [101]. In that study, various nucleophiles (HS^- , CH_3S^- , HSe^- and CH_3Se^-) and substrates (R_1SSeR_2 and R_1SeSeR_2 with R_1 and $\text{R}_2 = \text{H}$ or Me) were used to model gas-phase substitution at selenium and sulfur in diselenides and selenosulfides. Using MP2 and B3LYP, the PES of each reaction was investigated and in all cases it was found that nucleophilic attack at selenium is both kinetically and thermodynamically preferred over attack at sulfur. The implication of these results is that one needs to tackle thiol exchange reactions to design new, more effective GPx mimics.

20.3.2.2 Substituent Effects on the GPx-like Activity of Ebselen

Experimental investigation into potential organo-selenium therapeutic agents is often accomplished by generating a reasonably high number of unique organo-selenium compounds and testing them using various prescribed assays. The results will point to promising candidates for future drug molecules; however, a key disadvantage of such a methodology is that a proper understanding of the mechanism of action of any particular compound may be lost; that is, a full understanding of why a compound yields desired properties or not may not be fully explored and consequently the method may fail to uncover structures beyond the scope of the study whereas a knowledge-based approach could succeed in doing just that.

Consider the case of ebselen with a nitro substituent adjacent to the selenium atom. Experimentally this compound exhibits a GPx activity that is nine times greater than ebselen itself [102]. It is tempting to attribute this difference to the electron-withdrawing nature of the NO_2 and, indeed, this has been done [27]; however, in the



- (1) $R_1, R_2, R_3, R_4 = H$
 (15) $R_1 = H, R_2 = NO_2, R_3, R_4 = H$
 (16) $R_1, R_2, R_3 = F, R_4 = H$
 (17) $R_1, R_2, R_3 = H, R_4 = NO_2$
 (18) $R_1, R_2, R_3 = H, R_4 = HCO$
 (19) $R_1, R_2, R_3 = H, R_4 = t\text{-but}$
 (20) $R_1, R_2, R_3 = H, R_4 = OCH_3$

Figure 20.5 Electronic and steric modifications to the ebselen structure.

oxidation of organo-selenium compounds by peroxide it has been shown that there is a significant reduction in electron density at the selenium atom [88] and thus an electron-withdrawing substituent would not be expected to aid such a process.

We have therefore investigated substituent effects on the GPx-like activity of the ebselen molecule with the goal of uncovering a more feasible explanation for the observed increase in activity of the nitro derivative [103]. A series of seven structures (Figure 20.5) were chosen to probe the electronic and steric effects of substituents on the ebselen framework and the transition states of the reactions of each with hydrogen peroxide were located. Reactant complexes, product complexes and transition states were optimized using the B3LYP/6-31G(d,p), and the topology of the electron density was analyzed using QTAIM. Table 20.2 summarizes the results.

Any structure having a substituent in the R_4 position (i.e., 17–20) exhibits a significantly larger barrier to oxidation. While the decomposition of the electronic charge of the system into its atomic components indicates that the electron-withdrawing and electron-donating substituents do indeed contribute to the charge on the selenium atom (q_{Se}), the charge transfer during the reaction (indicated by Δq) is

Table 20.2 Electronic energy barriers ($\Delta E_e^\ddagger$), dihedral angles in the TS for the direct oxidation of ebselen and several derivatives by hydrogen peroxide. electronic charges on selenium (q_{Se}), changes in electronic charge on selenium and each peroxide oxygen atom ($\Delta q_{Se}^{(TS-RC)}$, $\Delta q_{Os}^{(TS-RC)}$, $\Delta q_{Ow}^{(TS-RC)}$) and

Structure	$\Delta E_e^\ddagger$ (kcal mol ⁻¹)	q_{Se} (e)	$\Delta q_{Se}^{(TS-RC)}$ (e)	$\Delta q_{Os}^{(TS-RC)}$ (e)	$\Delta q_{Ow}^{(TS-RC)}$ (e)	$\angle OOSeC$ (°)
1	25.7	0.57	0.36	-0.25	-0.25	-93.3
15	27.8	0.61	0.33	-0.25	-0.25	-93.3
16	26.6	0.60	0.34	-0.25	-0.26	-90.6
17	34.0	0.74	0.37	-0.23	-0.25	-128.2
18	33.1	0.71	0.37	-0.23	-0.26	-130.9
19	30.9	0.55	0.38	-0.23	-0.28	-122.0
20	30.9	0.57	0.37	-0.23	-0.27	-106.2

virtually unchanged in all cases. There is, however, a stark contrast between the geometry of the transition state structures for cases where the substituent is in the R₄ position (i.e., 17–20) and for cases where the substituent is anywhere else, as is illustrated by the dihedral angles in Table 20.2. These data strongly suggest a steric effect induced by substituents located close to the selenium atom that hinder the approach of incoming peroxides and thus raise the overall barrier of the reaction.

In the context of the full catalytic cycle of ebselen (Scheme 20.2) it is apparent that while a substituent near the selenium atom will hinder the approach of a peroxide molecule in the direct oxidation reaction (**1** → **6**), it must also hinder the approach of a thiol in the selenosulfide intermediate and thus prevent thiol exchange. Since it has been established that thiol exchange is to blame for poor activity in the case of ebselen, the experimentally observed increased activity of the nitro derivative of ebselen can be attributed to the steric effect of the substituent and thus a modification of the local steric environment of the selenium atom provides a tool for overcoming thiol exchange and increasing the GPx-like activity of the parent compound.

One might also expect that altering the framework of the parent compound would be advantageous to overcome thiol exchange. A much less obvious GPx mimic was discovered in 1,3-dihydro-1-methyl-2*H*-imidazole-2-selenol (MSeI, Figure 20.6) when Mugesh *et al.* found, experimentally, its GPx activity to be much greater than that of ebselen [56, 104]. Soujanya *et al.* subsequently carried out a B3LYP/6-31G(d) investigation into its mechanism [105]. The full catalytic cycle analogous to that in Scheme 20.2 was studied with MSeI as the substrate and, despite calculating prohibitively high barriers for a catalytic process, it was shown that the Mulliken charges on the sulfur and selenium atoms of the seleno-sulfide intermediate were 0.434 and –0.390, respectively, suggesting that the imidazole framework may provide a basis to overcome thiol exchange in such systems. Any substrate framework that has the potential to enhance nucleophilic attack at sulfur versus selenium in the selenosulfide intermediate affords the opportunity to overcome thiol exchange and, therefore, has a higher potency as a GPx mimic. This may explain the observed high GPx-like activity of MSeI and, consequently, give new insight into effective GPx mimics.

20.3.2.3 Effect of the Molecular Environment on GPx-like Activity

An intuitive choice for a GPx mimic would be the Sec residue itself. Cardey and Enescu have studied the reduction of hydrogen peroxide by selenolate and thiolate [106] as well as the anions of Sec and Cys [107] theoretically. This was with the goal of uncovering details regarding what the effect of the immediately bonded environ-

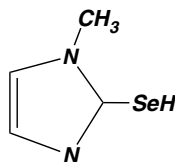


Figure 20.6 Structure of 1,3-dihydro-1-methyl-2*H*-imidazole-2-selenol (MSeI).

ment of the selenium and sulfur atoms were that contributed to the overall activity of GPx. In their study, the integrated molecular orbital + molecular orbital (IMOMO) method [108] was employed, which has the ability to combine two calculations at two different levels of theory, analogous to the QM/MM methods employed for the full GPx protein (see above). In an IMOMO calculation, one chooses a high-level method to be applied to a restrained part of the quantum system and a lower level to be applied to the entire system; in the case of Cardey and Enescu, QCISD(T) and MP2 were chosen, respectively. The complexes and transition states of each reaction with the Sec and Cys residues were optimized using the MP2/6-311+G(d,p) method in the gas phase, and in the aqueous phase with the use of a PCM dielectric, and the energetics of the reaction were subsequently calculated using the IMOMO method.

The results showed that Sec and Cys generally have very similar barriers to oxidation by hydrogen peroxide but are highly sensitive to the external dielectric modeling the effect of solvent. In addition, conformations of the amino acids that allowed for an intermolecular interaction between the peroxide and the NH group of the amino acid significantly reduce the reaction barrier. This indicates that the molecular environment is important for optimizing the efficiency of peroxide reduction with these species and reinforces the role played by the surrounding active site in the natural GPx enzyme as well as the direct interaction with solvent.

Because of the similarity in the overall GPx-like catalytic cycle of most small-molecule GPx mimics much information can be gained from studying a relatively small number of representative species. Many GPx mimics including ebselen are based on an aryl selenol framework and, as such, phenylselenol is a good approximation to most general aryl selenols. Bayse has used phenylselenol to study the effect of the theoretical treatment of the surrounding medium on the catalytic activity of aryl selenols [109]. Most quantum chemical calculations on such systems employ some adaptation of the polarizable continuum model where the effect of solvent is modeled by placing the system in a cavity with an external dielectric constant equal to that of the chosen medium, usually water. The advantage of such treatment is that it is relatively simple and robust; however, this comes at a cost of neglecting potentially important interactions with nearby solvent molecules. Clearly, these interactions are particularly important in the case of reactions where the solvent plays a direct role in facilitating some part of the process, say proton exchange. To capture such interactions, one must explicitly include the solvent in the calculation and so Bayse has constructed a network of water molecules to surround the substrate phenylselenol and each intermediate along its GPx-like catalytic cycle to allow proton exchange during the cycle to occur in a concerted process via the solvent.

These networks were constructed using two, three and four water molecules and each complex and transition state was optimized using the mPW1PW91 functional [42] in conjunction with relativistic effective core potentials. It was shown that when the solvent network facilitates proton exchange, the transition states for each step do not have to adopt the highly constrained geometries necessary for analogous gas-phase processes. As a result, the Gibbs energy barriers are significantly lowered and thus it is evident that explicit solvation is a key factor in obtaining realistic barriers for

the catalytic reduction of peroxides by small-molecule GPx mimics, just as it is for the full GPx enzyme.

20.4

Summary

In this chapter, we have presented a brief review of the quantum mechanical approaches to selenium biochemistry that have appeared over the past two decades. Much of the work focuses on understanding the mechanism of action of GPx and GPx-like small molecules to design potential therapeutic agents such as ebselen, which has been studied in great detail. It has been demonstrated that the role played by the molecular environment in the catalytic cycle is significant both for the entire GPx enzyme as well as for small molecule mimics of GPx.

Theoretical studies of GPx have helped to elucidate details of its structure and have answered questions concerning the presence of water molecules in the active site and what catalytic role these water molecules and neighboring residues of Sec play in the enzyme's function. Specifically, nearby groups help to stabilize transition states and facilitate proton transfer in the reduction of hydrogen peroxide, which prevents the formation of highly strained transition structures. Such activity has also been found for water in theoretical studies of small molecule GPx mimics, indicating a strong contribution from solvent and the need for explicit solvation to achieve reasonable barriers for these processes.

Small molecule GPx mimics may also suffer from thiol exchange reactions whereby the catalytic activity is stalled by a shuttling of thiol substituents. Several theoretical studies have shown potential avenues for overcoming thiol exchange reactions and, therefore, may lead to more effective GPx mimics and therapeutic agents. Such studies have shown that thiol exchange may be prevented through steric interference by substituents at specific sites of the ebselen structure and by electronic modification of the selenosulfide bond by using imidazoles as the catalytic substrate.

References

- 1 Berzelius, J.J. (1818) *Afhandl. Fys. Kemi Mineralogi*, **6**, 42.
- 2 Klayman, D.L. and Guenther, W.H. (eds) (1973) *Organic Selenium Compounds: Their Chemistry and Biology*, Wiley Interscience, New York.
- 3 Löwig, C.J. (1836) *Pogg. Ann.*, **37**, 552.
- 4 Schwarz, K. and Foltz, C. (1957) *J. Am. Chem. Soc.*, **79**, 3292.
- 5 Patai, S. and Rappoport, Z. (eds) (1986) *The Chemistry of Organic Selenium and Tellurium Compounds*, vol. 1, John Wiley & Sons Ltd., Chichester.
- 6 Liotta, D. (ed.) (1987) *Organo-Selenium Chemistry*, John Wiley & Sons Inc., New York.
- 7 Back, T.G. (ed.) (1999) *Organo-Selenium Chemistry: A Practical Approach*, Oxford University Press, Oxford.
- 8 Burk, R.F. (ed.) (1994) *Selenium in Biology and Human Health*, Springer-Verlag, New York.
- 9 Hope, E. and Levason, W. (1993) *Coord. Chem. Rev.*, **122**, 109.
- 10 Bochmann, M. (1996) *Chem. Vap. Deposition*, **2**, 85.

- 11 Flohe, L., Gunzler, W., and Schock, H. (1973) *FEBS Lett.*, **32**, 132.
- 12 Rotruck, J., Pope, A., Ganther, H., Swanson, A., Hafeman, D.G., and Hoekstra, W.G. (1973) *Science*, **179**, 588.
- 13 Davey, J., Becker, K., Schneider, M., Germain, D.L.S., and Galton, V.A. (1995) *J. Biol. Chem.*, **270**, 26786.
- 14 Croteau, W., Whittemore, S., and Schneider, M. (1995) *J. Biol. Chem.*, **270**, 16569.
- 15 Arthur, J.R., Nicol, F., and Beckett, G.J. (1990) *Biochem. J.*, **272**, 537.
- 16 Behne, D., Kyriakopoulos, A., Meinhold, H., and Kohrle, J. (1990) *Biochem. Biophys. Res. Commun.*, **173**, 1143.
- 17 Lescure, A., Gautheret, D., Carbon, P., and Krol, A. (1999) *J. Biol. Chem.*, **274**, 38147.
- 18 Tamura, T. and Stadtman, T.C. (1996) *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 1006.
- 19 Lee, S., Kim, J., Kwon, K., Yoon, H., Levine, R., Ginsburg, A., and Rhee, S.G. (1999) *J. Biol. Chem.*, **274**, 4722.
- 20 Watabe, S., Makino, Y., Ogawa, K., Hiroi, T., Yamamoto, Y., and Takahashi, S.Y. (1999) *Eur. J. Biochem.*, **264**, 74.
- 21 Mustacich, D. and Powis, G. (2000) *Biochem. J.*, **346**, 1.
- 22 Williams, C.H. Jr (2000) *Eur. J. Biochem.*, **267**, 6101.
- 23 Motsenbocker, M.A. and Tappel, A.L. (1984) *J. Nutr.*, **114**, 279.
- 24 Epp, O., Ladenstein, R., and Wendel, A. (1983) *Eur. J. Biochem.*, **133**, 51.
- 25 Rocher, C., Lalanne, J.-L., and Chaudiere, J. (1992) *Eur. J. Biochem.*, **205**, 955.
- 26 Milligan, J.R., Tran, N.Q., Ly, A., and Ward, J.F. (2004) *Biochemistry*, **43**, 5102.
- 27 Mughesh, G., du Mont, W., and Sies, H. (2001) *Chem. Rev.*, **101**, 2125.
- 28 Hinchliffe, A. (1981) *J. Mol. Struct. (THEOCHEM)*, **86**, 189.
- 29 Laitinen, R. and Pakkanen, T. (1983) *J. Mol. Struct. (THEOCHEM)*, **91**, 337.
- 30 Muller, J., Saethre, L.J., and Gropen, O. (1983) *Chem. Phys. Lett.*, **75**, 395.
- 31 Angyan, J.G., Csizmadia, I.G., Daudel, R., and Poirier, R.A. (1986) *Chem. Phys. Lett.*, **131**, 247.
- 32 Balasubramanian, K., Liao, M.Z., and Han, M. (1987) *Chem. Phys. Lett.*, **139**, 551.
- 33 Balasubramanian, K. and Liao, M.Z. (1988) *J. Phys. Chem.*, **92**, 4595.
- 34 Binning, R. Jr and Curtiss, L.A. (1990) *J. Chem. Phys.*, **92**, 1860.
- 35 Binning, R. Jr and Curtiss, L.A. (1990) *J. Chem. Phys.*, **92**, 3688.
- 36 Pearson, J.K., Ban, F., and Boyd, R.J. (2005) *J. Phys. Chem. A*, **109**, 10373.
- 37 Becke, A.D. (1993) *J. Chem. Phys.*, **98**, 1372.
- 38 Lee, C., Yang, W., and Parr, R.G. (1988) *Phys. Rev. B*, **37**, 785.
- 39 Perdew, J.P. and Wang, Y. (1992) *Phys. Rev. B*, **45**, 13244.
- 40 Bayse, C.A. (2005) *J. Chem. Theory Comput.*, **1**, 1119.
- 41 Bayse, C.A. (2004) *Inorg. Chem.*, **43**, 1208.
- 42 Adamo, C. and Barone, V. (1998) *J. Chem. Phys.*, **108**, 664.
- 43 Keal, T.W. and Tozer, D.J. (2004) *J. Chem. Phys.*, **121**, 5654.
- 44 Pryor, W.W. (1982) *Free Radicals in Biology*, vol. 5, Academic Press, New York.
- 45 Sies, H. (1985) *Oxidative Stress*, Academic Press, London.
- 46 Sies, H. (1986) *Angew. Chem.*, **98**, 1061.
- 47 McCord, J.M. and Fridovich, I. (1969) *J. Biol. Chem.*, **244**, 6056.
- 48 Maddipati, K.R. and Marnett, L.J. (1987) *J. Biol. Chem.*, **262**, 17398.
- 49 Stadtman, T.C. (1991) *J. Biol. Chem.*, **266**, 16257.
- 50 Maiorino, M., Aumann, K., Brigelius-Flohe, R., and Doria, D. *et al.* (1995) *Biol. Chem.*, **376**, 651.
- 51 Brigelius-Flohe, R. (1999) *Free Radical Biol. Med.*, **27**, 951.
- 52 Ren, B., Akesson, B., Ladenstein, R., and Huang, W. (1997) *J. Mol. Biol.*, **268**, 869.
- 53 Prabhakar, R., Musaev, D.G., Khavrutskii, I., and Morokuma, K. (2004) *J. Phys. Chem. B*, **108**, 12643.
- 54 Prabhakar, R., Vreven, T., Morokuma, K., and Musaev, D.G. (2005) *Biochemistry*, **44**, 11864.
- 55 Prabhakar, R., Vreven, T., Frisch, M.J., and Morokuma, K. (2006) *J. Phys. Chem. B*, **110**, 13608.
- 56 Roy, G., Nethaji, M., and Mughesh, G. (2004) *J. Am. Chem. Soc.*, **126**, 2712.
- 57 Raes, M., Michiels, C., and Remacle, J. (1987) *Free Radical Biol. Med.*, **3**, 3.
- 58 Michiels, C. and Remacle, J. (1988) *Biochim. Biophys. Acta*, **967**, 341.

- 59 Michiels, C., Toussaint, O., and Remacle, J. (1990) *J. Cell. Physiol.*, **144**, 295.
- 60 Paoletti, R. (ed.) (1994) *Oxidative Processes Antioxidants*, Raven Press Ltd, New York.
- 61 Muller, A., Cadenas, E., Graf, P., and Sies, H. (1984) *Biochem. Pharmacol.*, **33**, 3235.
- 62 Wendel, A., Fausel, M., Safayhi, H., Tiegs, G., and Otter, R. (1984) *Biochem. Pharmacol.*, **33**, 3241.
- 63 Parnham, M. and Kindt, S. (1984) *Biochem. Pharmacol.*, **33**, 3247.
- 64 Muller, A., Gabriel, H., and Sies, H. (1985) *Biochem. Pharmacol.*, **34**, 1185.
- 65 Safayhi, H., Tiegs, G., and Wendel, A. (1985) *Biochem. Pharmacol.*, **34**, 2691.
- 66 Wendel, A. and Tiegs, G. (1986) *Biochem. Pharmacol.*, **35**, 2115.
- 67 Sies, H. (1993) *Free Radical Biol. Med.*, **14**, 313.
- 68 Sies, H. (1994) *Methods Enzymol.*, **234**, 476.
- 69 Schewe, T. (1995) *Gen. Pharmacol.*, **26**, 1153.
- 70 Nakamura, Y., Feng, Q., Kumagai, T., and Torikai, K. *et al.* (2002) *J. Biol. Chem.*, **277**, 2687.
- 71 Zhang, M., Nomura, A., Uchida, Y., and Iijima, H. *et al.* (2002) *Free Radical Biol. Med.*, **32**, 454.
- 72 Lesser, R. and Weiss, R. (1924) *Ber. Dtsch. Chem. Ges.*, **57B**, 1077.
- 73 Zhao, R. and Holmgren, A. (2002) *J. Biol. Chem.*, **277**, 39456.
- 74 Back, T.G. and Moussa, Z. (2003) *J. Am. Chem. Soc.*, **125**, 13455.
- 75 Fischer, H. and Dereu, N. (1987) *Bull. Soc. Chim. Belg.*, **96**, 757.
- 76 Mughesh, G. and Singh, H.B. (2000) *Chem. Soc. Rev.*, **29**, 347.
- 77 Mughesh, G. and du Mont, W.-W. (2001) *Chem.-Eur. J.*, **7**, 1365.
- 78 Reich, H.J. and Jasperse, C.P. (1987) *J. Am. Chem. Soc.*, **109**, 5549.
- 79 Iwaoka, M. and Tomoda, S. (1996) *J. Am. Chem. Soc.*, **118**, 8077.
- 80 Bailly, F., Azaroual, N., and Bernier, J.-L. (2003) *Bioorg. Med. Chem.*, **11**, 4623.
- 81 Wirth, T. (1998) *Molecules*, **3**, 164.
- 82 Wilson, S.R., Zucker, P.A., Huang, R.-R.C., and Spector, A. (1989) *J. Am. Chem. Soc.*, **111**, 5936.
- 83 Mughesh, G., Pa, A., Singh, H., Puneekar, N., and Butcher, R.J. (1998) *Chem. Commun.*, 2227.
- 84 Galet, V., Bernier, J., Henichart, J., and Lesieur, D. *et al.* (1994) *J. Med. Chem.*, **37**, 2903.
- 85 Engman, L., Stern, D., Frisell, H., and Vessman, K. *et al.* (1995) *Bioorg. Med. Chem.*, **3**, 1255.
- 86 Back, T.G. and Moussa, Z. (2002) *J. Am. Chem. Soc.*, **124**, 12104.
- 87 Pearson, J.K. and Boyd, R.J. (2006) *J. Phys. Chem. A*, **110**, 8979.
- 88 Pearson, J.K. and Boyd, R.J. (2007) *J. Phys. Chem. A*, **111**, 3152.
- 89 Peng, C., Ayala, P., Schlegel, H.B., and Frisch, M.J. (1996) *J. Comput. Chem.*, **17**, 49.
- 90 Peng, C. and Schlegel, H.B. (1993) *Israel J. Chem.*, **33**, 449.
- 91 Gonzalez, C. and Schlegel, H.B. (1990) *J. Phys. Chem.*, **94**, 5523.
- 92 Gonzalez, C. and Schlegel, H.B. (1989) *J. Chem. Phys.*, **90**, 2154.
- 93 Barone, V. and Cossi, M. (1998) *J. Phys. Chem. A*, **102**, 1995.
- 94 Morgenstern, R., Cotgreave, I., and Engman, L. (1992) *Chem.-Biol. Interact.*, **84**, 77.
- 95 Sarma, B.K. and Mughesh, G. (2005) *J. Am. Chem. Soc.*, **127**, 11477.
- 96 (a) Bader, R.W.F. (1990) *Atoms in Molecules: A Quantum Theory*, Oxford University Press, Oxford, (b) Matta, C.F. and Boyd, R.J. (eds) (2007) *The Quantum Theory of Atoms in Molecules: From Solid State to DNA Drug Design*, Wiley-VCH Verlag GmbH, Weinheim.
- 97 Bhabak, K.P. and Mughesh, G. (2007) *Chem.-Eur. J.*, **13**, 4594.
- 98 Foresman, J., Keith, T., Wiberg, K., Snoonian, J., and Frisch, M.J. (1996) *J. Phys. Chem.*, **100**, 16098.
- 99 Wolinski, K., Hinton, J.F., and Pulay, P. (1990) *J. Am. Chem. Soc.*, **112**, 8251.
- 100 Reed, A.E., Curtiss, L.A., and Weinhold, F.A. (1988) *Chem. Rev.*, **88**, 899.
- 101 Bachrach, S.M., Demoin, D.W., Luk, M., and Miller, J.V. Jr (2004) *J. Phys. Chem. A*, **108**, 4040.
- 102 Parnham, M.J., Biedermann, J., Bittner, C., and Dereu, N. *et al.* (1989) *Agents Actions*, **27**, 306.
- 103 Pearson, J.K. and Boyd, R.J. (2008) *J. Phys. Chem. A*, **112**, 1013.

- 104 Roy, G. and Mugesh, G. (2005) *J. Am. Chem. Soc.*, **127**, 15207.
- 105 Soujanya, Y. and Sastry, G.N. (2007) *Tetrahedron Lett.*, **48**, 2109–2112.
- 106 Cardey, B. and Enescu, M. (2005) *ChemPhysChem*, **6**, 1175.
- 107 Cardey, B. and Enescu, M. (2007) *J. Phys. Chem. A*, **111**, 673.
- 108 Svensson, M., Humbel, S., and Morokuma, K. (1996) *J. Chem. Phys.*, **105**, 3654.
- 109 Bayse, C.A. (2007) *J. Phys. Chem. A*, **111**, 9070.

21

Catalytic Mechanism of Metallo β -Lactamases: Insights from Calculations and Experiments

Matteo Dal Peraro, Alejandro J. Vila, and Paolo Carloni

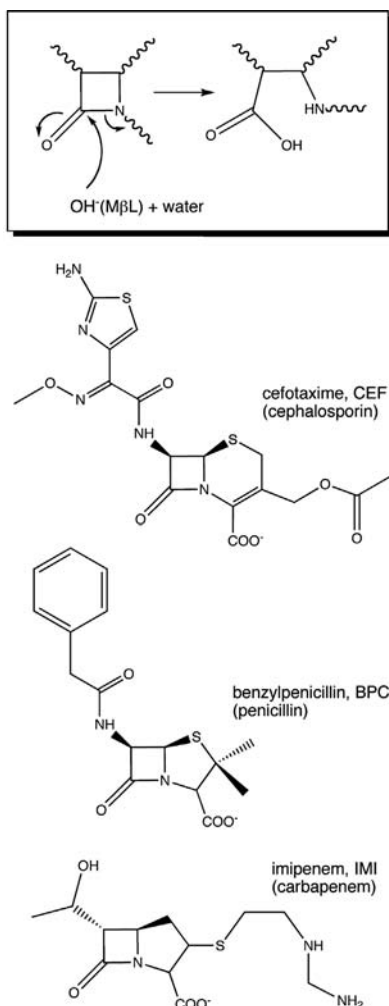
21.1

Introduction

β -Lactam antibiotics are the most widespread antibiotics on the market nowadays. They have exerted great evolutionary pressure on bacteria, triggering sophisticated resistance mechanisms. Among them, the most widespread is the expression of β -lactamases. This is a family of hydrolases that uses different protein scaffolds and catalytic architectures to inactivate β -lactam drugs (Scheme 21.1) [1].

β -Lactamases from classes A, C and D are serine hydrolases, whereas class B β -lactamases are characterized by the presence of Zn ions bound to their active sites [2]. Metallo β -lactamases (M β LS), despite not being as ubiquitous as serine β -lactamases, represent the largest group of carbapenemases, that is, they hydrolyze a very important class of antibiotics, the carbapenems. M β LS are increasingly spreading among pathogenic bacteria in the clinical setting and are resistant to all current clinical inhibitors on the market [3–6]. Thus, understanding their function at the molecular level is of paramount importance for designing effective drugs (either M β L inhibitors or antibiotics refractory to hydrolysis by M β LS).

M β LS are classified by sequence homology in three subclasses: B1, B2 and B3 [4–7]. The B1 subclass includes several chromosomally-encoded enzymes, such as the ones from *Bacillus cereus* (BcII) [8–13], *Bacteroides fragilis* (CcrA) [14–16], *Elizabethkingia meningoseptica* (BlaB) [17, 18], as well as the transferable VIM, IMP, SPM and GIM-type enzymes [19–25]. Subclass B2 includes the CphA and ImiS lactamases from *Aeromonas* species, and Sfh-I from *Serratia fonticola* [26–28]. Subclass B3, along with the extensively characterized enzyme L1 from *Stenotrophomonas maltophilia* [29–32], includes enzymes from environmental bacteria, such as CAU-1 from *Caulobacter crescentus* [33], and THIN-B from *Janthinobacterium lividum* [34], and from opportunistic pathogens like FEZ-1 from *Legionella gormanii* [35–37], and GOB from *E. meningoseptica* [38]. B1 and B3 enzymes display a broad substrate spectrum, being able to actively hydrolyze β -lactams of all three



Scheme 21.1 Metallo β -lactamase hydrolysis of a general β -lactam, and β -lactam substrates representative of major antibiotic families.

β -lactam families (e.g., penicillins, cephalosporins and carbapenems), whilst B2 M β Ls are selective carbapenemases.

The metal ions are essential for hydrolysis: Fast mixing techniques coupled to Trp fluorescence have shown that apo-BcII, despite being properly folded, is unable to bind any substrate [13]. Unfortunately, characterization of the catalytic mechanism in Zn-enzymes by means of experimental techniques is difficult: they are silent to most spectroscopic techniques, in contrast to transition metal-ion based enzymes.

Computational quantum mechanical methods may provide a full structural and energetics description of the reactive mechanism, along with the metal polyhedron reorganization, providing a valuable instrument to dissect enzymatic catalysis.

Here we discuss the enzymatic reaction mechanism of M β Ls from both a computational and experimental point of view. We mostly focus on subclass B1, which represents to date the best structurally and functionally characterized M β L species [6, 8, 10–12, 14–16, 28, 32, 39–43], based on its clinical relevance [5].

21.2

Structural Information

Atomistic structures of M β Ls from the three subclasses have been solved by X-ray crystallography (B1: BcII [43–45], BlaB [17], CcrA [46], IMP-1 [47], SPM-1 [48], VIM-2 (B1); B2: CphA [26]; and B3: L1 [30] and FEZ-1 [36]). The folding frame is characterized in all cases by a compact $\alpha\beta/\beta\alpha$ sandwich, accommodating an active site that can allocate one or two Zn ions [3, 4]. At the active pocket, the first metal site (Zn1) is coordinated in a tetrahedral geometry to three histidine ligands (His116, His118 and His196), and the nucleophilic hydroxide in B1 and B3 enzymes, thus being called the 3H site (Figure 21.1). The coordination of the second metal site (Zn2) is provided by the nucleophile, one water molecule and a ligand triad of protein residues that includes Asp120, His263 and a third residue, Cys221 (subclasses B1 and B2) or His121 (subclass B3). This site is called the DCH (B1, B2) or DHH site (B3) (Figure 21.1). B2 enzymes are active only with the metal ion at the DCH site, while binding of a second Zn ion inhibits the enzyme activity [49, 50]. (The identity of this inhibitory site is still matter of debate.)

The metal ion occupancy in B1 enzymes has been also a matter of debate. The first crystal structure of BcII (at 2.5 Å, PDB code 1bmc [45]) showed one zinc ion bound at the 3H site. Subsequent structures of BcII and CcrA revealed a conserved dinuclear

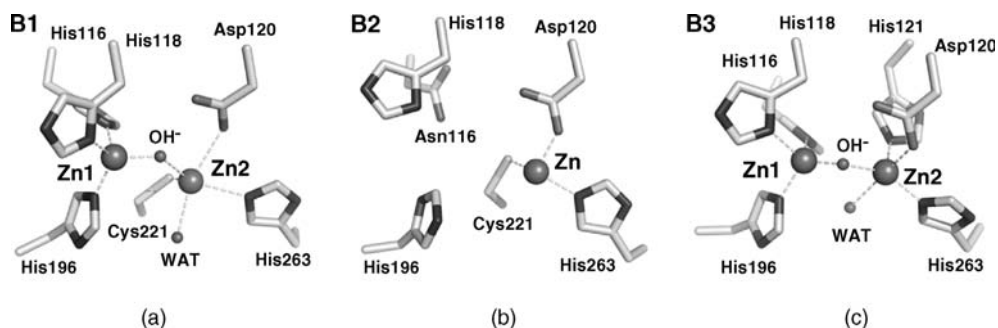


Figure 21.1 Metal site conformations in the M β L family. Representatives of (a) B1 subclass (di-Zn CcrA from *Bacteroides fragilis*), (b) B2 subclass (mono-Zn DCH CphA from *Aeromonas hydrophila*) and (c) B3 subclass (di-Zn FEZ-1 from *Legionella gormanii*).

metal center in all B1 enzymes [8, 44, 46], with the conserved Asp120-Cys221-His263 triad (DCH) binding the second Zn ion (Zn2, Figure 21.1a).

21.3

Computational Details

We briefly report the computational protocol generally adopted to study the reactivity of enzyme–substrate adducts (more detailed information can be found in the original papers and recent reviews on the subject) [51–58]. Substrates (e.g., in this case benzylpenicillin, cefotaxime, etc., Scheme 21.1) are docked at the active site of the M β L, based on the X-ray structure at the highest resolution. Particular attention is paid to the definition of the protonation state of active site residues, which plays a crucial role in binding and catalysis. Large-size DFT calculations have usually been used to determine the most plausible protonation state of the catalytic pocket, based on existing X-ray structures [52, 58]. The optimum protonation states are therefore identified based on the lowest energy protomers displaying small RMSD with the experimental structures. Based on this structural information, force field parameters for substrates and the metal coordination shell are usually developed *ad hoc* for the system of interest, using reported procedures to covalently treat metal–ligand interactions [54].

Adducts are immersed in a water box neutralized by counter-ions, which also reproduce physiological salt concentration, before undergoing MD simulations (the AMBER force field is used for the protein, and a TIP3P model is used for water molecules). MD simulations on the multi-nanosecond timescale are required to sample the thermal fluctuations at the active pocket that can be functional to the catalytic mechanism. Conformations differing by relevant structural determinants (e.g., substrate–nucleophile distances at the active site), which can identify stable states of the complex, are selected for subsequent QM/MM simulations. The reaction pathway for each of them is investigated by first principles MD in a fully Hamiltonian QM/MM scheme [59], which combines Car–Parrinello [60], DFT MD and classical force-field based MD. Although first principles calculations access usually short timescales (generally on the order of 10^2 ps) they offer the advantage of being transferable. Within the Car–Parrinello scheme statistical mechanics approaches to simulate rare events may be used to investigate reactions in biologic systems. The free energy of activation for the catalytic mechanism is evaluated in this case with the method of thermodynamics integration [61], using the distance between the nucleophilic O atom (on OH[−]) and carbonyl C atom (on the substrate β -lactam ring) as reaction coordinate (RC). The RC is gradually shortened (by 0.1/0.2 Å per step) to search for critical points in the one-dimensional free energy surface; forces on the constraint are averaged for every step, and then integrated to provide the free energy profile.

The free energy of activation can be related to experimental k_{cat} values through transition state theory. Estimation of ΔF provides, however, an approximate value of the catalytic rate due to problems related with DFT accuracy and sampling (amplify

discussed in Reference [62]), and because of the inherent limitations of rate constants definition.

21.4

Preliminary Comment on the Comparison between Theory and Experiment

The literature is plagued with misunderstandings between research groups studying an enzymatic reaction by means of an experimental approach (either structural, kinetic or spectroscopic) and those performing calculations. These problems are largely due to incomplete knowledge of the assumptions, limitations and/or short-cuts inherent in the experiments or the calculations themselves. In this joint effort, we examine some of these problems in the context of B1 MβLs:

- 1) The metal occupancy in MβLs might not be adequately addressed by the X-ray structures. In fact, partial occupancies can be described by large B factors when the resolution is not high enough, and metal sites partially populated can be modeled as solvent molecules [45, 63]. In addition, in B1 MβLs, the metal ion distribution between different sites at sub-stoichiometric metal ion concentrations differs from that observed in solution [64, 65]. Thus, the choice of the structure used as templates for sophisticated calculations should be a major concern.
- 2) The calculated activation free energy is usually compared with that matching experimental k_{cat} values. However, in multistep reactions, this comparison might be misleading. Calculated free energies usually stand for individual, mechanistic steps, usually accounting for the rate-determining step of the reaction. However, individual rate constants are not always available in the literature (nor easy to determine in all cases) and comparisons with the steady-state turnover number may not be the best way to validate the outcome of the calculations. One option to circumvent this problem is to identify experimentally the rate-determining step of the reaction. In the case of B1 MβLs, there are two possible rate-determining steps: nucleophilic attack by a metal-bound hydroxide or the protonation step (Scheme 21.1). Solvent kinetic isotope or proton inventory experiments have shown that the rate-determining step involves the protonation of the bridging nitrogen (Scheme 21.1) for all MβLs [10, 14, 16, 66, 67]. Kinetic information regarding the rate-limiting step ought to be considered when comparing calculations with experiments.
- 3) Substrate binding is usually (poorly) described by K_M values, that is, apparent dissociation constants, while data on K_S are less common in the literature [31, 42]. This may limit obtaining an accurate description of the affinity of the enzyme towards substrates. This is particularly an issue in the case of MβLs, which are multi-substrate enzymes. Fortunately, pre-steady state kinetics following intrinsic Trp fluorescence of the protein allow direct determination of K_S values for several substrates in different MβLs. In addition, these data should be considered in theoretical studies.
- 4) The spectroscopic trapping of intermediates is of great help to probe enzymatic reactions. In the case of Zn enzymes such as MβLs one can use suitable spectro-

scopic probes, such as Co(II) or Cd(II) ions. In addition, spectroscopic techniques following the fate of the substrate can be employed [16, 66, 68]. Thus, the identity of chemical species predicted by state-of-the-art calculations can be ascertained or validated against data emerging from experimentally trapped intermediates.

21.5

Michaelis Complex in B1 M β LS

21.5.1

Substrate Binding Determinants

Molecular dynamics simulations of mono-Zn and di-Zn B1 M β LS in complex with different types of β -lactams (e.g., benzylpenicillin, imipenem and cefotaxime) pointed to few crucial interactions for binding recognition across subclass B1 (Scheme 21.1, Figure 21.2), regardless of metal content (i.e., in both 3H mono-Zn BcII from *Bacillus cereus* and di-Zn CcrA from *Bacteroides fragilis*) [53, 54, 56].

- 1) A water-mediated electrostatic interaction is present between the β -lactam carboxylate group and the metal center (WAT in ES in Scheme 21.2 and Figures 21.2 and 21.3). WAT H-bonds to the β -lactam carboxylate moiety; it is highly persistent within the catalytic pocket, and it is spatially conserved in mono- and di-zinc systems and functional to binding recognition pattern found for different β -lactams in the pocket (Figure 21.2).
- 2) A water-mediated salt-bridge is maintained between the β -lactam carboxylate moiety and Lys224 (conserved in most B1 enzymes; Scheme 21.2 and Figure 21.2). Notice that the presence of this carboxylate is not sufficient for efficient binding to M β LS: monocyclic β -lactams possessing the carboxylate moiety are not hydrolyzed by BcII and do not even bind to its active site [69]. This reveals that the bicyclic ring of penicillins, cephalosporins and carbapenems are required for the binding event, regardless of their different side-chain decorations. Consistently, simulations point to several, non-specific contacts between the enzyme active site and the different substituents present in the antibiotics.
- 3) An H-bond is present between β -lactam carbonyl and residues in a loop flanking the active site groove (i.e., Asn233). This is the weakest electrostatic interaction, and may not be effective in some cases.

It is striking that these common minimal features are sufficient to accommodate different β -lactams, producing a productive conformation for the enzymatic reaction in all complexes; indeed, the putative reaction coordinate (i.e., the distance between the nucleophilic carbonyl oxygen and β -lactam carbonyl carbon) ranges from 3.2 to 3.5 Å, giving plausible models for the Michaelis complexes [53, 56]. In addition, these structural determinants align the β -lactam ring parallel to the Zn1-OH-(Zn2) plane in the mono-Zn (di-Zn) species. Finally, in such complexes, the HOMO of the nucleophilic agent (the OH⁻ group) and the LUMO localized on the β -lactam ring are correctly aligned for the nucleophilic attack [52, 53, 56].

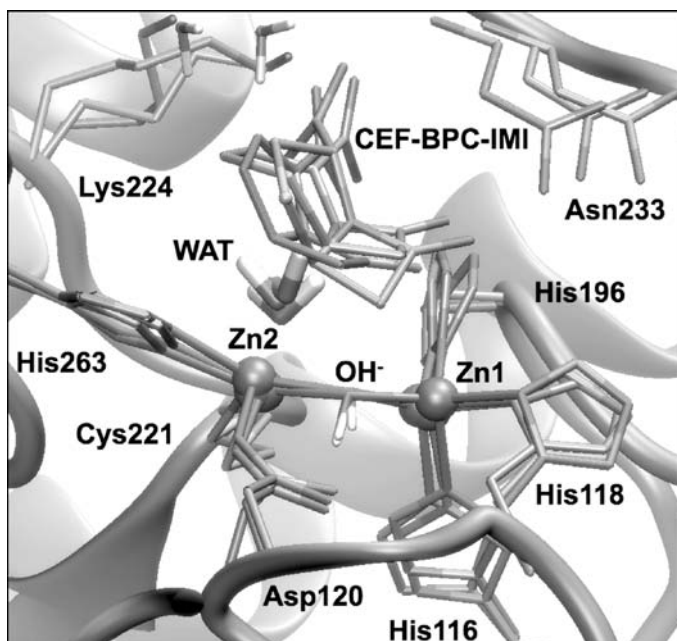
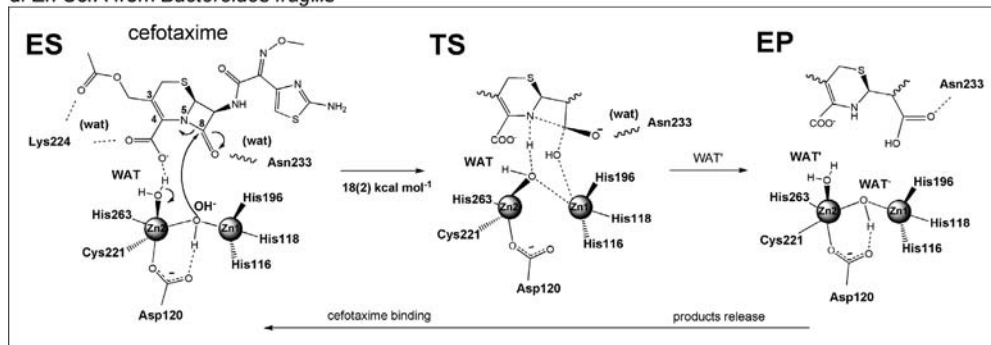
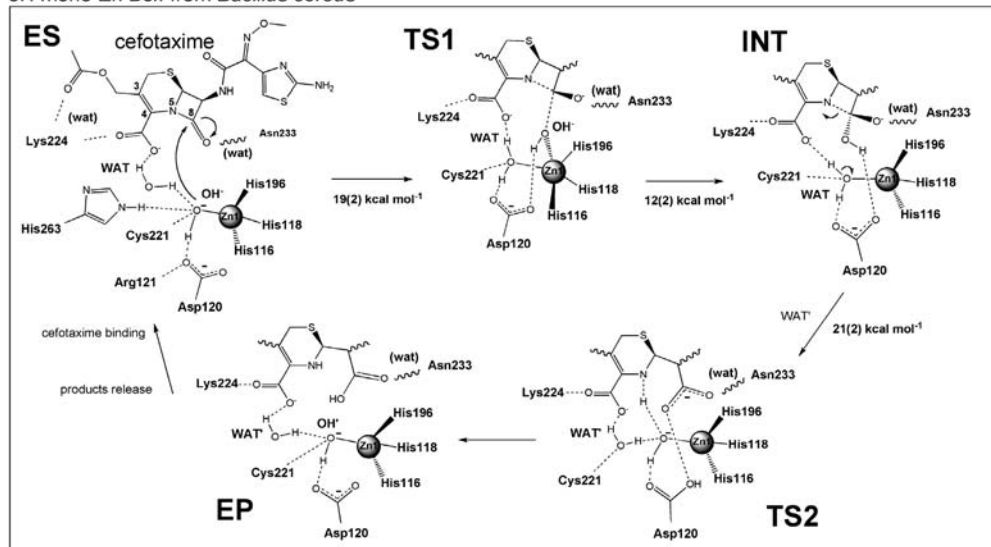


Figure 21.2 MD structural insights into the binding mode of di-Zn MβL CcrA in complex with benzylpenicillin, cefotaxime and imipenem β-lactam substrates (for clarity, only substrate bicyclic cores and residue heavy atoms are shown; see also Schemes 21.1 and 21.2) [56].

21.5.2

Nucleophile Structural Determinants

Classical MD and hybrid QM/MM simulations allow us to identify similar features on passing from the 3H mono-Zn (BcII) to the di-Zn species (CcrA): (i) in CcrA, the Asp120-Cys221-His263 triad binds the second Zn, which coordinates the nucleophilic OH^- and the water molecule, WAT; (ii) in BcII the same triad together with Arg121 forms an highly organized H-bond network that appears to be functionally equivalent, as it similarly orients the nucleophile and WAT for the reaction (Figure 21.1, Scheme 21.2). Thus, these residues appear to be modularly exploited according to the zinc content, to preserve the Michaelis complex architecture. In addition, the overall effective charge at the active site is preserved through point mutations: the charge of the second zinc ion in the di-Zn species is replaced in monozinc by the presence of Cys221 (in the neutral state), and Arg121, which forms a strong H-bond with Asp120.

di-Zn CcrA from *Bacteroides fragilis*3H mono-Zn BcII from *Bacillus cereus*

Scheme 21.2 Di-Zn and 3H mono-Zn reaction mechanisms in B1 M β LS that emerged from QM/MM studies [53, 56] (see also Figures 21.2 and 21.3).

21.6

Catalytic Mechanism of B1 M β LS

We used hybrid QM/MM calculations within the DFT/BLYP framework to investigate the hydrolysis of a commonly used cephalosporin (cefotaxime), which is actively degraded either by mono-Zn and di-Zn species [53, 56]. We summarize here our findings for both the possible conformations of the active site as suggested by the crystallographic data available to date. BcII from *B. cereus* shows that the metal ion is bound in the 3H zinc site, whereas CcrA from *B. fragilis* has the 3H and DCH sites both occupied (Figures 21.2 and 21.3, Scheme 21.2). As emerged from calculations, both the enzymes can (i) promote the nucleophilic attack by a metal-bound hydroxide

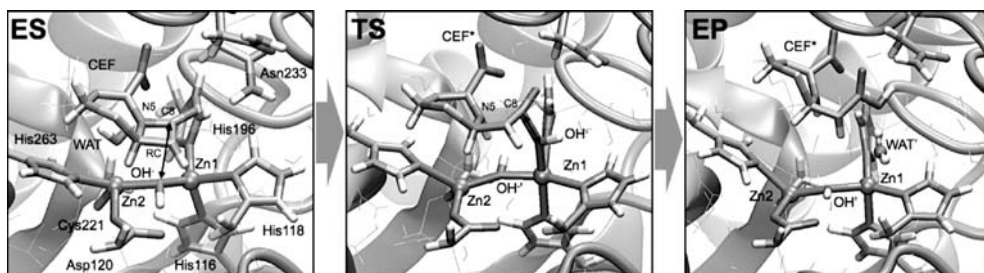


Figure 21.3 QM/MM structural insights into the reaction mechanism of di-Zn MβL CcrA from *B. fragilis* (atoms shown in ball- and -stick representation are included in the QM cell, see also Schemes 21.1 and 21.2) [56].

and (ii) catalytically activate a water (WAT) as a proton shuttle on nitrogen (N5) that finally triggers C–N cleavage in the β-lactam ring. Nonetheless, the chemistry and kinetics of the two reactions strongly depends on the zinc architecture and content (Figure 21.3).

21.6.1

Cefotaxime Enzymatic Hydrolysis in CcrA [56]

In this di-zinc MβL, the nucleophile OH^- approaches the β-lactam carbonyl carbon, maintaining its coordination with the metal ions. The whole adduct is not relevantly affected by any structural changes until the RC reaches the transition state (TS) (i.e., $\text{RC} = 2.0 \text{ \AA}$). At the TS a cascade of almost simultaneous events occurs (Figures 21.2 and 21.3 and Scheme 21.2): (i) OH^- moves out from the Zn1–Zn2 plane, attacking C8 while the Zn1– OH^- bond is being weakened ($2.2 \text{ \AA}$). (ii) The OH^- –Zn2 bond is lost upon nucleophilic attack, and the Zn1–Zn2 distance increases ($3.8 \text{ \AA}$). A hydroxide simultaneously bound to Zn1 and Zn2 is in fact expected to be a poor nucleophile, and thus it is reasonable that OH^- attacks while detaching from at least one Zn ion. As a consequence, the Zn2 coordination number changes from 5 to 4. (iii) The WAT ligand bonds tightly to Zn2, forming a bipyramidal polyhedron. The WAT–Zn2 distance gradually decreases, thus lowering the WAT pK_a . (iv) The Zn2-bound WAT gets closer to the β-lactam ring, consequently increasing the C–N distance. The partial negative charge on N together with the enhanced nucleophilicity of WAT produces a proton shuttle from WAT to the N5 atom that finally triggers the C8–N5 bond cleavage. (v) Deprotonated water binds Zn1, completing the tetrahedral coordination sphere of Zn2 and perfectly replacing the position of the OH^- nucleophile in ES state. Zn1 at this point switches to a pentacoordinated bipyramidal coordination from the initial tetrahedral one.

When the constraint of RC is released, the system evolves onto the products state (EP, Figure 21.3, Scheme 21.2), where OH^- reorients H-bonding to Asp120 and bridging the Zn1 and Zn2 ions as in the ES state. The CEF substrate is completely

hydrolyzed, the β -lactam ring is open, C8 acquires planar sp^2 hybridization, and it is finally detached from the metal center.

The estimated activation free energy for the single-step mechanism is $\Delta F = 18$ (2) kcal mol^{-1} (1 $\text{kcal} = 4.184 \text{ kJ}$), which is consistent with experimental evidence that suggests a single-step reaction for cephalosporin hydrolysis, and a free energy of activation of about 17 kcal mol^{-1} for similar reactions. The efficiency of the reaction mechanism seems to be modulated by second shell ligand residues: Asn233 H-bonds to the β -lactam carbonyl during turnover. This interaction fluctuates on the nanosecond timescale, as shown in MD simulations, and when it is not present the catalytic mechanism becomes impaired, showing a free energy of activation inconsistent with kinetic data ($\sim 30 \text{ kcal mol}^{-1}$). Moreover, this leads to the formation of a high-energy intermediate in which the WAT molecule is expelled from the active site and the β -lactam carboxylate moiety is directly bonded to the Zn2 in DCH site. This conformation is very similar to crystallographic results obtained by product inhibition in CphA (B2) and L1 (B3) enzymes, in which the substrate is trapped in a stable state where the carboxylate moiety is covalently bonded to Zn2. This further confirms the importance of the catalytic water WAT for the efficient progress of a single-step mechanism leading to β -lactam hydrolysis.

21.6.2

Cefotaxime Enzymatic Hydrolysis in BcII [53]

In 3H mono-Zn BcII, Zn1 also undergoes a large rearrangement of its coordination shell upon nucleophilic attack. When the hydroxyl nucleophile attacks the β -lactam carbonyl carbon, Zn1 distorts its polyhedron and coordinates WAT as fifth ligand. This leads to the formation of a stable high-energy intermediate state [INT, $12(2) \text{ kcal mol}^{-1}$, Scheme 21.2]. Thus, the presence of only one Zn equivalent in the 3H site forcedly leads to a two-step mechanism in BcII, so that WAT first replaces the nucleophile during the first step entering the Zn1 coordination shell, and then is activated in the second step as proton donor for β -lactam N5 (the latter being the rate limiting step of the reaction). In CcrA, the presence of Zn2 merges these two movements in a concerted unique step: Zn2 already activates WAT as soon as the OH–Zn1 bond is broken upon the nucleophilic attack. This also explains the improved catalytic efficiency of di-Zn versus 3H mono-Zn variants, as emerged from the calculation of the free energy of activation for di-Zn ($\Delta F = 18 \text{ kcal mol}^{-1}$) and mono-Zn ($\Delta F = 21 \text{ kcal mol}^{-1}$) species.

Recent mechanistic studies on BcII have shown the accumulation of an active mono-Zn variant during turnover that, in contrast to the X-ray structure, displays the metal ion localized in the DCH site [68, 70]. Thus, the active site would resemble the conformation found in mono-Zn B2 enzymes, where the DCH site holds the zinc ion and nucleophile activation is not metal-mediated. Unfortunately, the lack of structural data for such a conformation may hamper QM/MM investigations of this alternative pathway.

21.6.3

Zinc Content and Reactivity of B1 MβLs

Di-Zn CcrA showed a highly concerted single-step mechanism in which the role of the two metal ion is crucial. The coordination numbers of Zn1 and Zn2 change between 4 and 5 (Figure 21.3, Scheme 21.2): in particular at the transition state (TS) Zn1 is able to accept a fifth ligand (WAT) upon the nucleophilic attack, distorting the initial tetrahedral polyhedron to a trigonal bipyramidal coordination. Conversely, Zn2 simultaneously switches from 5 to 4 ligands: from bipyramidal to tetrahedral once the nucleophile is detached, and promotes the activation of a water molecule (WAT) to protonate the bridging N atom. After the products are released, Zn1 reacquires the tetrahedral coordination, whereas Zn2 can eventually switch back to the initial coordination once water molecules have completely diffused at the free active pocket.

These results highlight, especially by comparison to the mechanism proposed for mono-Zn BcII, the role of the Zn2 site in promoting protonation of the bridging nitrogen, which kinetic studies indicate to be the rate-determining step [10, 14, 16, 67]. A series of mutagenesis studies in di-Zn BcII confirms this view, since engineering a more buried position of the Zn2 site gives rise to an inactive β-lactamase, despite being dinuclear [71]. Instead, an optimized BcII obtained by *in vitro* evolution has been shown to be more efficient by the action of a second-shell ligand, which optimizes the position of Zn2 for the protonation step [72]. These studies provide an excellent example of closely related theoretical and experimental studies in a metalloenzyme, and point to an important and more flexible role of the DCH metal site in B1 species.

21.6.4

Reactivity of β-Lactam Antibiotics other than Cefotaxime

Because the zinc ligands and most active site residues are highly conserved among subclass B1 MβLs, the binding and mechanism found for CcrA and BcII may be likely for the entire B1 subclass, where β-lactams might follow similar catalytic pathways as all the groups involved in the catalysis do not depend on the substrate chemical diversity. Of course, slightly different activation free energies are expected, because of the substrate binding modes generated by different β-lactam substituents and non-identical long-range electrostatics of the MβL scaffold. This is also indirectly confirmed by the results obtained from classical MD simulations of different β-lactam antibiotics docked at the BcII and CcrA binding pocket (namely, imipenem: IMI and benzylpenicillin: BPC, Figure 21.2) [53, 54, 56]. The reaction coordinates for these additional adducts are always maintained relatively short [3.2(2) Å for IMI binding and 3.5(2) Å for BPC]. This is consistent with the fact that both are efficiently hydrolyzed by CcrA and BcII. The key interactions of the bicyclic core at the catalytic cleft are still preserved. More importantly, the catalytic water molecule, WAT is buried at the cleavage site, in an equivalent position as found for the adducts complexed with cefotaxime.

Thus, these common elements are flexible enough to accommodate different substrates and elicit a broad spectrum activity in M β LS, based on a similar water-assisted hydrolysis mechanism generally plausible for penicillins, cephalosporins and carbapenems. They might also provide a rationale as to why monobactams such as aztreonam, lacking the common β -lactam bicyclic core and carboxylate, are not efficiently hydrolyzed by M β LS.

21.7

Michaelis Complexes of other M β LS

In this section we discuss briefly the other two M β L subclasses, which share facets similar to those of the B1 subclass.

21.7.1

B2 Mono-Zn M β L Subclass

B2 M β LS, such as CphA, ImiS and Sfh-I, are active only in the mononuclear form, with Zn in the DCH site. Thus, these enzymes are meant to provide a non-metal activated nucleophile. Recent computational studies on substrate binding have shown indeed that the DCH site can orient two water molecules in the pocket [55]. Most importantly, as found for the role of WAT in di-Zn or mono-Zn B1 M β LS, one of the water anchors the β -lactam carboxylate to the metal center (Figure 21.2). This finding is in contrast with the proposal that the conserved β -lactam carboxylate group is covalently bonded to the metal ion (present in either the DCH or DHH sites). Our MD investigations consistently indicate that this interaction is always (i.e., in B1, B2 and B3 M β LS) mediated by a conserved water molecule (WAT). This in turn provides an optimal orientation of the substrate in the active site for the hydrolysis reaction, playing a crucial role in the reaction mechanism, as we have seen in the previous section.

21.7.2

B3 M β L Subclass

In B3 M β LS, where a His group usually substitutes Cys221, and the enzyme is commonly active in the di-Zn conformation, preliminary MD modeling of FEZ-1 binuclear adducts indicates a binding mode similar to CcrA and BcII systems. Furthermore, the presence of a water molecule equivalent to the catalytic water found in B1 subclass might in turn suggest that a water-assisted pathway is also shared with binuclear B3 enzymes [73].

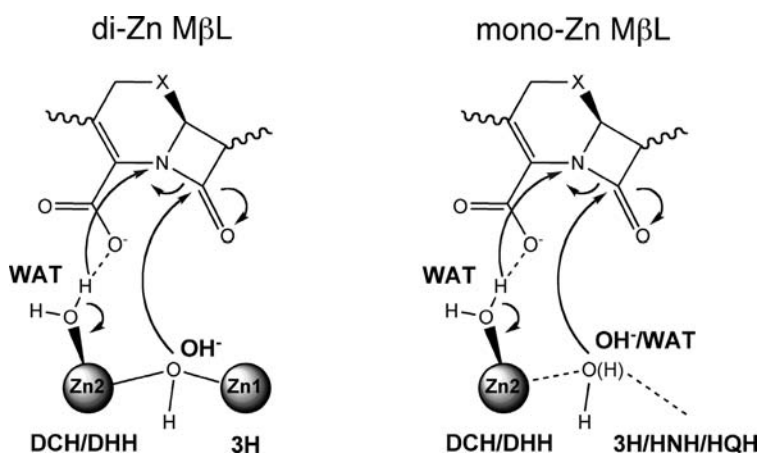
Moreover, recent studies suggest that representatives of B3 subclass, such as GOB M β L, can be active as a DHH mono-Zn enzyme, resembling in principle the active site architecture of B2 enzymes. Even if these cases may follow a different mechanism, these experimental data, together with computational work on B1 subclass, strongly suggest a central role of the DCH/DHH site in M β L-mediated catalysis, and call for more detailed studies in this direction.

21.8

Concluding Remarks

Our studies on MβLs' Michaelis complexes allow us to suggest that B1 and B3 subclass active sites are broad grooves almost devoid of protein elements designed to recognize a substrate, except for very few interactions present across different β-lactam antibiotics. This is fully consistent with the fact that B1 and B3 MβLs are multi-substrate enzymes [66]. Species belonging to B2 subclass present a much narrower binding pocket, specialized to recognizing only carbapenems, but seem to share conserved minimal features for promoting catalysis. In fact, a Zn-bound water molecule (WAT in Figures 21.2 and 21.3, and Schemes 21.2 and 21.3) turns out to be a common and crucial chemical feature across all MβLs: upon nucleophilic attack this water molecule can readily protonate the β-lactam nitrogen for an efficient cleavage of the substrate. The β-lactam carboxylate moiety, present in all β-lactam antibiotics actively hydrolyzed by MβLs, stabilizes WAT at the active site upon binding, so that the water/β-lactam entity should be considered as the favorite template for the design of new inhibitors.

Our investigations of the hydrolytic mechanism in B1 MβL subclass, combined with a wealth of experimental results, point to the crucial role of Zn in the second metal-binding site (i.e., DCH or DHH). This ion stabilizes the negative charge developed at β-lactam nitrogen upon nucleophilic attack and its detachment from the second zinc. This functional advantage is evident in the mechanism of di-Zn B1 MβLs, and is completely missing if one considers the mechanism of mono-Zn 3H enzymes, which are characterized by low-efficiency turnover and a step-wise mechanism. This suggests, along with recent kinetic findings [70], that in low-concentrations of zinc equivalent a more efficient catalysis could be achieved across all MβL subclasses when the metal ion is accommodated in the DCH/DHH site (Scheme 21.3).



Scheme 21.3 Role of DCH/DHH-bound Zn₂ in di-Zn and mono-Zn based MβL hydrolysis.

Another key ingredient that emerges for the enzymatic catalysis across the B1 subclass includes the flexibility of the zinc coordination polyhedra: The hydrolysis can be catalyzed only by conformational changes at the coordination metal sphere. In particular, in the binuclear enzyme the zinc ions rearrange their ligands in a concerted way to activate the reactive moieties – the nucleophilic hydroxide and the catalytic water.

In conclusion, transition state analogues could mimic these common minimal determinants developed at the active site during the reaction. Specifically, inhibitors targeting the B1 subclass should favor and promote a rearrangement of the metal sphere as seen during the transition state, inducing a similar shifting of zinc coordination number upon binding.

References

- 1 Fisher, J.F., Meroueh, S.O., and Mobashery, S. (2005) Bacterial resistance to beta-lactam antibiotics: compelling opportunism, compelling opportunity. *Chem. Rev.*, **105** (2), 395–424.
- 2 Hall, B.G. and Barlow, M. (2003) Structure-based phylogenies of the serine beta-lactamases. *J. Mol. Evol.*, **57** (3), 255–260.
- 3 Galleni, M., Lamotte-Brasseur, J., Rossolini, G.M., Spencer, J., Dideberg, O., Frere, J.M., and Grp, M.-B.-L.W. (2001) Standard numbering scheme for class B beta-lactamases. *Antimicrob. Agents Ch.*, **45** (3), 660–663.
- 4 Garau, G., Di Guilmi, A.M., and Hall, B.G. (2005) Structure-based phylogeny of the metallo-beta-lactamases. *Antimicrob. Agents. Ch.*, **49** (7), 2778–2784.
- 5 Walsh, T.R., Toleman, M.A., Poirel, L., and Nordmann, P. (2005) Metallo-beta-lactamases: the quiet before the storm? *Clin. Microbiol. Rev.*, **18** (2), 306.
- 6 Crowder, M.W., Spencer, J., and Vila, A.J. (2006) Metallo-b-lactamases: novel weaponry for antibiotic resistance in bacteria. *Acc. Chem. Res.*, **46** (10), 721–728.
- 7 Hall, B.G. and Barlow, M. (2005) Revised Ambler classification of beta-lactamases. *J. Antimicrob. Chemother.*, **55** (6), 1050–1051.
- 8 Orellano, E.G., Girardini, J.E., Cricco, J.A., Ceccarelli, E.A., and Vila, A.J. (1998) Spectroscopic characterization of a binuclear metal site in *Bacillus cereus* beta-lactamase II. *Biochemistry*, **37** (28), 10173–10180.
- 9 Prosperi-Meys, C., Wouters, J., Galleni, M., and Lamotte-Brasseur, J. (2001) Substrate binding, and catalytic mechanism of class B beta-lactamases: a molecular modelling study. *Cell Mol. Life Sci.*, **58** (14), 2136–2143.
- 10 Bounaga, S., Laws, A.P., Galleni, M., and Page, M.I. (1998) The mechanism of catalysis and the inhibition of the *Bacillus cereus* zinc-dependent beta-lactamase. *Biochem. J.*, **331**, 703–711.
- 11 Rasia, R.M. and Vila, A.J. (2002) Exploring the role and the binding affinity of a second zinc equivalent in *B. cereus* metallo-beta-lactamase. *Biochemistry*, **41** (6), 1853–1860.
- 12 Rasia, R.M., Ceolin, M., and Vila, A.J. (2003) Grafting a new metal ligand in the cocatalytic site of *B-cereus* metallo-beta-lactamase: structural flexibility without loss of activity. *Protein. Sci.*, **12** (7), 1538–1546.
- 13 Rasia, R.M. and Vila, A.J. (2004) Structural determinants of substrate binding to *Bacillus cereus* metallo-beta-lactamase. *J. Biol. Chem.*, **279** (25), 26046–26051.
- 14 Yanchak, M.P., Taylor, R.A., and Crowder, M.W. (2000) Mutational analysis of metallo-beta-lactamase CcrA from *Bacteroides fragilis*. *Biochemistry*, **39** (37), 11330–11339.
- 15 Wang, Z.G., Fast, W., and Benkovic, S.J. (1998) Direct observation of an enzyme-bound intermediate in the catalytic cycle of the metallo-beta-lactamase from *Bacteroides fragilis*. *J. Am. Chem. Soc.*, **120** (41), 10788–10789.
- 16 Wang, Z.G., Fast, W., and Benkovic, S.J. (1999) On the mechanism of the

- metallo-beta-lactamase from *Bacteroides fragilis*. *Biochemistry*, **38** (31), 10013–10023.
- 17 Garcia-Saez, I., Hopkins, J., Papamichael, C., Franceschini, N., Amicosante, G., Rossolini, G.M., Galleni, M., Frere, J.M., and Dideberg, O. (2003) The 1.5-angstrom structure of *Chryseobacterium meningosepticum* zinc beta-lactamase in complex with the inhibitor D-captopril. *J. Biol. Chem.*, **278** (26), 23868–23873.
 - 18 Rossolini, G.M., Franceschini, N., Riccio, M.L., Mercuri, P.S., Perilli, M., Galleni, M., Frere, J.M., and Amicosante, G. (1998) Characterization and sequence of the *Chryseobacterium* (Flavobacterium) *meningosepticum* carbapenemase: a new molecular class B beta-lactamase showing a broad substrate profile. *Biochem. J.*, **332**, 145–152.
 - 19 Docquier, J.D., Lamotte-Brasseur, J., Galleni, M., Amicosante, G., Frere, J.M., and Rossolini, G.M. (2003) On functional and structural heterogeneity of VIM-type metallo-beta-lactamases. *J. Antimicrob. Chemother.*, **51** (2), 257–266.
 - 20 Oelschlaeger, P., Mayo, S.L., and Pleiss, J. (2005) Impact of remote mutations on metallo-beta-lactamase substrate specificity: implications for the evolution of antibiotic resistance. *Protein. Sci.*, **14** (3), 765–774.
 - 21 Yamaguchi, Y., Kuroki, T., Yasuzawa, H., Higashi, T., Jin, W.C., Kawanami, A., Yamagata, Y., Arakawa, Y., Goto, M., and Kurosaki, H. (2005) Probing the role of Asp-120(81) of metallo-beta-lactamase (IMP-1) by site-directed mutagenesis, kinetic studies, and X-ray crystallography. *J. Biol. Chem.*, **280** (21), 20824–20832.
 - 22 Materon, I.C., Beharry, Z., Huang, W.Z., Perez, C., and Palzkill, T. (2004) Analysis of the context dependent sequence requirements of active site residues in the metallo-beta-lactamase IMP-1. *J. Mol. Biol.*, **344** (3), 653–663.
 - 23 Oelschlaeger, P., Schmid, R.D., and Pleiss, J. (2003) Insight into the mechanism of the IMP-1 metallo-beta-lactamase by molecular dynamics simulations. *Protein. Eng.*, **16** (5), 341–350.
 - 24 Moali, C., Anne, C., Lamotte-Brasseur, J., Gros Lambert, S., Devreese, B., Van Beeumen, J., Galleni, M., and Frere, J.M. (2003) Analysis of the importance of the metallo-beta-lactamase active site loop in substrate binding and catalysis. *Chem. Biol.*, **10** (4), 319–329.
 - 25 Toleman, M.A., Simm, A.M., Murphy, T.A., Gales, A.C., Biedenbach, D.J., Jones, R.N., and Walsh, T.R. (2002) Molecular characterization of SPM-1, a novel metallo-beta-lactamase isolated in Latin America: report from the SENTRY antimicrobial surveillance programme. *J. Antimicrob. Chemother.*, **50** (5), 673–679.
 - 26 Garau, G., Bebrone, C., Anne, C., Galleni, M., Frere, J.M., and Dideberg, O. (2005) A metallo-beta-lactamase enzyme in action: crystal structures of the monozinc carbapenemase CphA and its complex with biapenem. *J. Mol. Biol.*, **345** (4), 785–795.
 - 27 Valladares, M.H., Felici, A., Weber, G., Adolph, H.W., Zeppezauer, M., Rossolini, G.M., Amicosante, G., Frere, J.M., and Galleni, M. (1997) Zn(II) dependence of the *Aeromonas hydrophila* AE036 metallo-beta-lactamase activity and stability. *Biochemistry*, **36** (38), 11534–11541.
 - 28 Crawford, P.A., Yang, K.W., Sharma, N., Bennett, B., and Crowder, M.W. (2005) Spectroscopic studies on cobalt(II)-substituted metallo-beta-lactamase ImiS from *Aeromonas veronii* bv. *sobria*. *Biochemistry*, **44** (13), 5168–5176.
 - 29 Walsh, T.R., Hall, L., Assinder, S.J., Nichols, W.W., Cartwright, S.J., Macgowan, A.P., and Bennett, P.M. (1994) Sequence-analysis of the L1 metallo-beta-lactamase from *Xanthomonas maltophilia*. *BBA-Gene Struct. Expr.*, **1218** (2), 199–201.
 - 30 Ullah, J.H., Walsh, T.R., Taylor, I.A., Emery, D.C., Verma, C.S., Gamblin, S.J., and Spencer, J. (1998) The crystal structure of the L1 metallo-beta-lactamase from *Stenotrophomonas maltophilia* at 1.7 angstrom resolution. *J. Mol. Biol.*, **284** (1), 125–136.
 - 31 Spencer, J., Clarke, A.R., and Walsh, T.R. (2001) Novel mechanism of hydrolysis of therapeutic beta-lactams by *Stenotrophomonas maltophilia* L1 metallo-beta-lactamase. *J. Biol. Chem.*, **276** (36), 33638–33644.

- 32 Garrity, J.D., Carenbauer, A.L., Herron, L.R., and Crowder, M.W. (2004) Metal binding Asp-120 in metallo-beta-lactamase L1 from *Stenotrophomonas maltophilia* plays a crucial role in catalysis. *J. Biol. Chem.*, **279** (2), 920–927.
- 33 Docquier, J.D., Pantanella, F., Giuliani, F., Thaller, M.C., Amicosante, G., Galleni, M., Frere, J.M., Bush, K., and Rossolini, G.M. (2002) CAU-1, a subclass B3 metallo-beta-lactamase of low substrate affinity encoded by an ortholog present in the *Caulobacter crescentus* chromosome. *Antimicrob. Agents. Ch.*, **46** (6), 1823–1830.
- 34 Docquier, J.D., Lopizzo, T., Liberatori, S., Prenna, M., Thaller, M.C., Frere, J.M., and Rossolini, G.M. (2004) Biochemical characterization of the THIN-B metallo-beta-lactamase of *Janthinobacterium lividum*. *Antimicrob. Agents. Ch.*, **48** (12), 4778–4783.
- 35 Mercuri, P.S., Garcia-Saez, I., De Vriendt, K., Thamm, I., Devreese, B., Van Beeumen, J., Dideberg, O., Rossolini, G.M., Frere, J.M., and Galleni, M. (2004) Probing the specificity of the subclass B3FEZ-1 metallo-beta-lactamase by site-directed mutagenesis. *J. Biol. Chem.*, **279** (32), 33630–33638.
- 36 Garcia-Saez, I., Mercuri, P.S., Papamichael, C., Kahn, R., Frere, J.M., Galleni, M., Rossolini, G.M., and Dideberg, O. (2003) Three-dimensional structure of FEZ-1, a monomeric subclass B3 metallo-beta-lactamase from *Fluoribacter gormanii*, in native form and in complex with D-captopril. *J. Mol. Biol.*, **325** (4), 651–660.
- 37 Mercuri, P.S., Bouillenne, F., Boschi, L., Lamote-Brasseur, J., Amicosante, G., Devreese, B., Van Beeumen, J., Frere, J.M., Rossolini, G.M., and Galleni, M. (2001) Biochemical characterization of the FEZ-1 metallo-alpha-lactamase of *Legionella gormanii* ATCC 33297(T) produced in *Escherichia coli*. *Antimicrob. Agents Chemother.*, **45** (4), 1254–1262.
- 38 Bellais, S., Aubert, D., Naas, T., and Nordmann, P. (2000) Molecular and biochemical heterogeneity of class B carbapenem-hydrolyzing beta-lactamases in *Chryseobacterium meningosepticum*. *Antimicrob. Agents Chemother.*, **44** (7), 1878–1886.
- 39 Wang, Z.G., Fast, W., and Benkovic, S.J. (1999) Mechanistic studies of the metallo-beta-lactamase from *Bacteroides fragilis*. *J. Inorg. Biochem.*, **74** (1–4), 333.
- 40 Wang, Z.G., Fast, W., Valentine, A.M., and Benkovic, S.J. (1999) Metallo-beta-lactamase: structure and mechanism. *Curr. Opin. Chem. Biol.*, **3** (5), 614–622.
- 41 Fast, W., Wang, Z.G., and Benkovic, S.J. (2001) Familial mutations and zinc stoichiometry determine the rate-limiting step of nitrocefin hydrolysis by metallo-beta-lactamase from *Bacteroides fragilis*. *Biochemistry*, **40** (6), 1640–1650.
- 42 Rasia, R.M. and Vila, A.J. (2004) Structural determinants of substrate binding to *Bacillus cereus* metallo-beta-lactamase. *J. Biol. Chem.*, **279** (25), 26046–26051.
- 43 Davies, A.M., Rasia, R.M., Vila, A.J., Sutton, B.J., and Fabiane, S.M. (2005) Effect of pH on the active site of an Arg121Cys mutant of the metallo-beta-lactamase from *Bacillus cereus*: Implications for the enzyme mechanism. *Biochemistry*, **44** (12), 4841–4849.
- 44 Fabiane, S.M., Sohi, M.K., Wan, T., Payne, D.J., Bateson, J.H., Mitchell, T., and Sutton, B.J. (1998) Crystal structure of the zinc-dependent beta-lactamase from *Bacillus cereus* at 1.9 angstrom resolution: binuclear active site with features of a mononuclear enzyme. *Biochemistry*, **37** (36), 12404–12411.
- 45 Carfi, A., Pares, S., Duee, E., Galleni, M., Duez, C., Frere, J.M., and Dideberg, O. (1995) The 3-D structure of a zinc metallo-beta-lactamase from *Bacillus-cereus* reveals a new-type of protein fold. *EMBO J.*, **14** (20), 4914–4921.
- 46 Concha, N.O., Rasmussen, B.A., Bush, K., and Herzberg, O. (1996) Crystal structure of the wide-spectrum binuclear zinc beta-lactamase from *Bacteroides fragilis*. *Structure*, **4** (7), 823–836.
- 47 Concha, N.O., Janson, C.A., Rowling, P., Pearson, S., Cheever, C.A., Clarke, B.P., Lewis, C., Galleni, M., Frere, J.M., Payne, D.J., Bateson, J.H., and Abdel-Meguid, S.S. (2000) Crystal structure of the IMP-1 metallo beta-lactamase from *Pseudomonas aeruginosa* and its complex with a mercaptocarboxylate inhibitor: binding determinants of a potent, broad-

- spectrum inhibitor. *Biochemistry*, **39** (15), 4288–4298.
- 48 Murphy, T.A., Catto, L.E., Halford, S.E., Hadfield, A.T., Minor, W., Walsh, T.R., and Spencer, J. (2006) Crystal structure of *Pseudomonas aeruginosa* SPM-1 provides insights into variable zinc affinity of metallo-beta-lactamases. *J. Mol. Biol.*, **357** (3), 890–903.
 - 49 Costello, A.L., Sharma, N.P., Yang, K.W., Crowder, M.W., and Tierney, D.L. (2006) X-ray absorption spectroscopy of the zinc-binding sites in the class B2 metallo-beta-lactamase ImiS from *Aeromonas veronii* bv. *sobria*. *Biochemistry*, **45** (45), 13650–13658.
 - 50 Bebrone, C., Anne, C., Kerff, F., Garau, G., De Vriendt, K., Lantin, R., Devreese, B., Van Beeumen, J., Dideberg, O., Frere, J.M., and Galleni, M. (2008) Mutational analysis of the zinc- and substrate-binding sites in the CphA metallo-beta-lactamase from *Aeromonas hydrophila*. *Biochem. J.*, **414**, 151–159.
 - 51 Dal Peraro, M., Carloni, P., and Vila, A.J. (2002) Structural determinants and H-bond network of mononuclear zinc-beta-lactamase active site. *Biophys. J.*, **82** (1), 439A.
 - 52 Dal Peraro, M., Vila, A.J., and Carloni, P. (2003) Protonation state of Asp120 in the binuclear active site of the metallo-beta-lactamase from *Bacteroides fragilis*. *Inorg. Chem.*, **42** (14), 4245–4247.
 - 53 Dal Peraro, M., Llarrull, L.I., Rothlisberger, U., Vila, A.J., and Carloni, P. (2004) Water-assisted reaction mechanism of monozinc beta-lactamases. *J. Am. Chem. Soc.*, **126** (39), 12661–12668.
 - 54 Dal Peraro, M., Vila, A.J., and Carloni, P. (2004) Substrate binding to mononuclear metallo-beta-lactamase from *Bacillus cereus*. *Proteins*, **54** (3), 412–423.
 - 55 Simona, F., Magistrato, A., Vera, D.M., Garau, G., Vila, A.J., and Carloni, P. (2007) Protonation state and substrate binding to B2 metallo-beta-lactamase CphA from *Aeromonas hydrophila*. *Proteins*, **69** (3), 595–605.
 - 56 Dal Peraro, M., Vila, A.J., Carloni, P., and Klein, M.L. (2007) Role of zinc content on the catalytic efficiency of B1 metallo beta-lactamases. *J. Am. Chem. Soc.*, **129** (10), 2808–2816.
 - 57 Dal Peraro, M., Ruggerone, P., Raugei, S., Gervasi, F.L., and Carloni, P. (2007) Investigating biological systems using first principles Car–Parrinello molecular dynamics simulations. *Curr. Opin. Struct. Biol.*, **17** (2), 149–156.
 - 58 Dal Peraro, M., Vila, A.J., and Carloni, P. (2002) Structural determinants and hydrogen-bond network of the mononuclear zinc(II)-beta-lactamase active site. *J. Biol. Inorg. Chem.*, **7** (7–8), 704–712.
 - 59 Laio, A., Van de Vondel, J., and Rothlisberger, U. (2002) A Hamiltonian electrostatic coupling scheme for hybrid Car–Parrinello molecular dynamics simulations. *J. Chem. Phys.*, **116** (16), 6941–6947.
 - 60 Car, R. and Parrinello, M. (1985) Unified approach for molecular-dynamics and density-functional theory. *Phys. Rev. Lett.*, **55** (22), 2471–2474.
 - 61 Sprik, M. and Ciccotti, G. (1998) Free energy from constrained molecular dynamics. *J. Chem. Phys.*, **109** (18), 7737–7744.
 - 62 Spiegel, K., Rothlisberger, U., and Carloni, P. (2006) Duocarmycins binding to DNA investigated by molecular simulation. *J. Phys. Chem. B*, **110** (8), 3647–3660.
 - 63 Carfi, A., Duee, E., Galleni, M., Frere, J.M., and Dideberg, O. (1998) 1.85 Å resolution structure of the zinc(II) beta-lactamase from *Bacillus cereus*. *Acta Crystallogr. D*, **54**, 313–323.
 - 64 Llarrull, L.I., Tioni, M.F., Kowalski, J., Bennett, B., and Vila, A.J. (2007) Evidence for a dinuclear active site in the metallo-ss-lactamase BcII with substoichiometric Co(II) – A new model for metal uptake. *J. Biol. Chem.*, **282** (42), 30586–30595.
 - 65 de Seny, D., Heinz, U., Wommer, S., Kiefer, M., Meyer-Klaucke, W., Galleni, M., Frere, J.M., Bauer, R., and Adolph, H.W. (2001) Metal ion binding and coordination geometry for wild type and mutants of metallo-beta-lactamase from *Bacillus cereus* 569/H/9 (BcII) – A combined thermodynamic, kinetic, and spectroscopic approach. *J. Biol. Chem.*, **276** (48), 45065–45078.

- 66 McManus-Munoz, S. and Crowder, M.W. (1999) Kinetic mechanism of metallo-beta-lactamase L1 from *Stenotrophomonas maltophilia*. *Biochemistry*, **38** (5), 1547–1553.
- 67 Llarrull, L.I., Fabiane, S.M., Kowalski, J.M., Bennett, B., Sutton, B.J., and Vila, A.J. (2007) Asp-120 locates Zn²⁺ for optimal metallo-ss-lactamase activity. *J. Biol. Chem.*, **282** (25), 18276–18285.
- 68 Tioni, M.F., Llarrull, L.I., Poeylout-Palena, A.A., Marti, M.A., Saggu, M., Periyannan, G.R., Mata, E.G., Bennett, B., Murgida, D.H., and Vila, A.J. (2008) Trapping and characterization of a reaction intermediate in carbapenem hydrolysis by *B-cereus* metallo-beta-lactamase. *J. Am. Chem. Soc.*, **130** (47), 15852–15863.
- 69 Poeylout-Palena, A.A., Tomatis, P.E., Karsisiotis, A.I., Damblon, C., Mata, E.G., and Vila, A.J. (2007) A minimalistic approach to identify substrate binding features in B1 Metallo-beta-lactamases. *Bioorg. Med. Chem. Lett.*, **17** (18), 5171–5174.
- 70 Llarrull, L.I., Tioni, M.F., and Vila, A.J. (2008) Metal content and localization during turnover in *B-cereus* metallo-beta-lactamase. *J. Am. Chem. Soc.*, **130** (47), 15842–15851.
- 71 Gonzalez, J.M., Martin, F.J.M., Costello, A.L., Tierney, D.L., and Vila, A.J. (2007) The Zn²⁺ position in metallo-beta-lactamases is critical for activity: a study on chimeric metal sites on a conserved protein scaffold. *J. Mol. Biol.*, **373** (5), 1141–1156.
- 72 Tomatis, P.E., Fabiane, S.M., Simona, F., Carloni, P., Sutton, B.J., and Vila, A.J. (2008) Adaptive protein evolution grants organismal fitness by improving catalysis and flexibility. *Proc. Natl. Acad. Sci. U.S.A.*, **105** (52), 20605–20610.
- 73 Spencer, J., Read, J., Sessions, R.B., Howell, S., Blackburn, G.M., and Gamblin, S.J. (2005) Antibiotic recognition by binuclear metallo-beta-lactamases revealed by X-ray crystallography. *J. Am. Chem. Soc.*, **127** (41), 14439–14444.

22

Computational Simulation of the Terminal Biogenesis of Sesquiterpenes: The Case of 8-Epiconfertifin

José Enrique Barquera-Lozada and Gabriel Cuevas

22.1

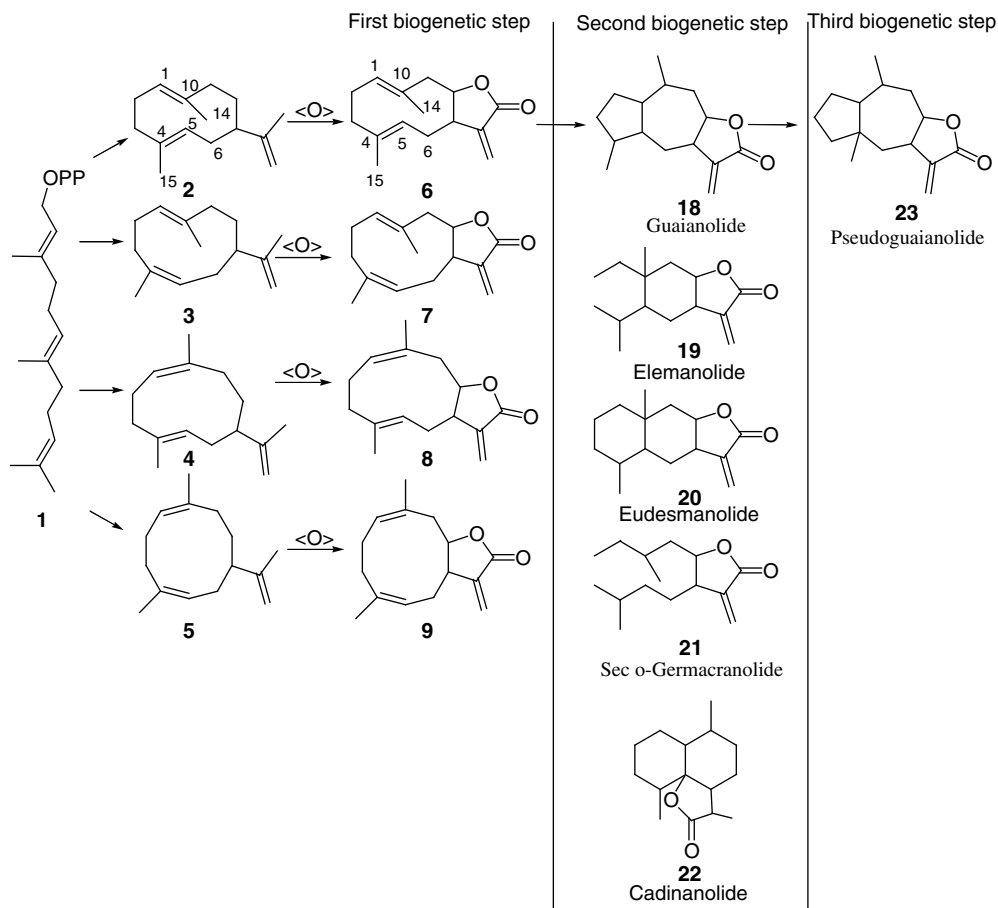
Introduction

Amongst chemists there has always been a fascination with the origin of compounds associated with life. Ever since Alfonso Herrera performed the first experiments designed to establish the origin of life on Earth it was noted that physicochemical principles are followed in both the reactions inside as flask and in cellular processes [1, 2]. The origin of several primary metabolites, including proteins, carbohydrates, lipids and others, has been studied extensively mainly because the processes obey physicochemical principles and, thus, it has been possible to rigorously establish the metabolic routes that form them [3]. This has been possible because many of the enzymes that catalyze these processes have been isolated and characterized. In addition, the genes that codify for their biosynthesis are well known and it has been possible to determine the mechanisms that regulate their activity.

Secondary metabolites are a group of chemical compounds that do not participate in the basic functions of growth and reproduction but participate in functions that facilitate the organism's adaptation to the environment [4]. Some compounds perform defense functions against pathogens and predators. Others attract insects that contribute to pollenization, disperse seeds and have allelopathic functions. Since secondary metabolites have diverse structures, they have been used as taxonomic markers and have been useful to determine the degree of evolution of different species [5]. Based on the biogenetic origin of secondary metabolites they can be classified in three groups, terpenoids, alkaloids and phenylpropanoids [6, 7], among other possible classifications [8].

Curiously, such a diverse variety of chemical compounds is produced using only a very limited number of biosynthetic routes [9]. This may be yet another example of the efficiency that commonly prevails in nature [3]. For example, mevalonic acid is the origin of all terpenes and terpenoids in superior organisms while 1-deoxy-D-xylulose-5-phosphate is used by lesser organisms [10, 11]. On the other hand, all the sesquiterpenes used by superior organisms have their origin in either farnesol

(specifically the farnesyl pyrophosphate, FPP, **1**, Scheme 22.1) or nerolidyl pyrophosphate [12–14].



Scheme 22.1 Biogenetic origin of pseudoguaianolides.

The first biogenetic hypothesis to explain the origin of terpenes is Ruzicka's rule of isoprene [15, 16]. According to this rule, three molecules of acetyl coenzyme A are successively condensed to produce 3-hydroxy-3-methylglutaryl coenzyme A through the action of HMG-CoA synthetase that through reduction generates mevalonic acid. A decarboxylation-elimination reaction generates isopentenyl pyrophosphate that can be isomerized to dimethylallyl pyrophosphate through the participation of isopentenyl-pyrophosphate isomerase [3, 4]. The latter two compounds allow the construction of all terpenoids in superior organisms. The so-called hemiterpenes derive directly from any of the two compounds while sesquiterpenes have their origin in the condensation of three of these molecules. This condensation also allows the

biosynthesis of farnesyl pyrophosphate (**1**, Scheme 22.1) or of its isomer nerolidyl pyrophosphate [17].

Nowadays, several fundamental aspects of the mevalonate and non-mevalonate pathways are well known, such as the enzymes that catalyze their transformation and the genes that codify for them [18]. The terminal biogenesis of the different natural products responsible for the great structural diversity of compounds has been rarely studied and there are still many unanswered questions on the mechanisms responsible for generating such a variety of molecules [13, 15, 16, 19–32].

Chemical simulation of biogenetic processes is based on the fact that the reactions that take place *in vivo* follow the same general principles of *in vitro* transformations [1]. Using this concept, one can demonstrate that, in many cases, a complex enzymatic system is not really necessary to maintain control of several aspects such as *in situ*, stereo- and regio-selectivity. Enzymes seem fundamental for the selection of the starting conformer and explain the high chemical yields of these transformations; however, one or two steps of the biogenetic process can be simulated *in vitro* [33–35].

Sesquiterpene synthetases generate cyclic compounds from FPP (**1**, Scheme 22.1) and control their subsequent transformation into more than 300 different sesquiterpenes [36]. Thus, it has been proposed that an enzyme of this type transforms FPP into a derivative of cyclodecadiene. This hypothesis has been proven through the isolation of (+)-germacrene A synthetase isolated from chicory roots (*Cichorium intybus* L.) [37]. These kinds of enzymes and their action have been widely studied.

trans,trans-Germacradiene (**2**, Scheme 22.1) is obtained from the direct cyclization process of *trans,trans*-FPP [or (*E,E*)-FPP], an intermediate that is transformed by enzymatic oxidative modification to yield the corresponding lactones. Two possible biogenetic routes have been suggested for the formation of the lactone ring of these sesquiterpenoids [13, 23, 42]. In the first biogenetic stage, there are four possible configurational isomers: the germacrolides of configuration C1-*E*, C4-*E* (**6**, Scheme 22.1), the heliangolides (**7**) (C1-*E*, C4-*Z*), the melampolides (**8**) (C1-*Z*, C4-*E*) and the *Z,Z* germacranolides (**9**). This was demonstrated by the isolation of compounds with four possible configurations [38]. Costunolide [39] (**10**) and tamaulipin [40] (**11**) are examples of (*E,E*)-germacradiene, frutescin [41] (**12**) and schkurliolide [41] (**13**) are examples of C1-*Z*, C4-*E* lactones, nobilin [42] (**14**) and eupaformonin [43] (**15**) are examples of C1-*E*, C4-*Z* lactones and melcanthin C(**16**) [44] and artemisiifolin [45] (**17**) are examples of (*Z,Z*)-germacranolides (Figure 22.1).

According to Fischer [13, 38], in the second biogenetic stage, five different types of skeletons are produced from the germacranolide. These are the guaianolides (**18**), elemanolides (**19**), eudesmanolides (**20**), seco-germacranolides (**21**), and cadinanolides (**22**) (Scheme 22.1). It has been proposed that, for example, guaianolides and eudesmanolides are produced through electrophilic attacks of the double bonds of cyclodecadiene or any of its epoxides while elemanolides require a Cope-type reaction.

Pseudoguaianolides (**23**, Scheme 22.1) are formed from the third biogenetic stage and typically contain a non-regular isoprene skeleton with a methyl group at C5. A fundamental fact that arises is that it is possible to isolate from nature both guaianolides with *cis* and *trans* fusion and epimeric pseudoguaianolides at position

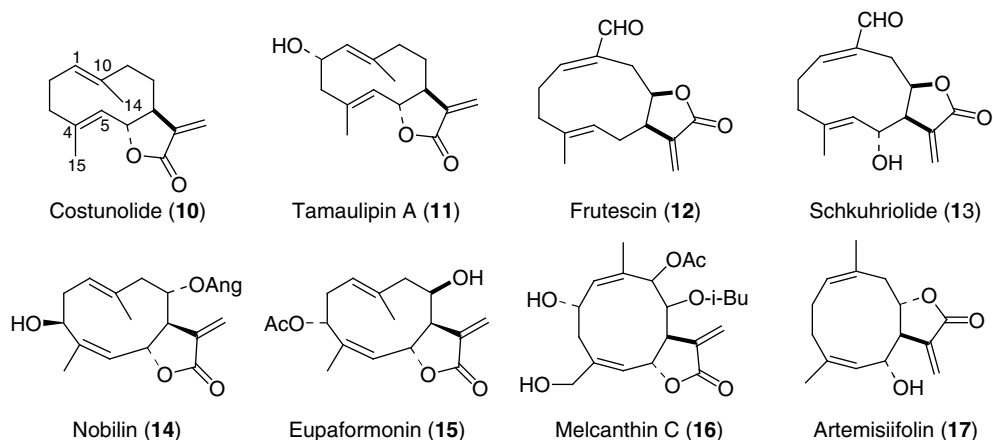


Figure 22.1 Natural sesquiterpenic lactones with the four possible configurations of the double bonds.

10. For example, when the methyl group at C10 adopts the β -position, a lactone of the ambrosianolide type is obtained (**24**, Figure 22.2). This compound is abundant in the Ambrosiinae sub-tribe and the genus *Parthenium*. However, if the methyl adopts an α -position, it is possible to obtain a skeleton typical of the helenanolides (**25**), which is related to some members of the Heleniae tribe.

Biogenetic hypotheses of natural products are based on the proposition of reasonable reaction mechanisms that allow the product under study to be obtained from a precursor that is usually found in the same natural source. Support for a biogenetic hypothesis usually requires the structure of the precursor and the product; however, available experimental information is very limited. Consequently, the use of computational models makes sense since it is possible to simulate the reaction mechanism, supported always by experimental results. Without the experimental data, the results of calculations are mere speculation.

In the laboratory, it is relatively easy to transform 4,5-epoxygermacranolides into guaianolides using cyclizations catalyzed by Lewis acids [46]. The same is true for the transformation of 4,5-epoxyguaianolide into pseudoguaianolides [47]. However, biomimetic transformations of a germacranolide into the third biogenetic stage are rare. The first successful experimental transformation of a germacradiene derivative

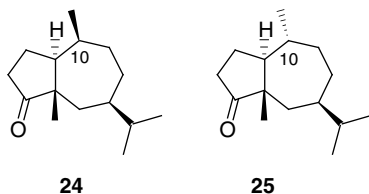
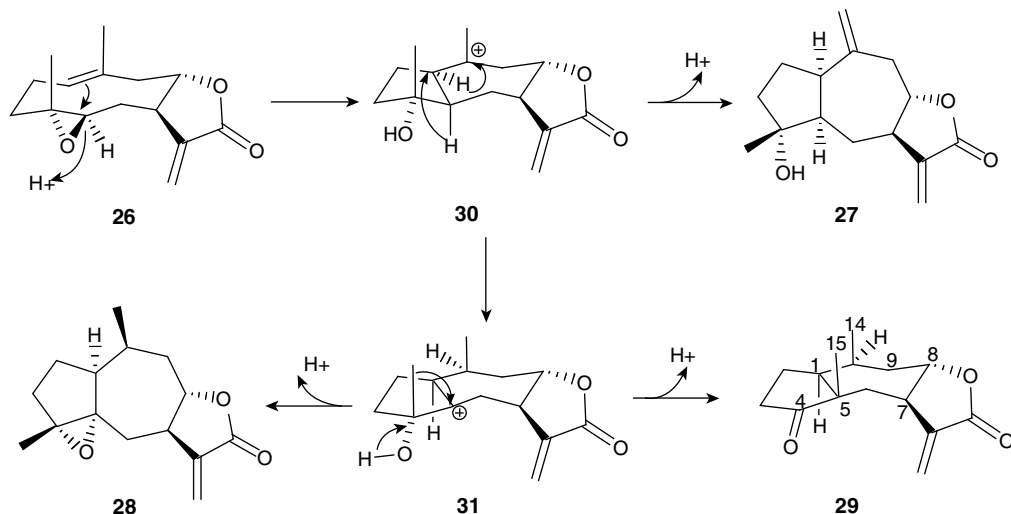


Figure 22.2 Absolute stereochemistry of the stereogenic center supporting the methyl group in the ambrosianolides is *S* (**24**) while in helenolides is *R* (**25**).

into a pseudoguaianolide was the transformation performed by Ortega and Maldonado [48] of $4\alpha,5\beta$ -epoxyinunolide (**26**, Scheme 22.2), isolated from *Stevia tephrophylla* Blake, into inuviscolide (**27**, yield 1.8%), $4\alpha,5\alpha$ -epoxy- $10\alpha,14H$ -inuviscolide (**28**, yield 1%) and 8-epiconfertine (**29**), a pseudoguaianolide (yield 4.1%), using a bentonitic earth.



Scheme 22.2 Proposed mechanism for the transformation of $4\alpha,5\beta$ -epoxyinunolide (**26**) into inuviscolide (**27**), $4\alpha,5\alpha$ -epoxy- $10\alpha,14H$ -inuviscolide (**28**) and 8-epiconfertine (**29**) through cations **30** and **31** [48].

The stereochemistry of the cyclization products is expected if it is assumed that the most stable conformation of the precursor **26** is similar to the conformation of laurenobiolide [49], in which both methyl groups are above the plane of the cyclodecadiene and the double bond and the epoxide have a crossed orientation, which in Samek's nomenclatures is described as $^{15}D_{5,1}D^{14}$ (**26**) [50].

22.2

Reaction Mechanism

The proposed mechanism is presented in Scheme 22.2 and is supported in the structure of the isolated byproducts [48]. Within this mechanism, it is proposed that in the first step is a cyclization of the $^{15}D_{5,1}D^{14}$ conformer, which produces a *cis*-fused carbocation **30** that yields compound **27** if the methyl group takes part in the elimination reaction. This is the main product when Lewis or Brønsted acids are used.

Carbocation **30**, after two consecutive hydrogen shifts (Scheme 22.2), produces carbocation **31**, which through an intramolecular nucleophilic attack generates epoxide **28**. Finally, carbocation **31** suffers a transposition driven by the formation

of the carbonyl group that produces the methyl group migration. Because Ortega and Maldonado [48] had no further proof of the existence of any other carbocation that takes part in the process, they did not mention other cationic intermediates that would be likely to take part. Finally, Herz has proposed that the transformation of pseudoguaianolides into guaianolides occurs in a concerted fashion [22], in other words, the transformation of alcohol into ketone and the migration of the methyl and the two hydrogen atoms occurs in a single step without the participation of intermediates.

Then, the issue is which carbocations really exist as intermediates and what is the nature of the transition states (TSs) involved in the process in terms of structure and energy. The other issue to be answered is if this reaction needs enzymatic catalysis or if it is only a “cascade” pathway in which the main role of the enzyme is to start the process by activating the electrophilic reaction and stabilizing cationic intermediates to enable the required transpositions. In this case, it is interesting to analyze the role of bentonitic earth because it has two major effects. The first effect is selectivity. Although it lacks the complementary structure, similar to the pattern characteristic of proteins, the bentonitic earth can select the starting conformer of the germacradiene. The second effect is the capability to stabilize the intermediate cations that might increase their average lifetime. This is why the system may undergo transposition reactions before undergoing eliminations and, as a consequence, the formation of more advanced biogenetic derivatives. It is important to establish the conformational properties of compound **26** since, according to Fischer [38], the formation of the epimers at C10 depend on conformer $^{15}\text{D}_5, ^1\text{D}_{14}$ (**26**) transforming into conformer $^{15}\text{D}_5, ^1\text{D}_{14}$ (**32**, Figure 22.3).

Regarding the establishment of the terminal biogenesis of pseudoguaianolides, it is very important to study the previous biogenetic proposals. First, it has been proposed that the electrophilic closure that generates the fused rings of five and seven members occurs in a concerted fashion. In other words, without the existence of intermediates [3]. Most interestingly, Hendrickson’s scheme does not explain the stereochemistry of the group of lactones known as heleanolide [19]. Parker *et al.* [51] proposed a first hypothesis that stated that these molecules should come from a germacradiene with a *Z*-configuration in the double bond C1–C10 (melampolide, **8**, Scheme 22.1). This is the origin of the relative stereochemistry associated with the *cis* and *trans* fusion of the skeleton of guaianolides associated with the geometry of double bonds in the starting compounds. Thus, the *cis* guaianolides originate in germacradienes with olefins of 1-*E*,4-*E* configuration (**6**, Scheme 22.1) and the *trans* guaianolides have their origin in germacradienes with geometry 1-*Z*,4-*E* (**8**, Scheme 22.1). In contrast, a third hypothesis put forth by Fischer [38, 52] gives another alternative, suggesting that a different conformer of the same germacradiene **6** is responsible for the second series of compounds.

A recently published [53] study has demonstrated that this type of problem can be approached using computational methods and that the mPW1B95 functional developed by Truhlar *et al.* [54] and the double split valence polarized and diffuse base 6-31+G(d,p) constitute a reasonable level of theory with better performance than the MP2 method. All energies were reported with zero-point energy corrections and

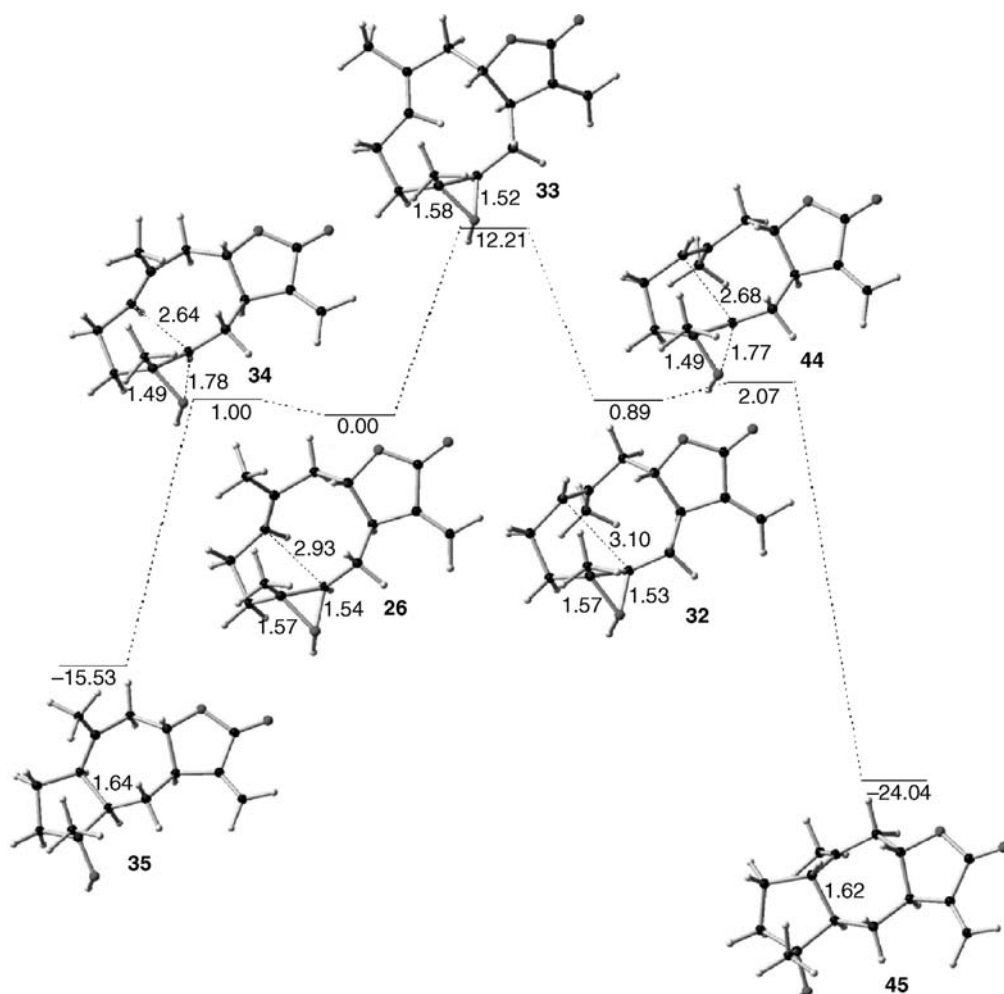


Figure 22.3 Conformational barrier for the interchange between $^{15}\text{D}_{5,1}\text{D}^{14}$ (**26**) and $^{15}\text{D}_5, ^1\text{D}^{14}$ (**32**) conformers and the first electrophilic ring closure; relative energy in kcal mol^{-1} . Carbon in black, oxygen in gray and hydrogen in white.

are not scaled for comparative purpose. The 6-31+G(d,p) basis functions were used because addition of diffuse functions to a double split valence basis has been shown to be more important than increasing to a triplet split valence basis when calculating reaction energies and activation energies with DFT [55]. All the calculations included in this chapter were performed using the Gaussian 03 program [56].

A recent study has examined the conformational process that converts the two conformers of the protonated $4\alpha,5\beta$ -epoxyinunolide at the epoxide's oxygen through the rotation of the bonds neighboring the C1–C10 double bonds; in other words, the process that interconverts the $^{15}\text{D}_{5,1}\text{D}^{14}$ (**26**) conformer from a chair–chair

conformation, which is, in fact, the most stable, into the $^{15}\text{D}_5, ^1\text{D}_{14}$ (**32**) conformer that is less stable by $0.89 \text{ kcal mol}^{-1}$ ($1 \text{ kcal} = 4.184 \text{ kJ}$) and maintains its boat–chair conformation [53].

The corresponding conformational TS (**33**, Figure 22.3) maintains the trisubstituted ethylene on the molecule's plane and is located $12.21 \text{ kcal mol}^{-1}$ from the most stable conformer. This barrier is reasonable in regard to the inter-conversion energy of cyclodecane [57] if it is considered that both the double bond as well as the lactone of *trans* fusion and the epoxide confer rigidity to the ring.

The transformation mechanism of conformer $^{15}\text{D}_5, ^1\text{D}_{14}$ into the corresponding protonated pseudoguaianolide implies a group of stationary states that are presented on Figure 22.4 and have been fully described [53]. Figure 22.5 presents the intermediates that lead to the formation of an epimer at C10 of the 8-epiconferfin. Figure 22.6 presents the relative energy of each stationary state in relation to compound **26**, which is used here as a reference.

One of the consequences derived from the conformation adopted by **26** and **32** is the different trans-annular distance ($2.93 \text{ \AA}$ in **26** and $3.10 \text{ \AA}$ in **32**), even when the O–C5 distance is the same. The transition state for the electrophilic closure associated with each conformer (**34** and **44**, respectively) shows a decrease of the trans-annular distance and the increase of the C5–OH distance that ends in the *cis* fusion product (**35**) with a C1–C5 distance of $1.64 \text{ \AA}$ and the *trans* fusion product (**45**) with a C1–C5 distance of $1.62 \text{ \AA}$. This is the first elemental step in the mechanism, with product **45** being more stable than **35** by $8.41 \text{ kcal mol}^{-1}$. The transition states for the trans-annular closure barely differ, by $1.07 \text{ kcal mol}^{-1}$. The lower stability of **35** in relation to **45** can be associated to the fact that **35** presents a *cis* fusion that implies a partial eclipsing of the C10–C1–C5–C6 ($\tau = 5.37^\circ$), H–C1–C5–H ($\tau = 9.18^\circ$) and C2–C1–C5–C4 ($\tau = 2.92^\circ$) segments. In addition, the methyl group joined to C10, an atom that has lost its charge, adopts an arrangement close to methylene C6. This arrangement would be expected to be more energetic than that of **45** where the methyl is oriented in an opposing direction with respect to the C6 methylene and the segments associated to the fusion maintain an anti disposition. In the next stage, the reaction mechanisms are different and will not be similar again until the hydrogen atom at position 1 is located at position 10, where it is the base for the methyl group at C14. For the system originated in the β -epimer, this transformation requires two elemental steps [53] while only one elemental step is required for the α -epimer. Intermediates **35** and **45** lack the proper conformational arrangement to allow the required {1,2}-hydrogen shift, the next step in the transformation. The plane that forms C1–C10–C14 is parallel to the C–H that must be transferred; this chemical process requires that the transferring bond be perpendicular to the plane of the referred atoms. Thus, for **35** and **45** the H–C1–C10–C9 angle is 31.2° and 23.6° , respectively. It has been stated that conformational analysis requires considering the conformational process as a sequence of elementary conformational steps [58, 59]. Hence, it is necessary to establish the way in which the stationary states of minimum energy are interconnected through the TS. This is feasible for cation **35**, which is transformed into conformer **37** through TS **36** (Figure 22.4). In this TS, the angle C14–C10–C1–C5 reaches 145° . The stabilization of cation **36** is due to

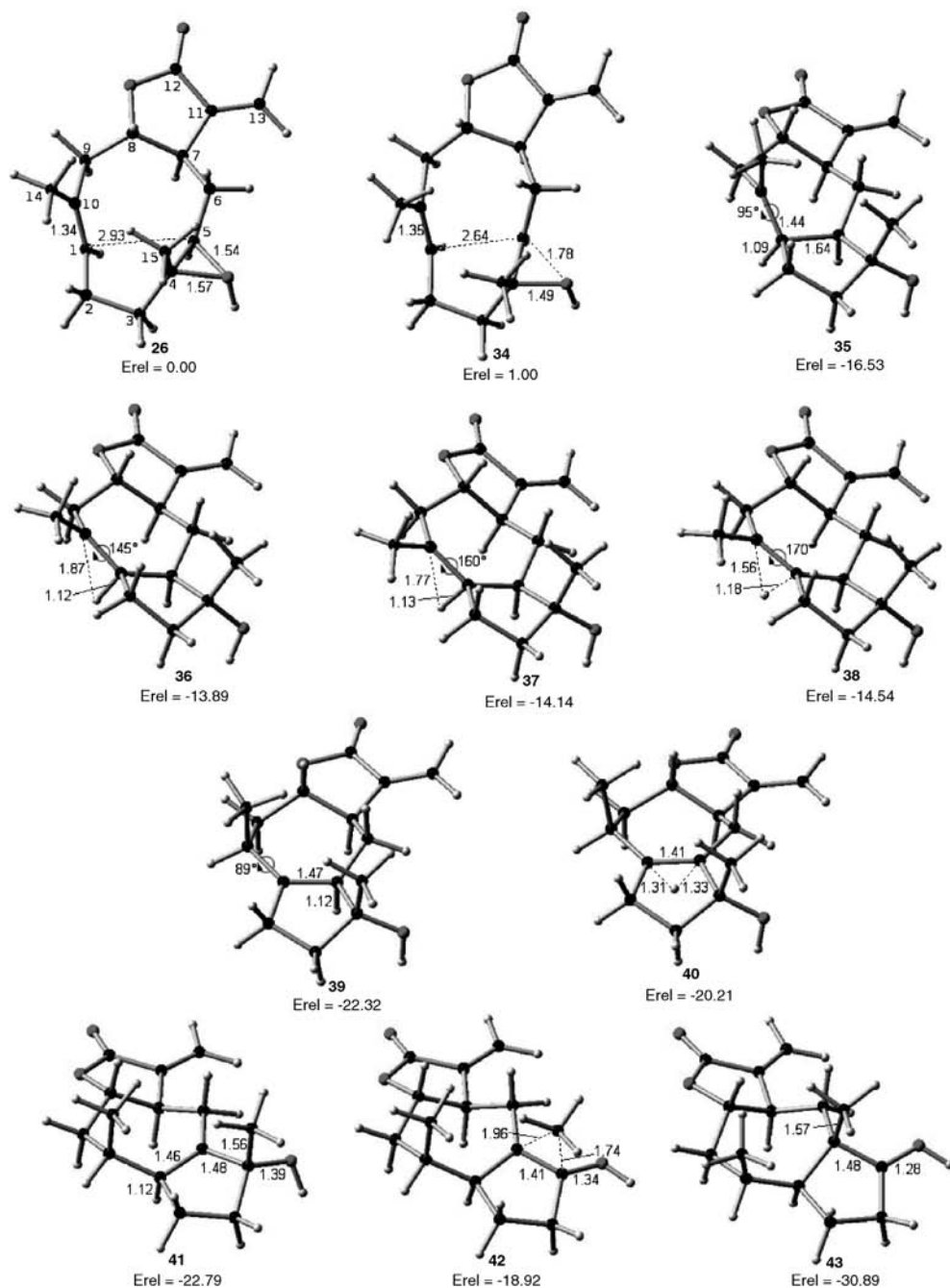


Figure 22.4 Reaction intermediates and TS structures for the conversion of the conformer $^{15}\text{D}_{5,1}\text{D}^{14}$ of epoxinunolide cation (**26**) into the 8-epiconfertin cation (**43**) at the mPW1B95/6-31+G(d,p) level. Selected distances and

C14–C10–C1–C5 dihedral angles are shown in Å and in degrees, respectively; relative energy in kcal mol⁻¹. Carbon in black, oxygen in gray and hydrogen in white.

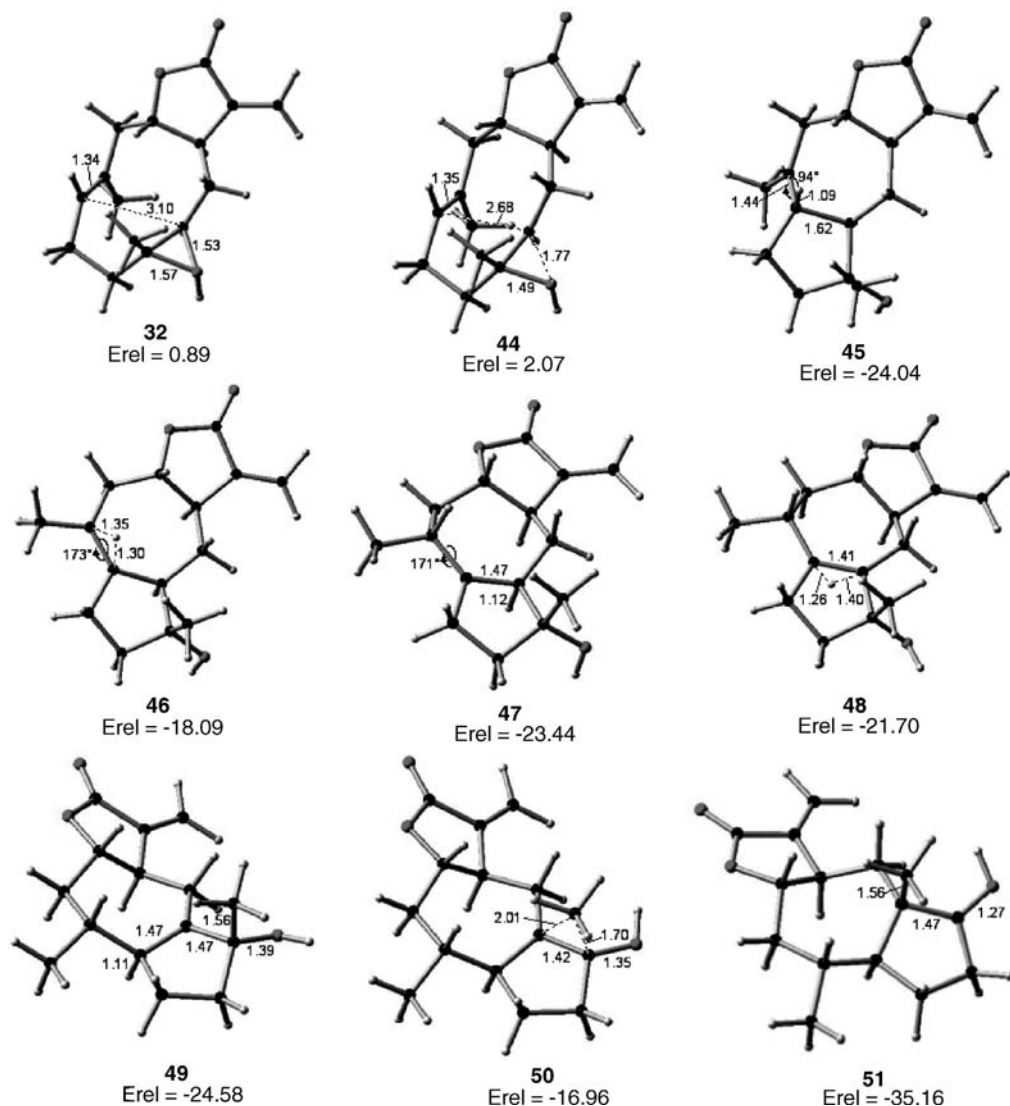


Figure 22.5 Reaction intermediates and TS structures for the conversion of the conformer $^{15}\text{D}_5$, $^{15}\text{D}_{14}$ of the epoxynunolide cation (32) into the C10-epimer of 8-epiconfertifin cation (51) at the mPW1B95/6-31+G(d,p) level. Selected

distances and C14–C10–C1–C5 dihedral are shown in Å and in degrees, respectively; relative energy in kcal mol $^{-1}$. Carbon in black, oxygen in gray and hydrogen in white.

hyperconjugation [60] by the participation of two $\sigma_{\text{C}-\text{C}}$ bonds (C1–C5 and C8–C9) and a $\sigma_{\text{C}-\text{H}}$ bond of the methyl group [61]. At the mPW1B95/6-31+G(d,p) level the distance C1–C5 is 1.64 Å, C8–C9 is 1.55 Å and C15–H is 1.10 Å, which is larger than C15–H distance of the other hydrogen atoms of the methyl group. NBO analysis [62]

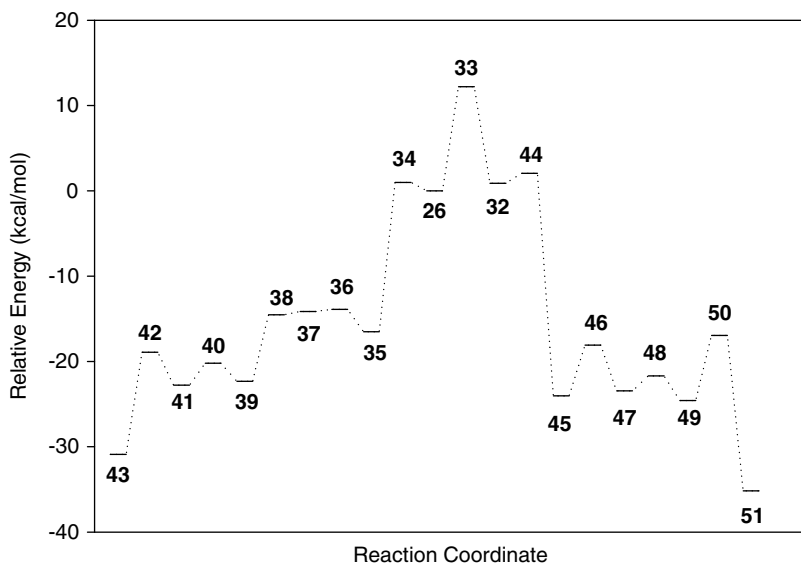


Figure 22.6 Energetic pathway for the formation of 8-epiconfertifin cation and its C10 epimer from $^{15}\text{D}_5, ^1\text{D}_{14}$ (**26**) and $^{15}\text{D}_5, ^1\text{D}_{14}$ (**32**) conformers of $\alpha,5\beta$ -epoxyinunolide; relative energy in kcal mol^{-1} .

of these $\sigma_{\text{C-H}} \rightarrow \pi$ interactions establishes values of 25.4, 13.4 and $15.8 \text{ kcal mol}^{-1}$ respectively for each interaction.

In cation **37** (Figure 22.4), the product of the conformational change, there are three hydrogen atoms that hyperconjugate to produce stabilization. The C1–H distance is $1.13 \text{ \AA}$, C9–H is $1.10 \text{ \AA}$ and C15–H is $1.10 \text{ \AA}$ and coincides with the energy stabilization values determined by NBO analysis at the mPW1B95/6-31+G(d,p) level of 54.8, 14.8 and $18.0 \text{ kcal mol}^{-1}$, respectively.

With the adequate conformation, an {1,2}-hydrogen shift is experienced through TS **38** (Figure 22.4), where the hydrogen atom is found at a distance of $1.18 \text{ \AA}$ [at the mPW1B95/6-31+G(d,p) level], barely $1.56 \text{ \AA}$ from the destination carbon. Concurrently with migration, the C14–C10–C1–C5 angle is adjusted, going to 170° in TS **38** and finishing at 89° in product **39**. The asynchrony of TS **38** is noteworthy. The natural charge of the hydrogen atom, which migrates, is 0.42 in TS **38**, while in **39** it is 0.31 and in **37** it is 0.32. At C1 the charge is -0.38 , -0.25 and 0.51 and at C10 it is 0.48 , 0.32 and -0.38 for **37**, **38** and **39**, respectively, at the mPW1B95/6-31+G(d,p) level of theory. From the above, it can be concluded that in the TS the charge is delocalized mainly between the two carbons; even if it is also delocalized towards the hydrogen atom. Apparently, this intermediate is stabilized through a process of three-center, two-electrons bond (3c–2e); the hydrogen atom, instead of migrating as a hydride as is usually accepted [63, 64], migrates as a proton. In chemical language, this is denominated a {1,2}-hydride shift.

To confirm this, the charges adjusted to the electrostatic potential were determined [65]. For **38** the hydrogen atom under study has a charge of 0.17, while C1 has a charge of -0.08 and C10 of 0.32.

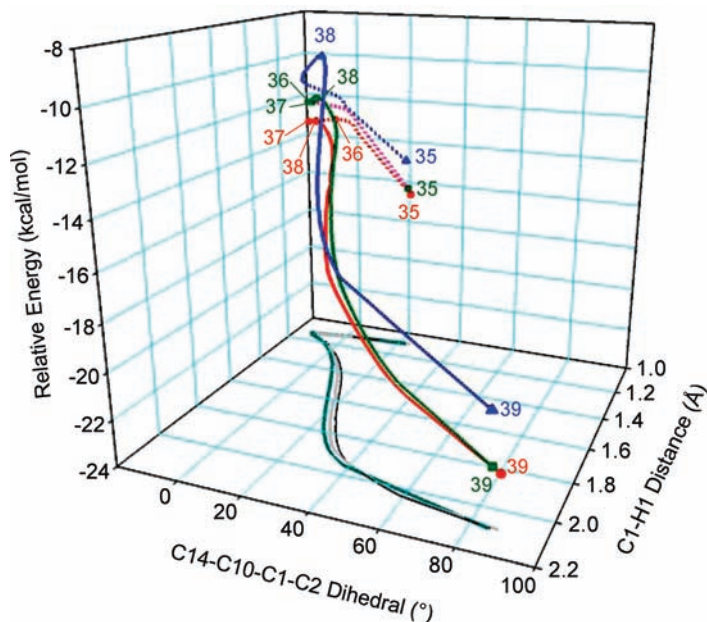


Figure 22.7 Intrinsic reaction coordinate (IRC) 6-31+G(d,p) (green) levels of theory. Projection of the IRC at mPW1B95/6-31+G(d,p) in dark gray, projection of the IRC at B3LYP/6-31+G(d,p) in gray and projection of IRC at M05-2x/6-31+G(d,p) (blue) and M05-2x/6-31+G(d,p) in black.

For the α -epimer, it is possible to find only one transition state (**46**) for the whole process. It would seem then that the stereo-electronic requirement [66] that must be satisfied by the C–H bond that migrates is not necessary.

Figure 22.7 presents the reaction trajectory for the transformation of intermediate **35** into **39** calculated at various levels of theory. At the mPW1B95/6-31+G(d,p) level it is possible to locate the stationary states associated with TSs **36** and **38**. As expected, as the conformational change occurs, the C1–H distance remains practically unmodified and only suffers an important change during the second stage of the process. At the B3LYP/6-31+G(d,p) level, this process is completely different. In the first segment of the trajectory, the conformational change occurs with minor changes in the C1–H distance and a constant increase in the energy until TS **38** (**52**, Figure 22.8) is reached. This is the point of maximum energy for the whole segment. From this moment on, relaxation of the system continues until minimum **39** is reached. The method is not capable of describing the conformational process previous to the migration of the hydrogen atom and locating an intermediate point of maximum energy that could be associated with the transition state of the conformational process. The third curve corresponds to the calculation at the M05-2x/6-31+G(d,p) [67] level and is similar to that obtained with the mPW1B95/6-31+G(d,p) functional because it separates the conformational process from the migration.

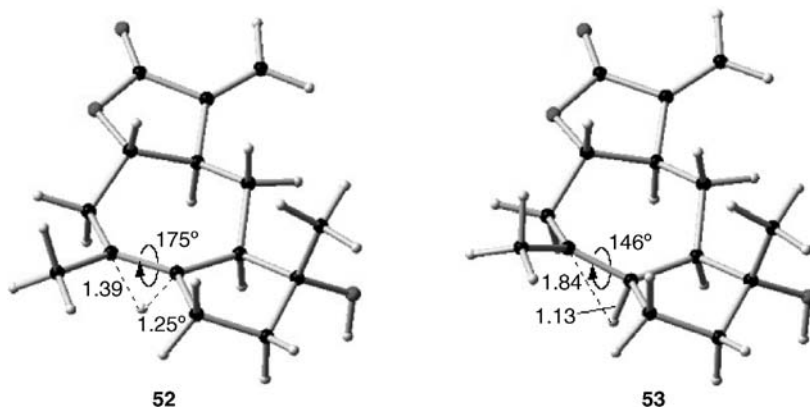


Figure 22.8 Transition states located in the reaction path of the reaction at the B3LYP/6-31+G(d,p) (**52**) and MP2/6-31+G(d,p) (**53**) levels; carbon in black, oxygen in gray and hydrogen in white.

The last segment of the curves describes the conformational readjustment with a pronounced change in the dihedral angle C14–C10–C1–C2 and a slight change in the C1–H distance once the migration has occurred.

Finally, when the mechanism is studied at the MP2 level, it is not possible to determine the transition state associated with the migration of the hydrogen atom. However, it is possible to determine the stationary state associated with the conformational change (**53**, Figure 22.8). Thus, the profile of the trajectory of reaction is very sensitive to the level of theory. This is because the potential energy surface is too flat in the TS region. It must be acknowledged, that there is a reaction pattern common to all curves. This shows that the physical phenomena are similar in all cases but the methods of calculation show different types of deficiencies when the phenomenon is described. The reaction trajectories described by the three levels of theory are shown in Figure 22.7 and are approximately parallel even when, for the case of the B3LYP functional, a new trajectory with a single TS could be generated. This trajectory could, in fact, be more direct or shorter to connect intermediates **35** and **38**. Now, a new question arises: why are these trajectories parallel? Even if the different methods do not locate the same stationary states, they do recognize the fact that the first segment of the reaction corresponds to the conformational change (this aligns the orbitals until the stereo-electronic requirements are satisfied) and the second segment corresponds to the transfer of a hydrogen atom. Evidently, a trajectory where the proton transference occurs at the same time as the conformational change should be discarded because it requires a reaction trajectory that is completely different from the previous because the changes in the C1–H bond length would happen at the same time as the changes in the C14–C10–C1–C2 angle, which is not observed here. This hypothetical case is illustrated in Figure 22.9. If the migration and the conformational change were simultaneous, the expected trajectory would be totally different.

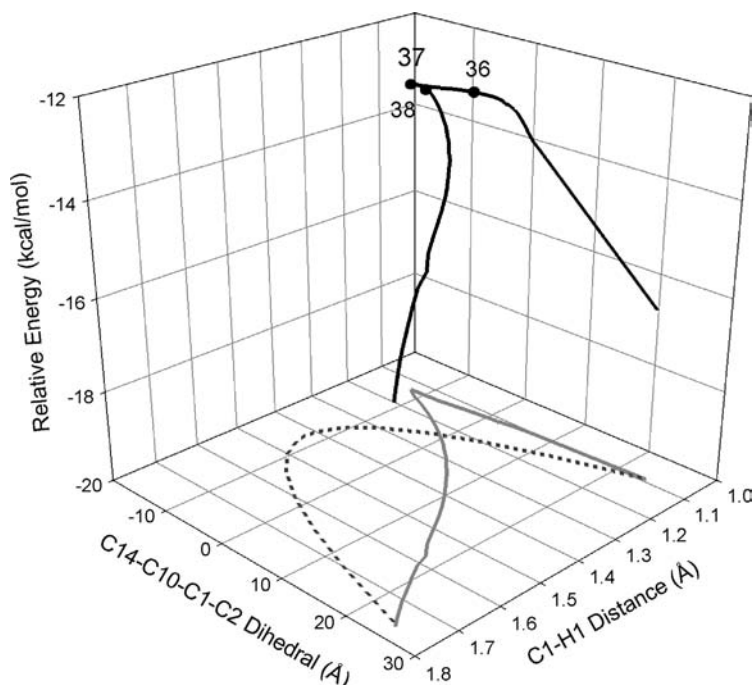


Figure 22.9 Hypothetical concerted (dotted line) and calculated trajectory for the transformation of **35** into **39**.

Although the B3LYP functional does not allow establishing stationary states for the conformational change, the trajectory involves the alienation of orbitals. If this alienation were not relevant, as proposed by Tantillo [68], there would be a reaction trajectory completely independent of the conformational change between stationary states **35** and **39**. Thus, the misconceived suggestion that the migration of the hydrogen atom is not subject to stereo-electronic control has its origin in the analysis of only one, or just a few, points of the reaction trajectories obtained using methodologies that do not adequately describe the potential energy surface, such as the B3LYP functional. In addition, a group of types of reaction coordinates denominated nearly plateau, well-defined shoulder, slight shoulder or no shoulder has been proposed in an attempt to systematize their description [68]. However, the trajectories depend so much on the level of theory that the controversy that arises can only be resolved in two ways. One is to ensure that the functional is correct for a specific surface of potential energy, which is impossible due to the small energetic difference between the different stationary states, and the second is to obtain experimental information about the intermediaries, which is difficult due to the close energetic proximity of the stationary states.

Figure 22.10 presents the reaction trajectory generated by the β -epimer calculated at the mPW1B95/6-31+G(d,p) level. As can be observed, the reaction trajectory shows that the hydrogen atom migrates only after the stereo-electronic requirements have been adjusted. Tantillo has stated that the condition a concerted mechanism

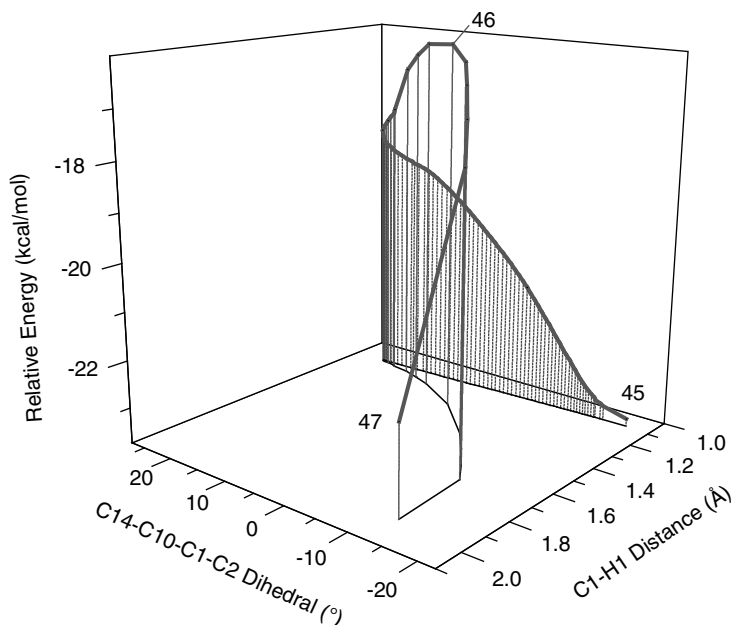


Figure 22.10 Intrinsic reaction coordinate (IRC) for the transformation of cation **45** into cation **47** at mPWb95/6-31 + G(d,p). In this case, once again the conformational arrangement is followed by the transference of the hydrogen atom.

requires simultaneous alignment for all relevant orbitals is not always true. However, this does not seem to be the case here since the migration does not happen until the conformational change happens [68].

At this point, the conditions that favor the occurrence of the second migration are present in intermediates **39** and **47**. For **39**, the hydrogen atom that will take part in that migration has the required geometry for this purpose as it maintains a *syn* relation with respect to the hydrogen atom that migrates from C1 to C10. In this case, **39** and **41** are almost isoenergetic and the barrier height is smaller than for other similar cases [69–71]. This migration is highly symmetric, since in the TS the C1–H distance is 1.31 Å and the C5–H distance is 1.33 Å. This stems from the fact that the two cations connected by TS **40** have a similar substitution pattern at the level of first neighbors. The existence of cation **41** is certainly proven by the isolation of epoxide **28** (Scheme 22.2). In contrast, in TS **48** the symmetry associated with its epimer **40** is lost. Again, the epimer with the C14 methyl in pseudo-equatorial position is more stable than the one that maintains it in a pseudo-axial position, by 1.49 kcal mol^{−1}. The charges adjusted to the electrostatic potential of the hydrogen atoms that migrate are again positive for both cases, 0.11 for **40** and 0.20 for **48** [65]. Now, it is relevant to examine whether the species that migrates is a hydride or a proton. To analyze the nature of the migrating atom, the TS of **54** to **56** that interconvert the corresponding cations were studied and are shown in Figure 22.11. For all cases, the charges adjusted to the electrostatic potential [65] show that the

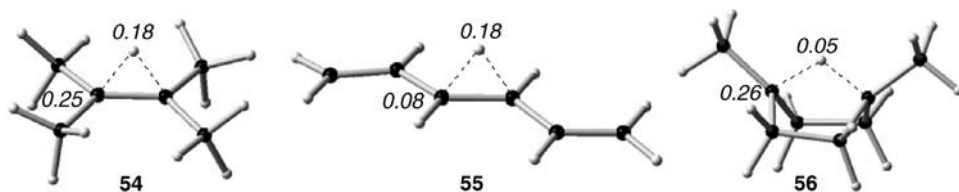


Figure 22.11 Transition states associated with the {1–2} transference of hydrogen. In all cases the atom's charge is positive.

hydrogen atom that migrates has lost the charge and this is distributed between the carbon atoms that participate in the process. It is a case of stabilization by 3c–2e intermediates and not the migration of a hydride. This is an example of where the application of so-called chemical language violates fundamental principles of physics and should be corrected.

Finally, protonated 8-epiconfertifin (**43**) is accessed through {1,2}-methyl shift from **41**. The methyl group in the TS is 1.74 Å from the origin and 1.96 Å from the destination atom. This makes it more similar to **41** than to the product **43**. Therefore, in accordance with Hammond's postulate, an exothermic reaction is expected when the product is generated. This is evident by the observation that the process concludes with the hydrogen atom joined to the carbonyl group. In fact, this bond evolved from 1.39 Å in the intermediate, to 1.34 Å in the TS and finally to 1.28 Å in the product. Notably, the depletion of charge in cation **43** is concentrated between C4 and O, since the C4 natural charge goes from 0.17 in **41** to 0.74 in **43** and the charge of the oxygen atom goes from –0.74 in **41** to –0.59 in **43**.

A similar migration was determined in the elemental step that includes the stationary states **49**–**51**. As in the case of migration of proton **48**, in **50** the transition state loses symmetry but gains stability. In all stationary states hereby analyzed the disposition of the O–H bond was restricted to the conformer that maintains an antiperiplanar arrangement in relation to a C4–Me bond, a product of the approximation of the proton below the epoxide, since it is 0.58 kcal mol^{–1} more stable than the cation derived from the approximation of the proton above the epoxide. Nevertheless, it must be expected that in some steps of the mechanism the participation of the OH bond is superseded by participation of the lone pair at the oxygen atom. It was found that when a lone pair at the oxygen atom is antiperiplanar to methyl C15 in the stationary states **41**–**43** the energy decreased by 2.54, 1.80 and 0.78 kcal mol^{–1}, respectively, at the mPW1B95/6-31+G(d,p) level of theory. These changes generate an increase in the energetic barrier associated with **41** and **42** from 2.12 to 2.86 kcal mol^{–1}, while the energetic barrier of **42** to **43** goes from 3.87 to 4.65 kcal mol^{–1}. Stationary state **41** is stabilized mainly, according to the deletion energies obtained by NBO analysis, by the interaction of σ_{C5-H} with the empty orbital of the carbocation, going from 31.61 to 36.16 kcal mol^{–1}, when the lone pair at oxygen is antiperiplanar to C15. In contrast, **43** is not stabilized for this reason, since interaction of the σ_{C4-C15} bond with the empty orbital of the carbocation decreases, going from 18.15 to 16.73 kcal mol^{–1} at mPW1B95/6-31+G(d,p) level of theory when the referred lone

pair is maintained antiperiplanar in relation to C15. Because in **42** the hyperconjugation with the empty orbital at the carbocation is weakened, **41** is stabilized more than **43**.

22.3

Conclusions

The Hendrickson [19] and Fischer [3, 38] theories of the biogenetic origin of guaianolides and pseudoguaianolides have full computational support since they characterize all the intermediates and TSs associated with the reaction mechanism. The computational results explain the origin of the intermediates isolated for the first time by Ortega and Maldonado [48], who in a one-pot transformation employing bentonitic earth realized three biogenetic stages of these compounds, as is described in Section 22.1: FPP $\rightarrow$ germacranolide, germacranolide $\rightarrow$ guaianolide, guaianolide $\rightarrow$ pseudoguaianolide (Scheme 22.1). In this instance, the transformation occurs through the more stable conformer, which is $^{15}D_5$, $^{14}D_{14}$. According to our results, some of the roles of the enzymes that catalyze these reactions at a biologic level would be the selection of the starting conformer and the stabilization of the intermediate cations. This makes it possible for them to undergo transposition instead of elimination reactions, as is the case for the origin of compound **27** (Scheme 22.2). Hence, only the next biogenetic stage is observed experimentally.

The transformation mechanism to produce pseudoguaianolides from cation **35** cannot occur in a concerted manner because the conformational adjustment and the migration of hydrogen atoms occur in elemental steps and a single TS that explains the transformation could not be found. The mechanism is driven since the reaction product is always more stable than its antecessor and because barriers are low. The higher barrier is for the migration of the {1,2}-methyl group with the formation of the carbonyl group, the final step in the biogenesis of pseudoguaianolides.

The mechanisms described herein support the hypothesis of Fischer [3, 38], who proposed that the reaction occurs through a different conformer. There is an additional proposal suggesting that the transformation of guaianolide to pseudoguaianolide occurs in only one step [23]. The migration of the methyl group at C4 as well as migration of the two hydrogen atoms with the elimination of a nucleofuge previously added to cation **35** requires only one TS. This is impossible, based on the results presented in this chapter, since several attempts to optimize such a singular TS were fruitless. This is because carbon C10 does not satisfy the stereochemical requirement that allows the migration of the hydrogen atom from position 1. It is necessary to remember that at the mPW1B95/6-31+G(d,p) level intermediate **35** must be transformed into conformer **37** through TS **36**. In this case a concerted mechanism, such as proposed by Herz [22], requires the fulfillment of all stereochemical requirements and this is not the case, or even worse, **35** is not transformed into **43** in one step. In contrast, Tantillo [68] has described that this is not always true and that the concerted mechanism requires

simultaneous alignment for all relevant orbitals. The reaction trajectories described by the mPW1B95 and B3LYP levels are somewhat parallel even when in the case of the B3LYP functional a new trajectory could be formed. This new trajectory could be more direct or even shorter between intermediates **35** and **39**. During the course of the reaction a conformational change happens. This change allows the correct alignment of the orbitals that favors the migration of the hydrogen atom under stereo-electronic control [66]. Even when B3LYP is not able to establish stationary states in this case, the trajectory brings about the alignment of the orbitals involved in the migration. If the alignment were not relevant, the reaction trajectory would be completely independent of the conformational change between stationary states **35** and **39**. The proper orientation of the orbitals involved is required for the hydrogen migration. Finally, the Ortega and Maldonado proposal is fully supported by computational methods.

References

- Herrera, A.L. (1942) *Science*, **96**, 14.
- Negrón-Mendoza, A. (1994) *J. Biol. Phys.*, **20**, 11–15.
- Lehninger, A.L. (1970) *Biochemistry*, Worth Publisher, New York, p. 483.
- Manitto, P. (1981) *Biosynthesis of Natural Products*, John Wiley & Sons, Inc., New York, p. 10.
- Clement, J.S., Mabry, T.J., Wyler, H., and Dreiding, A.S. (1994) Chemical review and evolutionary significance of the betalains, in *Caryophyllales* (eds H.D. Behnke and T.J. Mabry), Springer, Berlin, pp. 247–261.
- Croteau, R., Kutchan, T.M., and Lewis, G.N. (2000) Natural products (secondary metabolites), in *Biochemistry and Molecular Biology of Plants* (eds B.B. Buchanan, W. Gruissem, and R.L. Jones), American Society of Plant Physiologists, Rockville, Maryland, New York, Chapter 24.
- Taiz, L. and Zeiger, E. (eds) (2006) Secondary metabolites and plant defense, in *Plant Physiology*, 4th edn, Sinauer Associates, Inc., Chapter 13.
- Judd, W.S., Campbell, C.S., Kellogg, E.A., Stevens, P.F., and Donoghue, M.J. (2002) Secondary plant compounds, *Plant Systematics: A Phylogenetic Approach*, Sinauer Associates, New York, Chapter 4.
- Richards, J.H. and Hendrickson, J.B. (1964) *The Biosynthesis of Steroids, Terpenes and Acetogenins*, Benjamin, New York.
- Kuzuyama, T. and Seto, H. (2003) *Nat. Prod. Rep.*, **20**, 171–183.
- Rohmer, M., Seemann, M., Horbach, S., Bringer-Meyer, S., and Sahm, H. (1996) *J. Am. Chem. Soc.*, **118**, 2564–2566.
- Cane, D.E. (1989) *Pure Appl. Chem.*, **61**, 493–496.
- Fischer, N.H., Olivier, E.J., and Fischer, H.D. (1979) The biogenesis and chemistry of sesquiterpene lactones, in *Progress Chemistry Organic Natural Products* (eds W. Herz, H. Grisebach, and G.W. Kirby), Springer-Verlag, Wien, p. 47.
- Romo de Vivar, A. (1977) *Rev. Latinoam. Quim.*, **8**, 63–73; *Chem. Abstr.* 1997, **87**, 18934.
- Ruzicka, L. (1953) *Experientia*, **9**, 357–367.
- Ruzicka, L. (1963) *Pure Appl. Chem.*, **6**, 493–523.
- Cane, D.E. (1989) *Pure Appl. Chem.*, **61**, 493–496.
- Cane, D.E. (1990) *Chem. Rev.*, **90**, 1089–1103.
- Hendrickson, J.B. (1959) *Tetrahedron*, **7**, 82–89.
- Barton, D.H.R., Bockmann, O.C., and de Mayo, P. (1960) *J. Chem. Soc.*, 2263–2271.
- Parker, W., Roberts, J.S., and Ramaje, R. (1967) *Quart. Rev.*, **21**, 331–363.
- (a) Herz, W. (1971) Sesquiterpene lactones biogenesis, in *Pharmacognosy and Phytochemistry* (eds H. Wagner and L. Horhammer), Springer-Verlag, West Berlin

- and Heidelberg, p. 64; (b) Herz, W. (1977) *Israel J. Chem.*, **16**, 32–44.
- 23 Hanson, J.R. (1979) Terpenoid biosynthesis, in *Comprehensive Organic Chemistry* (ed. D.H.R. Barton), Academic Press, New York, Chapter 29.
 - 24 Geissman, T.A. (1973) *Recent Adv. Phytochem.*, **6**, 65–95.
 - 25 Fischer, N.H., Oliver, E.J., and Fischer, H.D. (1979) *Prog. Chem. Org. Nat. Prod.*, **38**, 77–390.
 - 26 (a) Cane, D.E. (1999) Isoprenoids including carotenoids and steroids, in *Comprehensive Natural Products Chemistry*, vol. 2 (ed. D.E. Cane), Elsevier, Oxford, p. 155; (b) Felicetti, B. and Cane, D.E. (2004) *J. Am. Chem. Soc.*, **126**, 7212.
 - 27 (a) Goodwin, T.W. (ed.) (1970) *Natural Substances formed Biologically from Mevalonic Acid, Biochemical Symposia*, No. 29, Academic Press; (b) MacMillan, J. (1974) Recent aspects on the chemistry and biosynthesis of the Gibberellins. *Recent Adv. Phytochem.*, **7**, 1–19.
 - 28 Delgado, G. (2006) Investigación sobre la Química de productos naturales en el Instituto de Química de la UNAM. Estudios iniciales y química de eremofilanos, bisabolanos y sesquiterpenos relacionados, in *Química de la Flora Mexicana* (ed. A. Romo de Vivar), Instituto de Química, Mexico.
 - 29 Christianson, D.W. (2006) *Chem. Rev.*, **106**, 3412–3442.
 - 30 (a) Dewick, P.M. (1999) *Nat. Prod. Rep.*, **16**, 97–130; (b) Dewick, P.M. (2002) *Nat. Prod. Rep.*, **19**, 181–272.
 - 31 Steel, C.L., Crock, J., Bohlman, J., and Croteau, R. (1998) *J. Biol. Chem.*, **273**, 2078.
 - 32 Starcks, C.M., Back, K., Chapell, J., and Noel, J.P. (1997) *Science*, **277**, 1815–1820.
 - 33 Coates, R.M. (1976) *Prog. Chem. Org. Nat. Prod.*, **33**, 73–230.
 - 34 Goldsmith, D. (1971) *Prog. Chem. Org. Nat. Prod.*, **29**, 363–394.
 - 35 Money, T. (1973) *Prog. Chem. Org. Chem.*, **8**, 29–77.
 - 36 (a) Allermann, R.K., Young, N.J., Ma, S., Truhlar, D.G., and Gao, J. (2007) *J. Am. Chem. Soc.*, **129**, 13008–13013; (b) Chapell, J. (1995) *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **46**, 521–547; (c) McCaskill, D. and Croteau, R. (1995) *Planta*, **197**, 49–56; (d) van Klink, J., Becker, H., Anderson, S., and Boland, W. (2003) *Biomol. Chem.*, **1**, 1503–1508.
 - 37 de Kraker, J.-W., Franssen, M.C.F., de Groot, A., König, W.A., and Bouwmeester, H.J. (1998) *Plant Physiol.*, **117**, 1381–1392.
 - 38 Fischer, N.H. (1978) *Rev. Latinoam. Quím.*, **9**, 41–46.
 - 39 Rao, A.S., Keilkar, G.R., and Bhattacharyya, S.C. (1960) *Tetrahedron*, **9**, 275–283.
 - 40 Fischer, N.H., Mabry, T.J., and Kagan, H.B. (1968) *Tetrahedron*, **24**, 4091–4097.
 - 41 Delgado, G., Tejeda, V., Salas, A., Chavez, M.I., Guzmán, S., Bolaños, A., Aguilar, M.I., Navarro, V., and Villareal, M.L. (1998) *J. Nat. Prod.*, **61**, 1082–1085.
 - 42 Benesova, V., Samek, Z., Herout, V., and Sorm, F. (1970) *Tetrahedron Lett.*, **11**, 5017–5020.
 - 43 McPhail, A.T., Onan, K.D., Lee, K.H., Ibuka, T., and Huang, H.-Ch. (1974) *Tetrahedron Lett.*, **15**, 3203–3206.
 - 44 Fischer, N.H., Seaman, F.C., Wiley, R.A., and Haegele, K.D. (1978) *J. Org. Chem.*, **43**, 4984–4987.
 - 45 González-González, A., Arteaga, J.M., and Breton-Funes, J.L. (1973) *Phytochemistry*, **12**, 2997.
 - 46 For the transformation of epoxygermacranolides into guaianolides see: (a) White, E.H. and Winter, R.E.K., (1963) *Tetrahedron*, **19**, 137–141; (b) Govindachari, T.R., Josi, B.S., and Kamat, V.N. (1965) *Tetrahedron*, **21**, 1509–1519; (c) Gaissman, T.A. and Ellestad, G.A. (1971) *Phytochemistry*, **10**, 2475–2485; (d) Griffin, T.S., Geissman, T.A., and Winters, T.W. (1971) *Phytochemistry*, **10**, 2487–2495; (e) Irwin, M.A., Lee, K.H., Simpson, R.F., and Geissman, T.A. (1969) *Phytochemistry*, **8**, 2009–2012.
 - 47 For the transformation of 4, 5-epoxyguaianolide into pseudoguaianolides see: Fischer, N.H., Wiley, R.A., and Perry, D.L., (1976) *Rev. Latinoam. Quím.*, **7**, 87–93.
 - 48 Ortega, A. and Maldonado, E. (1989) *Heterocycles*, **29**, 635–638.
 - 49 Takeda, K. (1974) *Tetrahedron*, **30**, 1525–1534.
 - 50 Samek, Z. and Harmatha, J. (1978) *Coll. Czech. Chem. Commun.*, **43**, 2779–2799.

- 51 Parker, W., Roberts, J.S., and Ramaje, R. (1967) *Quart. Rev.*, **21**, 331.
- 52 Fischer, N.H., Wu-Shih, Y.F., Chiari, G., Fronczek, F.R., and Watkins, S.F. (1981) *J. Nat. Prod.*, **44**, 104–110.
- 53 Barquera-Lozada, J.E. and Cuevas, G. (2009) *J. Org. Chem.*, **72**, 874–883.
- 54 (a) Zhao, Y. and Truhlar, D.G. (2004) *J. Phys. Chem. A*, **108**, 6908–6918; (b) Zhao, Y. and Truhlar, D.G. (2008) *Acc. Chem. Res.*, **41**, 157–167.
- 55 Lynch, B.J., Zhao, Y., and Truhlar, D.G. (2003) *J. Phys. Chem. A*, **107**, 1384–1388.
- 56 Frisch, M.J., Trucks, G.W., Schlegel, H.B., et al. (2004) Gaussian 03, Revision D. 01, Gaussian, Inc., Wallingford CT.
- 57 Pawar, D.M., Smith, S.V., Mark, H.L., Odom, R.M., and Noe, E.A. (1998) *J. Am. Chem. Soc.*, **120**, 10715–10720.
- 58 Fernández-Alonso, M.C., Asensio, J.L., Cañada, F.J., Jiménez-Barbero, J., and Cuevas, G. (2003) *Chem. Phys. Chem.*, **4**, 748–753.
- 59 Fernández-Alonso, M.C., Cañada, J., Jiménez-Barbero, J., and Cuevas, G. (2005) *Chem. Phys. Chem.*, **6**, 671–681.
- 60 Juaristi, E. and Cuevas, G. (2008) *Acc. Chem. Res.*, **40**, 961–970.
- 61 Laube, T. (1995) *Acc. Chem. Res.*, **28**, 399–405.
- 62 (a) Carpenter, J.E. and Weinhold, F. (1988) *J. Mol. Struct. (THEOCHEM)*, **169**, 41–62; (b) Foster, J.P. and Weinhold, F. (1980) *J. Am. Chem. Soc.*, **102**, 7211; (c) Carpenter, J.E. (1987) PhD Thesis, University of Wisconsin, Madison, WI. (d) Reed, A.E. and Weinhold, F. (1983) *J. Chem. Phys.*, **78**, 4066–4073; (e) Reed, A.E. and Weinhold, F. (1985) *J. Chem. Phys.*, **83**, 1736–1740; (f) Reed, A.E., Weinstock, R.B., and Weinhold, F. (1985) *J. Chem. Phys.*, **83**, 735–746; (g) Reed, A.E., Curtiss, L.A., and Weinhold, F. (1988) *Chem. Rev.*, **88**, 899–926.
- 63 Smith, M.B. and March, J. (2007) *March's Advanced Organic Chemistry: Reactions, Mechanisms and Structure*, John Wiley & Sons Inc., Hoboken, p. 1581.
- 64 Carey, F.C. and Sundberg, R.J. (1990) *Advanced Organic Chemistry (Part A)*, Plenum, p. 317.
- 65 Breneman, C.M. and Wiberg, K.B. (1990) *J. Comput. Chem.*, **11**, 361–373.
- 66 (a) Deslongchamps, P. (1983) *Stereoelectronic Effects in Organic Chemistry*, Pergamon Press, Oxford; (b) Szarek, W.A. and Horton, D. (eds) (1979) *Anomeric Effect: Origin and Consequences (ACS Symposium Series, No. 87)*, American Chemical Society, Washington DC; (c) Kirby, A.J. (1983) *The Anomeric Effect and Related Stereoelectronic Effects at Oxygen*, Springer, New York; (d) Juaristi, E. and Cuevas, G. (1992) *Tetrahedron*, **48**, 5019–5087; (e) Thatcher, G.R.J. (ed.) (1993) *The Anomeric and Associated Stereoelectronic Effects*, American Chemical Society, Washington DC; (f) Graczyk, P.P. and Mikolajczyk, M. (1994) *Top. Stereochem.*, **21**, 159–349; (g) Juaristi, E. and Cuevas, G. (2002) *The Anomeric Effect*, CRC Press, Boca Raton, FL; (h) Chattopadhyaya, J. (1999) *Stereoelectronic Effects in Nucleosides and their Structural Implications*, Uppsala University Press, Uppsala; (i) Perrin, C.L. (2002) *Acc. Chem. Res.*, **35**, 28–34.
- 67 Zhao, Y., Schultz, N.E., and Truhlar, D.G. (2006) *J. Chem. Theory Comput.*, **2**, 364.
- 68 Tantillo, D.J. (2008) *J. Phys. Org. Chem.*, **21**, 561–570.
- 69 Gutta, P. and Tantillo, D.J. (2006) *J. Am. Chem. Soc.*, **128**, 6172–6179.
- 70 Vrček, I.V., Vrček, V., and Siehl, H. (2002) *J. Phys. Chem. A*, **106**, 1604–1611.
- 71 Vrček, V., Vrček, I.V., and Siehl, H. (2006) *J. Phys. Chem. A*, **110**, 1868–1874.

23

Mechanistics of Enzyme Catalysis: From Small to Large Active-Site Models

Jorge Llano and James W. Gauld

23.1

Introduction

Density functional theory (DFT), along with the steady increase in computing power and algorithmic efficiency, has opened new realms for the application of quantum chemistry to the study of complex materials and biomolecules. Computational enzymology, computational modeling of the structure and catalytic mechanism of enzymes at the atomistic level have benefited profusely from such advances. In particular, it has allowed for gradually more accurate and realistic chemical models. This in turn has enabled researchers to elucidate and predict the chemical behavior of many enzymes from their molecular structure. This chapter reviews the factors that influence the catalytic performance of enzymes, their relative energetic weight and the computational methods applied in their estimation. Then, our attention is focused on highlighting, through mechanistic studies carried out in our group, how careful selection of the chemical model can accurately grasp much of the chemistry taking place within the enzyme and, more generally, within a family of enzymes.

23.1.1

Factors Influencing the Catalytic Performance of Enzymes

The origin, rate-accelerating power, specificity and efficiency of enzymatic catalysis have been theoretically attributed to several factors:

- 1) transition-state stabilization [1, 2];
- 2) enzyme–substrate binding entropy [3–7];
- 3) desolvation [8–12];
- 4) steric effects (e.g., reaction intra-molecularity [4], geometric complementarity of the enzyme to both substrate and reaction transition state [1, 2], alignment of catalytic functional groups, substrate distortion, destabilization and near-attack conformations [13–17]);

- 5) electrostatic pre-organization of the active site [18–27];
- 6) enzyme conformational dynamics (e.g., coupled protein motions, reorganization or relaxation [28–40]);
- 7) quantum mechanical and dynamical effects (e.g., orbital steering [41–44], tunneling [45–54] and barrier recrossing [47–49, 55]);
- 8) ionic low-barrier hydrogen bonds [56–59] and donor–acceptor matching pK_a s [60, 61].

The dominance of one of these factors over the others as well as their relative contribution to the rate-enhancement, specificity and efficiency of enzymes has been the subject of much debate over the years [5, 16, 47, 62–65]. In general, however, it appears that none of these factors on its own can qualitatively and quantitatively account for the catalytic power of a single enzyme. Instead, the combination of two or more of these effects can describe, to certain accuracy, the catalytic behavior of a single enzyme or its family or both. Therefore, it is worth examining the most recent energetic estimates for several of these effects and discussing the distinct features of the enzymes in which they manifest with a marked contribution.

Based on kinetic studies of hydrolysis, hydration, decarboxylation, isomerization, migration, deamination and aminolysis reactions following the same reaction mechanism, both in dilute aqueous solution and in the presence of the enzymes [66–81], Wolfenden has gathered valuable data to quantify enzymatic rate enhancement. In particular, binding affinities of a typical enzyme for its transition state ($-\Delta_{\text{bind}}G_{\text{TS}}^0$) and for its substrate ($-\Delta_{\text{bind}}G_{\text{S}}^0$) were assessed from the estimates of the catalytic constant or turnover number (k_{cat}), the Michaelis constant (K_{m}) and the rate constant of the uncatalyzed reaction (k_{uncat}) in aqueous solution. The value of $-\Delta_{\text{bind}}G_{\text{TS}}^0$ was found to range from approximately 32 to 160 kJ mol⁻¹, while $-\Delta_{\text{bind}}G_{\text{S}}^0$ was on average approximately 23 kJ mol⁻¹. The parameters k_{cat} , k_{uncat} and K_{m} are often combined in relations that measure catalytic performance, such as the specificity constant or catalytic efficiency ($k_{\text{cat}}/K_{\text{m}} \approx 10^5\text{--}10^9 \text{ M}^{-1} \text{ s}^{-1}$ in water) [70, 78, 82], rate enhancement ($k_{\text{cat}}/k_{\text{uncat}}$) and catalytic proficiency [$(k_{\text{cat}}/K_{\text{m}})/k_{\text{uncat}}$] [70, 75, 78]. The rate of diffusion of substrate molecules to the enzyme (up to $10^9 \text{ M}^{-1} \text{ s}^{-1}$ in water) is the upper bound for the specificity constant, at which point K_{m} becomes the dissociation constant of the enzyme–substrate complex and is estimated to typically be around 10^{-4} M [83]. Finally, first-order rate constants of reactions catalyzed by typical enzymes were found to be in the range $10^{-20}\text{--}10^{-1} \text{ s}^{-1}$ [75, 77, 78]. When all these kinetic data are recast into energetic terms, a Gibbs energy diagram can be outlined for the potential energy surface (PES) of an enzyme-catalyzed reaction (Figure 23.1). This profile illustrates enzyme–substrate specificity and transition-state stabilization, and thereby confirms Pauling’s hypothesis about the catalytic power of enzymes.

Basically, $\Delta_{\text{bind}}G_{\text{TS}}^0$ is composed of an enthalpic component ($\Delta_{\text{bind}}H_{\text{TS}}^0$), which is related to the strength of the stabilizing forces that keep the transition structure held to the active site, and an entropic component ($T\Delta_{\text{bind}}S_{\text{TS}}^0$), which is related to the positioning of the substrate in the active site during reaction. For first-order

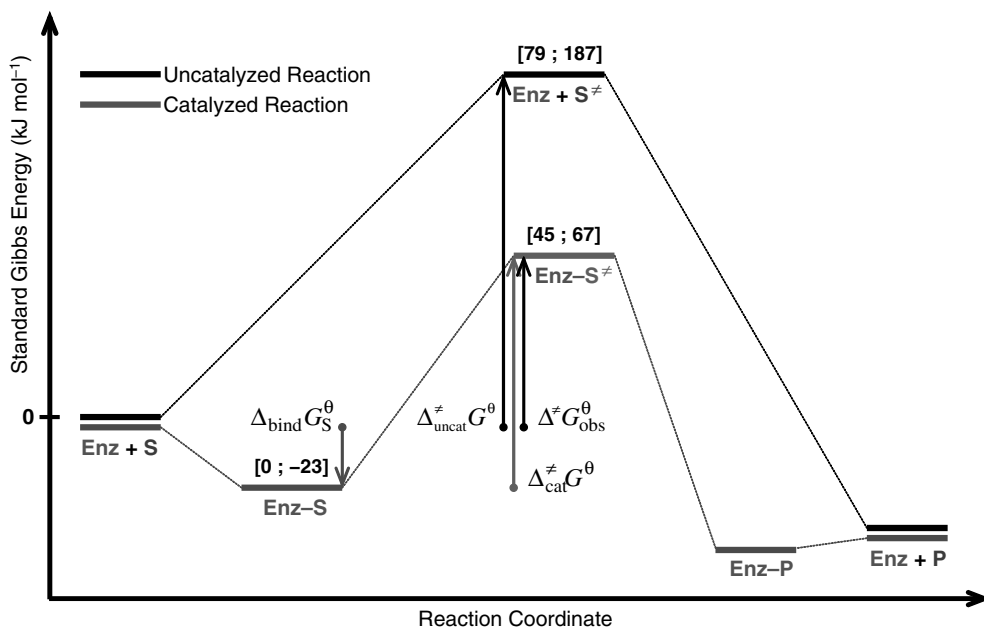


Figure 23.1 Standard Gibbs energy profiles of a typical uncatalyzed reaction in aqueous solution and the enzyme-catalyzed counterpart that fulfils a Michaelis–Menten kinetic scheme at 298 K. The observed standard Gibbs energy of activation for the catalyzed reaction ($\Delta^\ddagger G_{\text{obs}}^\theta$) is equal to the barrier to formation of product through an

enzyme-bound transition state ($\Delta^\ddagger G_{\text{cat}}^\theta$) minus the enzyme–substrate binding energy ($\Delta_{\text{bind}} G_S^\theta$). The standard Gibbs energy of activation observed for a typical enzyme ($\Delta^\ddagger G_{\text{obs}}^\theta$) is much smaller than that of the reference reaction occurring in aqueous solution ($\Delta^\ddagger G_{\text{uncat}}^\theta$). Typical Gibbs energy ranges are given in square brackets.

and pseudo first-order reactions (i.e., one-substrate enzymatic reactions), Wolfenden has shown that the enthalpic contribution dominates over the entropic contribution [69, 79, 84]. However, for second-order reactions (i.e., two-substrate enzymatic reactions), the catalytic effect can be highly entropic in origin. For instance, $\Delta_{\text{bind}} G_{\text{TS}}^\theta$ for the aminoacyl transfer step in the peptidyl transferase centre of the ribosome is 40 kJ mol^{-1} , with enthalpic and entropic contributions of 34 and 74 kJ mol^{-1} , respectively. These values suggest that the aminolytic transition structure fits a highly organized active site that tightly aligns the peptidyl- and aminoacyl-tRNAs, thereby reducing the translational and rotational entropy costs associated with formation of the equivalent transition structure in aqueous solution [74, 79, 85]. The mechanism and transition state of peptide bond formation in the ribosome is reviewed in Chapter 16.

In a similar vein, the entropic component of substrate binding to the enzyme $T\Delta_{\text{bind}} S_S^\theta$ has been credited with catalytic enhancing power, since the number of degrees of freedom for the reacting fragments in water are drastically restricted in the enzyme's active site [3, 4]. Hence, the entropic contribution to the Gibbs energy

of activation, $-T(\Delta_{\text{bind}}S_{\text{S,cat}}^0 - \Delta_{\text{bind}}S_{\text{S,uncat}}^0)$, should turn out to be larger in the enzyme than in the reference aqueous solution. This entropic effect has been estimated to be about -10 kJ mol^{-1} for the hydrolysis of the Tyr-Gly dipeptide in the active site of subtilisin [6, 7], which is in good agreement with the standard Gibbs energy cost ($RT\ln 55$) of bringing the reactants to form the enzyme–substrate complex in an infinite-dilute aqueous solution at 298 K.

In another view of substrate binding, the desolvation hypothesis proposes that some enzymes may solvate substrate molecules less than water does [4, 9, 21, 65]. Hence, polarizing a bond of the substrate by a charged group would be easier within a nonpolar active site than in water. In contrast, other authors view desolvation as the larger stabilization of the substrate relative to the transition structure, both within the enzyme and in water, with respect to the gas phase; with the effect being stronger within the enzyme [10, 12, 47, 86]. According to the latter view, desolvation has been estimated to amount to $25\text{--}34 \text{ kJ mol}^{-1}$ for the S_N2 displacement of a chloride ion by a carboxylate in the esterification step of haloalkane dehalogenase [11, 12].

In many cases, the process of substrate binding to the enzyme is believed to proceed in concert with a structural reorganization that pre-aligns and orients both the enzyme's catalytic groups and the substrate's reacting groups into a reactive configuration resembling the transition state. This is the basis of the near-attack conformation (NAC) hypothesis of catalytic enhancing power of enzymes [13–17]. In particular, estimates in the range of about $2\text{--}6 \text{ kJ mol}^{-1}$ from the NAC in enzymes have been calculated from computer simulations as the difference in the binding Gibbs energies of the substrate in the near-attack conformation ($\Delta_{\text{bind}}G_{\text{S(NAC)}}^0$) and in the relaxed bound conformation ($\Delta_{\text{bind}}G_{\text{S}}^0$). This effect is enthalpic and entropic in origin, with an enhancement ($\Delta_{\text{bind}}G_{\text{S(NAC),cat}}^0 - \Delta_{\text{bind}}G_{\text{S(NAC),uncat}}^0$) ranging from -10 to -40 kJ mol^{-1} on average, for the following steps: S_N2 transmethylation of catechol catalyzed by catechol O-methyltransferase [14, 16, 87], nucleophilic attack of methanol to the carbonyl carbon of *N*-methylacetamide catalyzed by trypsin [16, 88], isomerization of chalcone catalyzed by chalcone isomerase [15] and the Claisen rearrangement of chorismate catalyzed by chorismate mutase [15].

Regarding the enzyme's capability of aligning and positioning the reacting chemical groups, Warshel's electrostatic hypothesis proposes that the protein makes up a pre-organized electrostatic environment that provides a driving force for substrate binding, orientation and the subsequent catalytic cascade of steps occurring within the enzyme [18–22, 24–27]. More specifically, reactions occurring in water are subjected to an isotropic polar environment with a homogenous dielectric constant of about 80. In contrast, in the active site, the substrate finds a highly complementary and anisotropic electrostatic field that is made from permanent dipoles and charges of the residue's side-chains and the protein backbone. Hence, the protein environment can be assigned an inhomogeneous dielectric constant that fluctuates approximately between 2 to 10 depending on the site of the protein considered [23, 26]. The protein electrostatic stabilization effect with respect to water has been estimated to be in the range $20\text{--}105 \text{ kJ mol}^{-1}$ [27].

In another perspective, structural, computational and kinetic experiments (e.g., NMR spin relaxation and molecular dynamics simulation techniques) have shown that certain conformational transitions in the domain motions of some enzymes occur in time scales (typically within 10^{-6} – 10^{-3} s) that are near or of the order of the enzyme's catalytic constant [28–30, 32, 37]. In those cases, the enzyme conformational dynamics may play a decisive role in the catalytic performance, or even play a rate-limiting role in the enzyme's catalytic cycle. This hypothesis has been thoroughly examined in hydride-transferring enzymes, which display domain motions that allow opening and closing of the pocket wherein the active site is buried, as in the cases of dihydrofolate reductase (DHFR [33, 35, 36, 39]) and liver alcohol dehydrogenase (LADH [35]). Among these domain motions, the specific conformational transition bringing the donor and acceptor together for hydride transfer has been found to have a significant role in the catalytic performance of these enzymes, even if the conformational transition is not mechanistically coupled with the hydride transfer step. For instance, the pH-independent catalytic rate constant of DHFR is 950 s^{-1} and the conformational transition bringing the hydride donor and acceptor together has a rate constant of approximately 400 s^{-1} [33, 35].

The existence of quantum mechanical effects has also been invoked to rationalize the catalytic enhancing power of enzymes. One of the earliest quantum chemical proposals was the orbital steering hypothesis, which assumes that maximization of molecular orbital overlap in the bonding interaction of two reacting species governs the stereochemistry of their approach [41–44]. In fact, orbital steering is perhaps the hypothesis that enjoys the least acceptance, possibly because the underlying phenomenon behind it cannot be confirmed experimentally, and because simpler hypotheses have unambiguously provided clearer quantitative evidence of catalytic enhancing performance.

In contrast, the quantum phenomena of transfer of electrons and hydrogen cations (H^+), atoms ($\text{H}^\bullet$) and anions (H^-) have enjoyed widespread attention, since electron and nuclear transfers are the reaction type that defines the oxidoreductase class of enzymes. Under certain conditions, quantum particles having a total energy slightly lower than that required to cross a finite potential energy barrier can pass through it, instead of over it, by the phenomenon of tunneling [46–49, 51, 53]. The short- and long-range electron transfer mechanisms in biomolecules have been well investigated by applying Marcus theory and variants of it [89–91]. This interest has extended to the study of nuclear transfer reactions. Several examples of enzymes show that H^+ , $\text{H}^\bullet$ and H^- particles with energies in a range of approximately 5 – 12 kJ mol^{-1} lower than classical barrier heights of 20 – 60 kJ mol^{-1} may penetrate the barrier at 298 K [45–50, 54]. Thus, the inclusion of nuclear quantum tunneling in the theoretical prediction of catalytic constants may increase the rate by one (and up to two) order(s) of magnitude.

For one thing, proton transfer between hydrogen bonded groups ($\text{D-H}\cdots\text{A} \rightarrow \text{D}\cdots\text{H-A}$) typically occurs at physiologically relevant rates when there is a difference of no greater than 2 to 3 pH units in the pK_a s of the donor D-H and the protonated form of the acceptor H-A . Inspired by the Marcus formalism, the ionic low-barrier hydrogen bond (LBHB) [56–59] and donor–acceptor matching

pK_a s [60, 61] hypotheses propose that, unlike water, the protein environment can match the pK_a s of D–H and H–A, thereby creating a single-well potential or a low-barrier double-well potential for the proton. Consequently, the proton would essentially end up shared by the two centers ($D \cdots H \cdots A$) and, thus, the rate of proton transfer in chains of LBHBs would be significantly enhanced. These hypotheses have been postulated to explain the unusually short and strong hydrogen bonds observed in the NMR derived structures of serine proteases and certain isomerases that catalyze keto–enol conversions [56, 61]. However, it has been argued that the structure $D \cdots H \cdots A$ has a highly covalent component that would be favored in a non-polar environment but not in the anisotropically polar environment of an active site, in which the ionic structures $D-H \cdots A$ or $D \cdots H-A$ would be stabilized and prevail [57, 59].

Finally, catalytic enhancing power has also been ascribed to the chemical dynamical effect of barrier recrossing. This effect is formulated as the recrossing transmission coefficient, which is defined as the probability that decay of the transition state occurs in the direction of product formation rather than back to the reactants [47–49]. For several enzyme-catalyzed reactions, transmission coefficients have been found to be in the range 0.36–0.99 with a median value of approximately 0.75 [48]. This implies that the effective Gibbs energy of activation could be lowered by up to 2.5 kJ mol^{-1} , which is an insignificant amount compared to the errors of many computational simulations used to estimate these effects.

23.1.2

Computational Modeling in Enzymology

Undoubtedly, the quantitative weight of each of the contributions previously discussed cannot be obtained from experiment alone. Hence, computational simulations and chemical models have been designed to calculate energetic and kinetic estimates of those catalytic effects. These simulations have the challenge of reaching an accuracy lower than 5 kJ mol^{-1} , as an error of 5.7 kJ mol^{-1} in the Gibbs energy of activation brings about an error of one order of magnitude in any estimated rate constant at 298 K. Thus, the theoretical approaches applied in computational enzymology have explored several alternatives, always combining elements of the following categories:

- 1) chemical models suitable to represent the enzymatic system and catalytic process (i.e., full protein domain models, active-site models and small model-compound approaches);
- 2) physical theory used to describe the system's total energy (i.e., classical mechanics, electrodynamics, quantum mechanics and hybrids thereof);
- 3) time-dependence of the chemical process and system's total energy (i.e., stationary or dynamical treatments of the physical problem).

Computational simulations of reactions mechanisms including full protein domains surrounded by solvent have been possible by using QMMM methods [92, 93] that combine a quantum mechanical (QM) description for the reactive

subsystem or chemically active region (e.g., substrate, side-chains of active-site residues and backbone peptide bonds) with a molecular mechanical (MM) description for the surrounding environment (e.g., the reminder of the protein body and solvent). Equations 23.1 and 23.2 represent the system's total energy partitions for the original QMMM [18, 21, 94, 95] and ONIOM(QM:MM) [96–105] schemes, respectively (QM:MM and ONIOM methods are also reviewed in Chapters 2–4):

$$E_{\text{tot}}^{\text{QMMM}} = E_{\text{RS}}^{\text{QM}} + E_{\text{SE}}^{\text{MM}} + E_{\text{RS}\cap\text{SE}}^{\text{QM}\cap\text{MM}} \quad (23.1)$$

$$E_{\text{tot}}^{\text{ONIOM}} = E_{\text{RS}\cup\text{SE}}^{\text{MM}} - E_{\text{RS}}^{\text{MM}} + E_{\text{RS}}^{\text{QM}} \quad (23.2)$$

where for the original QMMM scheme, $E_{\text{RS}}^{\text{QM}}$ is the QM energy of the reactive subsystem, $E_{\text{SE}}^{\text{MM}}$ is the MM energy of the surrounding environment and $E_{\text{RS}\cap\text{SE}}^{\text{QM}\cap\text{MM}}$ is the energy of coupling and interaction of the QM and MM layers. For the ONIOM(QM:MM) scheme, $E_{\text{RS}\cup\text{SE}}^{\text{MM}}$ is the MM energy of the whole system and $E_{\text{RS}}^{\text{MM}}$ and $E_{\text{RS}}^{\text{QM}}$ are the MM and QM energies of the reactive subsystem of atoms, respectively. In fact, the ONIOM scheme has shown to be particularly useful in the stationary treatment of some environmental effects on the PES of large active-site models [106] and to calculate, for instance, the binding energy of dioxygen to the non-heme Fe(II)-dependent isopenicillin N synthase, as it contains a significant component due to van der Waals interactions [107].

The original QMMM scheme has proved to be suitable for integration into dynamical treatments of molecular motions. In Warshel's and Åqvist's approach [21, 24, 25, 65, 108, 109], classical potential energy functions for the Lewis resonance structures that describe reactant, intermediate and product states are parameterized using high-accuracy methods based on wavefunction theory (WFT) or density functional theory (DFT). These functions constitute an empirical valence bond (EVB) force field for the reaction coordinate. The EVB parameters are then incorporated into the topology of the force field for the whole system. Next, the PES of the enzymatic reaction is calibrated using the PES of the reaction in solution by applying the molecular dynamics (MD) free energy perturbation (FEP) technique. In Kollman's approach [16, 87, 94, 110], quantum mechanical and classical free energy calculations (QM-FE) are combined in the following manner: residues from the chemically active region are anchored to their original positions within the protein's frame of reference; then, the reaction PES for the chemically active subsystem is computed with a high-level quantum Hamiltonian; and finally, charges for the atoms of the reactive subsystem are generated at the reactant, transition, intermediate and product states in order to be incorporated into the force field to perform the MD-FEP simulation for each state of the reaction path. Applying the original QMMM scheme [18, 94, 95], Rothlisberger has devised the QM-CPMD method [111–113]. This approach treats the chemically active subsystem with DFT, plane wave basis functions, pseudo-potentials and Car–Parrinello molecular dynamics techniques (CPMD), and the surrounding environment with a classical force field.

For catalytic mechanisms in which the protein dynamics and strain as well as long-range van der Waals and electrostatic interactions do not have distorting influences

on the reaction PES, an active-site model of the reactive subsystem has shown to be a remarkably accurate description [114–124]. For example, such an approach has been found to be particularly applicable for a number of enzymes that employ covalent and metal ion catalysis. Indeed, in many cases, essentially the same PES was found when the long-range interactions were included in larger active-site models and full protein domain models [125]. An active-site model consists of a whole substrate molecule or the chemically active part of it, and amino acid side-chains conveniently anchored to their crystal structure positions within the protein's frame of reference [120]. The purpose of the active-site model is to include all electrostatic, quantum and dynamical effects of the interactions of the substrate with the nearby amino acid side chains and protein backbone. Computationally, QMMM and high-level DFT combined with self-consistent reaction field (SCRf) methods have been applied in the framework of the active-site model approach [114, 118, 119, 122]. In fact, as previously discussed, an active-site model calculation is a prerequisite in the routine of a number of stationary and dynamical QMMM schemes [16, 87, 92, 93, 106, 107, 110].

Finally, the study of reaction mechanisms combining highly accurate WFT or DFT methods with SCRf techniques, and small model-compounds allows dissection of the active site into single interactions as well as their gradual combination in order to estimate their mutual cooperativity and modulation [126–128]. Indeed, such an approach has been shown to be able to give detailed insight into the cumulative and competing effects of individual interactions that occur in the active site.

23.2

Active-Site Models of Enzymatic Catalysis: Methods and Accuracy

The enzymatic mechanistic studies discussed in this chapter followed the computational procedure developed by Siegbahn and Blomberg for the investigation of metalloenzymes [114–116, 119–124]. The protocol applied herein consisted of the hybrid density functional B3LYP [129], a combination of Becke's three-parameter hybrid exchange functional [130] and the Lee–Yang–Parr correlation functional [131], as available in the programs Gaussian 03 [132] and Jaguar 5.5 [133]. The atomic electron densities of metal centers were described with the Wadt and Hay core-valence effective core potentials [134–136] and the valence double- ζ (LACVP) and triple- ζ (LACV3P) contractions of the basis functions. Hydrogen, carbon, nitrogen, oxygen, phosphorus and sulfur were described with a modified variant of Pople's all-electron basis sets [137, 138] compatible with the LACV pseudopotentials. Polarization and diffuse functions were also included in the basis sets.

The long-range electrostatic influence of the protein environment was implicitly included through the integral equation formalism polarizable continuum model (IEF-PCM [139, 140]) and the Poisson–Boltzmann self-consistent reaction field model (PB-SCRf [141, 142]). Both models use sets of interlocking spheres to

define a cavity for the solute molecule within the continuum dielectric medium. However, sphere radii and polarization charges on the cavity surface are defined differently. The IEF-PCM approach computes point charges by a procedure of fitting the electrostatic potential and solving the resulting integral equations by means of the Green function method. There are several choices available to define the solute cavity. The sphere radii used in our IEF-PCM calculations correspond to the “united atom topological model” applied on atomic radii of the UFF force field [139, 140]. In contrast, the PB-SCRF approach computes point charges by fitting the electrostatic potential according to a least-squares criterion. The sphere radii are chosen from the CHARMM and OPLS force fields and some radii are allowed to be self-consistently optimized in the process of energy minimization [141, 142]. As previously discussed, the microscopic dielectric constant that can be assigned to a protein environment depends on the site of the protein considered, and can fluctuate between 2 and 10. An average value of 4 is usually assigned for SCRF calculations.

Chemical models of the active site were extracted from relevant crystallographic structures and a minimum number of appropriate atomic centers were anchored to their original positions within the protein’s frame of reference. The typical amino acid residue replacement is as follows: lysine by methylamine, arginine by a guanidinium group, histidine by an imidazole ring, aspartate and glutamate by either formate or acetate, asparagine and glutamine by either formamide or acetamide, serine and threonine by methanol, cysteine by methyl thiol, methionine by dimethyl sulfide, tyrosine by phenol, and tryptophan by indole. In the cases in which a backbone peptide bond is part of the chemically active region, it can be replaced by formaldehyde or formamide. Explicit water molecules are often included in the active site, especially if they are present in the crystal structure. Moreover, the protonation states of catalytic and substrate groups are set according to the experimental data available for each particular enzyme.

The reaction PESs were computed by geometry optimization of the chemical models in the gas phase or in the presence of the SCRF, always using double- ζ quality basis sets that include polarization functions. Gas-phase harmonic frequency analyses were performed to assess the character of the stationary point and to obtain zero-point vibrational energy corrections (ZPVEs). Finally, accurate total energies including solvent effects were computed on the stationary structures of reactant, transition, intermediate and product states using triple- ζ quality basis sets that include polarization functions and diffuse functions in certain cases.

Siegbahn has reported that for mechanistic studies the above approach exhibits several systematic errors. More specifically, errors in B3LYP total relative energies along a PES (i.e., in bond dissociation energies, barrier heights and energies of reaction) can amount to up to 12 kJ mol^{-1} for systems composed of first- and second-row atoms only, and up to 21 kJ mol^{-1} for systems including transition metals [122]. In addition, a shift of the entire PES by $\sim 5\text{--}10 \text{ kJ mol}^{-1}$ has been noted when formate rather than acetate is used to model the side chains of glutamate and aspartate [117].

23.3

Redox Catalytic Mechanisms

The catalytic mechanisms of two oxidase iron enzymes are discussed in this section. Nitric oxide synthase is a heme-containing enzyme that catalyzes the generation of nitrogen oxide, while AlkB is a non-heme enzyme that catalyzes the oxidative dealkylation of damaged nucleobases. Although both enzymes utilize iron in their redox mechanisms, the mechanisms are radically distinct owing to the effects of the protein environment.

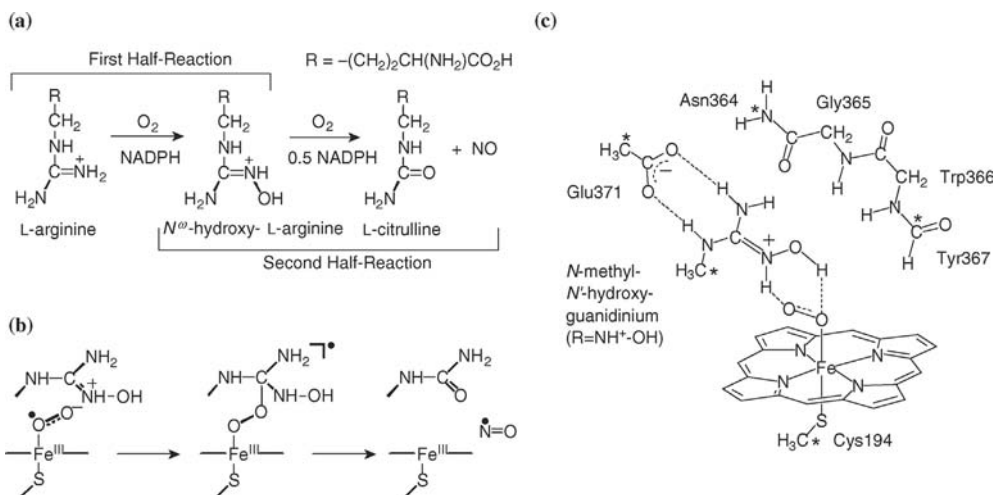
23.3.1

NO Formation in Nitric Oxide Synthase

Nitrogen oxide (NO) is a cell-signaling molecule that has three main effects: protective (e.g., antioxidant and leukocyte adhesion inhibitor), regulatory (e.g., neurotransmission and vasodilation) and deleterious (e.g., enzyme inhibitor and mediator in the generation of cytotoxic agents) [143]. NO is produced by nitric oxide synthases, a family of tetrameric enzymes composed of two monomers of nitric oxide synthase (NOS) and two monomers of calmodulin (CaM). An NOS monomer consists of a C-terminal reductase domain and an N-terminal oxygenase domain. When electrons donated by NADPH at a reductase domain active-site are transferred to the heme-iron centre at the oxygenase domain, NOS catalyzes the reaction of O₂ with L-arginine to generate citrulline and NO as products at the heme-iron centre. The enzyme-catalyzed reaction occurs in two oxidation cycles [144, 145]: in the first half-reaction, one O₂ molecule oxidizes L-arginine (R = NH) to give N_ω-hydroxy-L-arginine (R = NOH), while in the second half-reaction another O₂ molecule oxidizes the intermediate R = NOH to yield L-citrulline (R = O) and NO (Scheme 23.1a).

Owing to some similarities, the first half-reaction is thought to follow a mechanism analogous to that of the P450 monooxygenases [144, 145]. In contrast, there is no known analogous mechanism for the second half-reaction. As a result, it has attracted considerable attention, which has led to several mechanistic proposals [144–147]. While each of these is different, they generally share several common features. In particular, each proposal suggests: (i) that the reaction begins with binding of O₂ subsequently or concertedly with addition of one electron from the reductase domain and (ii) that the resulting ferric-superoxido species [Fe^{III}–(η¹-O₂^{•–})] attacks N_ω-hydroxy-L-arginine, or a derivative, at its guanidino carbon (C_{Gdn}) center to form an Fe^{III}–O–O–C_{Gdn} crosslinked “tetrahedral” species, which then collapses to give citrulline and NO (Scheme 23.1b).

Our computational studies focused on the second half-reaction, using *N*-methyl-*N'*-hydroxyguanidine (R = N–OH) as a substrate model for N_ω-hydroxy-L-arginine (Scheme 23.1c) [148–150]. However, rather than investigate each and every proposed mechanism, we began by considering key questions of all proposed mechanisms: (i) the protonation state of the R = N–OH substrate [148, 149], (ii) the thermochemistry of possible first steps of the mechanism [149] and (iii) the most feasible, if any, “tetrahedral” species [149].



Scheme 23.1 Nitric oxide synthase: (a) overall reaction catalyzed by NOS; (b) commonly proposed intermediates in the catalytic cycle; (c) chemical model of the active site. Atoms with an asterisk were held fixed in geometry optimizations.

With respect to the first question, the protonation state of the substrate at binding, the computational results suggested that it was in fact protonated; that is, the substrate in fact had an $\text{R} = \text{NH}^+ - \text{OH}$ structure whose protons were both involved in hydrogen bonding to the oxygen atoms of the $[\text{Fe}^{\text{III}} - (\eta^1\text{-O}_2^{\bullet-})]$ moiety [148]. All possible initial steps were then considered. In particular, we examined the thermochemistry for all possible H^+ and $\text{H}^\bullet$ transfers from the N' -hydroxyguanidinium cation's $\text{R} = \text{NH}^+ - \text{OH}$ moiety to the $[\text{Fe}^{\text{III}} - (\eta^1\text{-O}_2^{\bullet-})]$ group. Remarkably, these findings suggested that the most thermochemically feasible first step was a “double” H^+ and $\text{H}^\bullet$ transfer from the substrate to the $[\text{Fe}^{\text{III}} - (\eta^1\text{-O}_2^{\bullet-})]$ group, resulting in formation of $[\text{Fe}^{\text{III}} - (\eta^1\text{-H}_2\text{O}_2)]$, the ferric-dihydroperoxo complex [148, 149]. Notably, this intermediate was reminiscent of peroxidases, the enzymes that utilize hydrogen peroxide to oxidize their substrates.

Similarly, the mechanistic feasibility of a range of proposed and possible “tetrahedral” intermediates was also examined by calculating their energies relative to the initial bound active-site complex. These results suggested that any $\text{Fe}^{\text{III}} - \text{O} - \text{O} - \text{C}_{\text{Gdn}}$ crosslinked species would lie too high in energy [149].

On the basis of these findings, we then investigated the complete catalytic mechanism (Figure 23.2), by systematically identifying the lowest energy pathway for each subsequent intermediate [150]. The processes of binding of the $\text{R} = \text{NH}^+ - \text{OH}$ substrate and activation of oxygen were found to proceed more favorably on the spin-singlet PES, albeit the spin-septet PES is only 7 kJ mol^{-1} higher in energy (Figure 23.2) [150]. As a result, the first step of the second half-reaction involves the concerted deprotonation of $\text{R} - \text{NH}^+ - \text{OH}$ by $[\text{Fe}^{\text{III}} - (\eta^1\text{-O}_2^{\bullet-})]$ to form the $[\text{Fe}^{\text{III}} - (\eta^1\text{-H}_2\text{O}_2)]$ species and the nitrosyl radical intermediate $\text{R} = \text{N} - \text{O}^\bullet$. A thermodynamically favorable water elimination step makes possible the formation

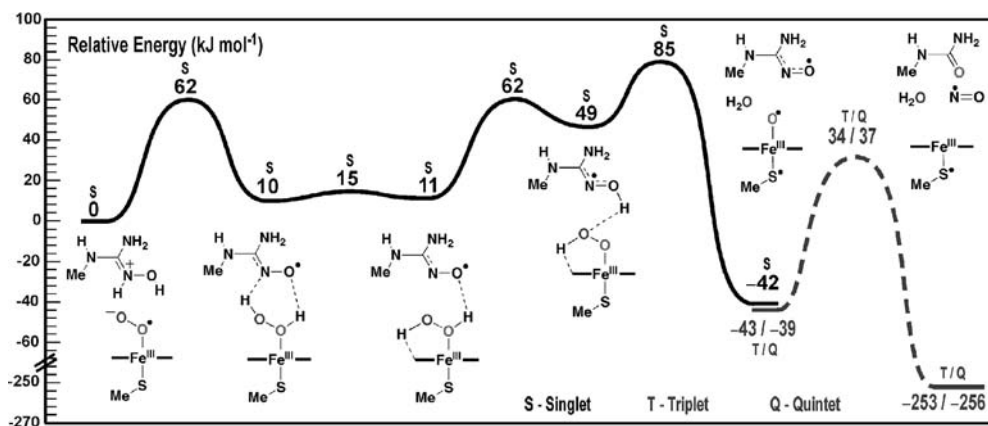


Figure 23.2 Potential energy surface for the oxidation of the *N*-methyl-*N'*-hydroxyguanidinium cation to *N*-methylurea and NO.

of the electrophilic ferric-oxyl species $\text{Fe}^{\text{III}}-\text{O}^\bullet$. At this point of advancement in the mechanism, spin crossing is likely to occur, since the electrophilic attack of the ferric-oxyl group to the guanidino carbon becomes far more favorable on the overall spin-triplet and -quintet PESs than on the spin-singlet surface. Finally, the $\text{Fe}^{\text{III}}-\text{O}^\bullet$ group attacks the guanidino carbon bringing about an oxygen atom transfer from the Fe(III) centre to the substrate's C_{Gdn} with concomitant release of NO [150].

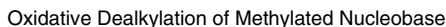
More generally, by using a systematic approach to elucidate several probable mechanistic pathways for the second half-reaction [150], we were able to identify possible alternate high-energy pathways that may help explain why experimental observations suggested that direct proton abstraction from the hydroxy group of the $\text{R}=\text{NH}^+-\text{OH}$ is not essential to the mechanism. Specifically, why production of NO still occurs, though at a slower rate, when modified substrates containing an $\text{R}=\text{NH}^+-\text{O}-\text{R}'$ ($\text{R}' = -\text{CH}_3$, $-\text{CH}_2-\text{CH}=\text{C}(\text{CH}_3)_2$ [147]) end are employed.

23.3.2

Oxidative Dealkylation in the AlkB Family

The AlkB family of enzymes (e.g., *Escherichia coli* AlkB and its human homologues ABH2 and ABH3) is able to repair a variety of both methylated and larger alkylated nucleobases [151–154]. These enzymes repair single-stranded DNA and RNA (*E. coli* AlkB [155]), double-stranded DNA (ABH2 [155]) and double-stranded DNA and RNA (ABH3 [155]) by a unique mechanism of oxidative dealkylation (Scheme 23.2) [156]. AlkB proteins belong to the α -ketoglutarate-Fe(II)-dependent dioxygenase superfamily and is the only family of enzymes with an oxidative dealkylation catalytic mechanism in this group [157, 158].

The chemical model used to investigate this mechanism (Figure 23.3) consisted of a seven amino acid side chains and the peptide bonds from Ser129 and Phe185.



Scheme 23.2 Proposed dealkylation mechanism catalyzed by AlkB with 1-methyladenine ($\text{Ade}^+ \text{CH}_3$) as the substrate.

The substrate 1-methyladenine was modeled as the 3-methyl-4-amino pyrimidinylation (3me4amPym⁺) and the α -ketoglutarate (α -KG) co-substrate was modeled as pyruvate. Dioxygen was bound to the Fe(II) ion trans to His131 along the reference axis. In the crystal structure, this position is occupied by a ligated water. The mechanism consists of four stages: (i) binding of O₂ and α -KG, (ii) oxygen activation, (iii) ferryl-oxo reorientation and (iv) oxidative dealkylation of the methylated nucleobase. The PESs of these stages were explored for the three overall



Figure 23.3 Structural models of the active site: (a) arrangement of the catalytically active residues of AlkB from the crystal structure (PDB accession code: 2FD8); (b) catalytic groups

replacing the residues and 3me4amPym⁺
replacing 1meAde⁺ in the computational
model. Atoms in red and with an asterisk were
held fixed in geometry optimizations.

possible spin-states of the system: triplet, quintet and septet. The electron flow was estimated through the calculated spin densities and Mulliken charges on the relevant atomic centers [159].

For the first stage, it was found that binding of dioxygen to iron in the active-site complex occurs concerted with electron transfer. Complexes in the overall sextet and quintet spin-states are thermodynamically stable in the ferric-superoxido [$\text{Fe}^{\text{III}}(\eta^1\text{-O}_2^{\bullet-})$] form regardless of the binding mode of dioxygen, because $\text{Fe}(\text{II})$ readily reduces $^3\text{O}_2$ to yield the highly reactive $\text{O}_2^{\bullet-}$. In contrast, the overall triplet and singlet spin-states are stable in the ferrous-dioxygen [$\text{Fe}^{\text{II}}(\eta^1\text{-O}_2^2)$] form, and charge transfer from the $\text{Fe}(\text{II})$ to $^3\text{O}_2$ moiety does not take place. In addition, no η^2 -coordinated species were found without charge transfer while for all η^1 -coordinated structures the bent binding mode is always preferred over any linear arrangement.

For the second stage of the reaction (Figure 23.4a), the activation of oxygen had slightly differing mechanisms (not fully shown here, see Reference [159]). It was found that the activation of oxygen through formation of a spin-quintet ferryl-oxo $\text{Fe}^{\text{IV}}=\text{O}$ intermediate is thermodynamically preferred over activation through formation of a spin-septet ferric-oxyl $\text{Fe}^{\text{III}}-\text{O}^{\bullet}$ intermediate, even though both pathways are exoergic. In fact, it was also found that the initiation of oxygen activation can be mechanistically competing on the septet and quintet PESs. This is in contrast to the preference for the spin-quintet activation pathway previously found in computational studies on related enzymes [160–163]. This is probably due to the inclusion of polarization functions on the heavy atoms in the present study compared to their non-inclusion in previous studies. Notably, by using a fragment molecular orbital approach, we were able to gain insights into how electrons are transferred from the co-substrate, which undergoes C–C cleavage in this step, to the O–O fragment.

Reorientation of the activated oxygen ferryl-oxo ($\text{Fe}^{\text{IV}}=\text{O}$; spin-quintet and -triplet pathways) and ferric-oxyl ($\text{Fe}^{\text{III}}-\text{O}^{\bullet}$, spin-septet pathway) species is necessary for the oxidative dealkylation to proceed (Figure 23.4b). This third stage was found to be nearly thermoneutral on all three spin-state PESs, with the lowest barrier ($\sim 50 \text{ kJ mol}^{-1}$) being found for the quintet surface. For the spin-septet pathway it was found that during reorientation, the $\text{Fe}^{\text{III}}-\text{O}^{\bullet}$ group accepts a proton from the nearby Arg210 active site residue to form $\text{Fe}^{\text{III}}-\text{OH}$. As a result, it is rendered catalytically inactive.

Finally, the rate-controlling step for both the spin-triplet and -quintet pathways was located at the oxidative dealkylation stage (Figure 23.4c). Specifically, it is the abstraction of the axial hydrogen atom from the methyl group of the alkylated nucleobase. The lowest barrier for this step (87.4 kJ mol^{-1}) occurs on the spin-quintet surface and is in good agreement with the experimental activation energy for the catalyzed reaction ($\sim 83 \text{ kJ mol}^{-1}$). The final undamaged nucleobase is then generated via a typical radical-rebound-type mechanism. Specifically, the $\text{Fe}^{\text{III}}-\text{OH}$ group resulting from the $\text{H}^{\bullet}$ abstraction transfers its $-\text{OH}$ to the substrate to form a $-\text{CH}_2\text{OH}$ group. The final step is cleavage of the $\text{N}-\text{CH}_2\text{OH}$ bond, which occurs in concert with deprotonation of the $-\text{CH}_2\text{OH}$ moiety to give H_2CO and the undamaged nucleobase.

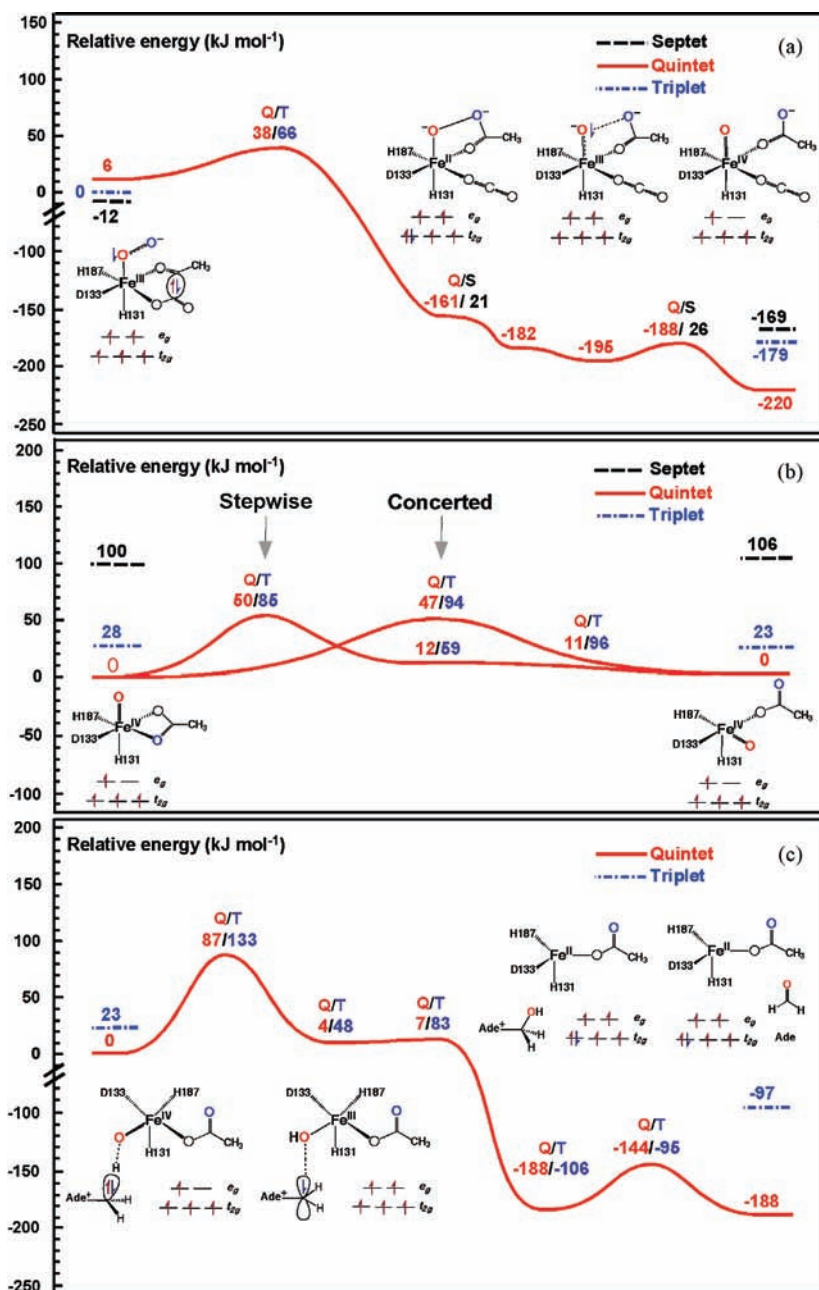


Figure 23.4 Spin-quintet potential energy surfaces for the repair mechanism of 1-methyladenine by AlkB: (a) activation of oxygen through the ferryl-oxo intermediate; (b) conformational reorientation of the ferryl-oxo group; (c) oxidative dealkylation of 1-methyladenine.

23.4

General Acid–Base Catalytic Mechanism of Deacetylation in LpxC

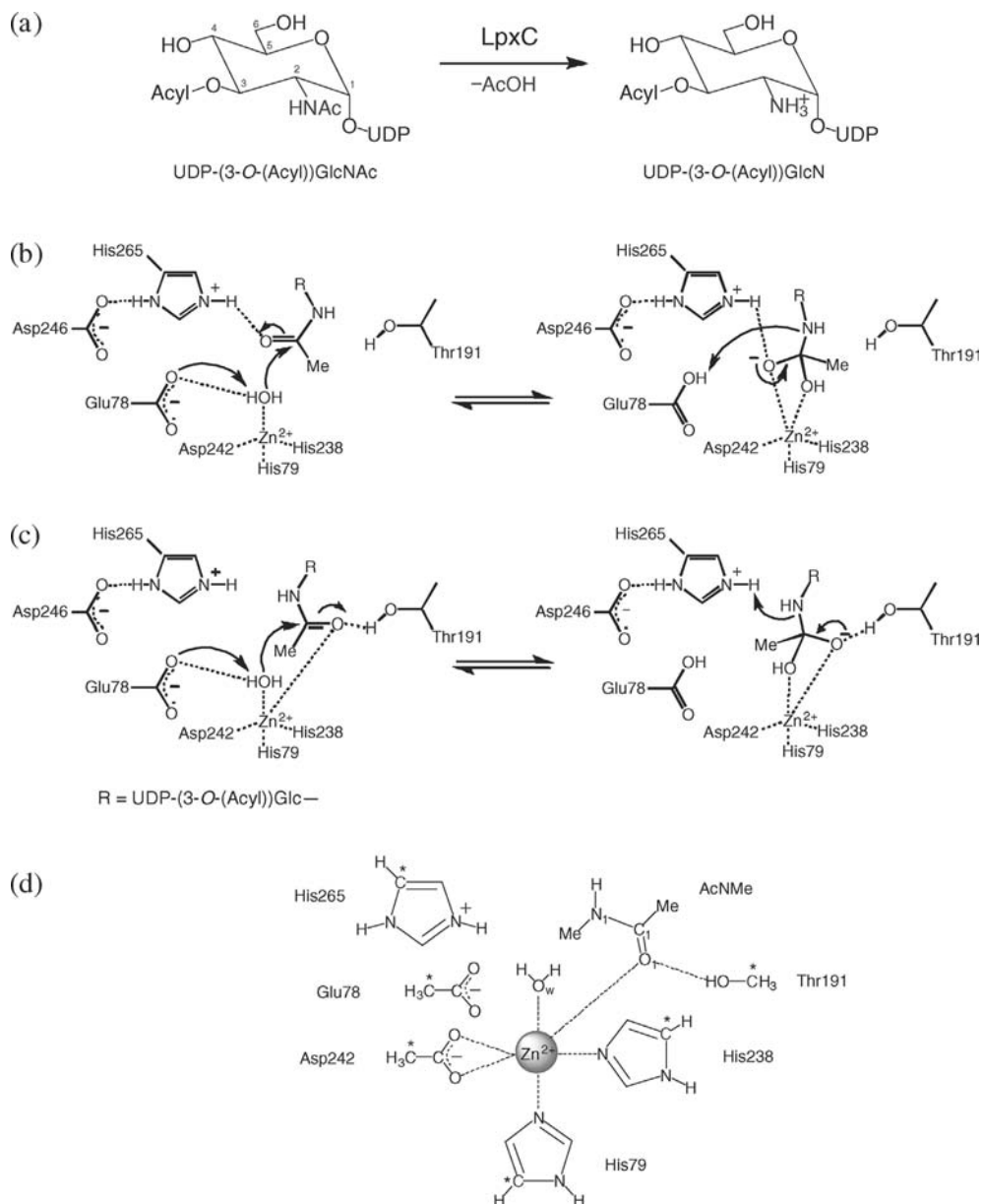
The outer membrane of most Gram-negative bacteria (e.g., *Chlamydia trachomatis*, *Escherichia coli*, *Hemophilus influenzae*, *Neisseria gonorrhoeae*, *Pseudomonas aeruginosa*, *Salmonella typhimurium* and *Vibrio cholerae*) contains antigenic lipopolysaccharides (LPSs) that are able to raise severe immune responses in humans. LPSs consist of three components that extend along the direction normal to the plane of the bacterium's outer membrane, from the extracellular space to the outer monolayer, in the order: (i) a repeating oligosaccharide known as the O-antigen, (ii) a non-repeating core polysaccharide and (iii) the lipid A or endotoxin [164, 165]. LPSs anchor to the outer monolayer of the outer membrane of most Gram-negative bacteria through Lipid A, the disaccharide GlcN4P(β 1 $\rightarrow$ 6)GlcN1P hexaacylated with N- and O-linked fatty acids that may contain saturated C10, C12, C14 and C16 chains. The core and O-antigen domains are highly variable in their monomer composition whereas Lipid A is rather conserved. As a result, the metabolic pathway of LPS synthesis is quite complex and variable. However, the second step in the biosynthesis of Lipid A has been identified to be highly conserved amongst Gram-negative bacteria and, thus, an excellent target for the development of more specific and potent antibiotics [164, 165]. More precisely, this step is the deacetylation reaction of the intermediate compound UDP-[3-O-(Acyl)]-GlcNAc by the metalloenzyme LpxC (Scheme 23.3a).

LpxC is a zinc hydrolase with a typical general acid–base mechanism for Zn(II)-dependent enzymes [166, 167]. However, on the basis of the experimental evidence, it was impossible to elucidate which residues acted as the general acid and base. One mechanistic proposal suggested that a single residue (Glu78) acted as both the general acid and base (Scheme 23.3b) [168, 169], whereas a second proposal suggested that a pair of residues (His265 and Glu78) acted as the general acid and base, respectively (Scheme 23.3c) [169].

Accordingly, we chose a chemical model of the active site (Scheme 23.3d) that included all residues directly coordinated to the Zn^{2+} center (Asp242, His79 and His238). Moreover, it also included those residues that have been proposed to directly hydrogen bond with the substrate's acetamide group (Thr191, Glu78 and His265). The latter Glu78 and His265 were also thought to be directly involved in the general acid–base catalysis according to both mechanistic proposals outlined above.

The PESs of both the reaction in aqueous solution and in the active site (Scheme 23.3c) were explored using *N*-methylacetamide (AcNMe) as the substrate [170]. The hydrolysis in solution was modeled by using two explicit water molecules with the system embedded in a continuum dielectric medium with $\epsilon = 80$. For the hydrolysis in solution, concerted and stepwise mechanistic pathways were found with barriers of 175 and 172 kJ mol⁻¹, respectively.

The first step in the mechanism is the binding of AcNMe to the active site. In this step, the carbonyl groups ligate to the Zn(II) center in concert with proton transfer from the Zn-bound water to Glu78. Some parallels between the hydrolysis in solution



Scheme 23.3 LpxC deacetylase hydrolysis of the amide bond: (a) overall reaction catalyzed by LpxC; (b) single-residue acid–base catalyzed mechanistic proposal (Glu78); (c) double-residue acid–base catalyzed mechanistic

proposal (protonated His265 and Glu78); (d) chemical model of the active site—atoms with an asterisk were held fixed in geometry optimizations.

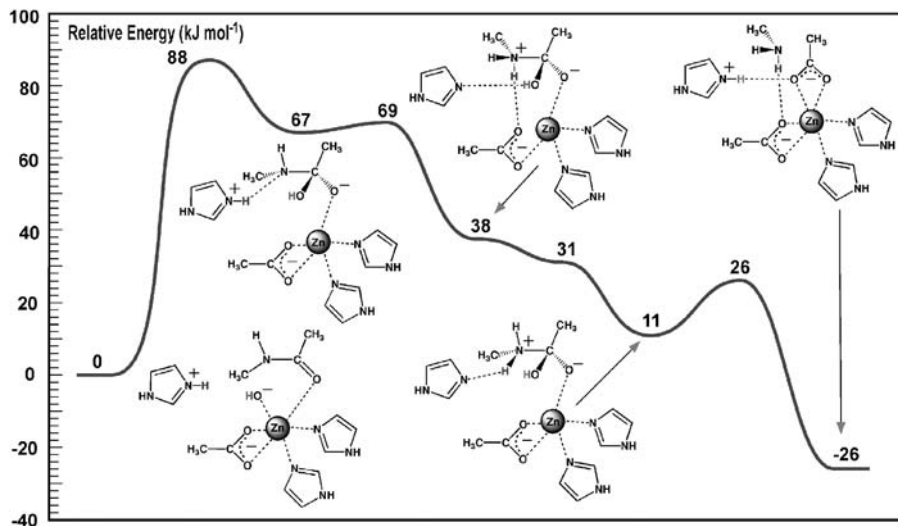


Figure 23.5 Potential energy surface for the deacylation of *N*-methylacetamide in the active site of LpxC.

and in the enzyme were observed. First, the rate-determining step in the stepwise hydrolysis in both water and enzyme is the initial nucleophilic attack of the hydroxy group to the substrate's carbonyl carbon. However, the enzymatic electrostatic environment was able to lower the barrier of this reaction to 88 kJ mol^{-1} , almost half that of the reaction in solution (Figure 23.5). This can be explained by the increased nucleophilicity of OH^- coordinated to the Zn^{2+} centre, compared to OH^- hydrogen bonding water. It was also found that the initial nucleophilic attack occurred in concert with protonation of the amide nitrogen by an active site histidine, forming a rather stable tetrahedral intermediate. This is then followed by a reorganization of the interactions involving the substrate and active-site residues. Specifically, the His265–substrate hydrogen bond switches from the substrate's protonated nitrogen center to its vicinal hydroxy group as the geminal oxygen atoms ligate to the Zn^{2+} ion. Cleavage of the substrate's C–N bond then occurs with a concerted proton transfer from the hydroxy group to His265, thereby forming methylamine and acetate.

23.5

Summary

This chapter presents updated energetic estimates of some effects thought to be responsible for the origin, efficiency and rate-accelerating power of enzyme catalysis. It seems quite clear that none of these effects alone can qualitatively and quantitatively account for the catalytic power of a single enzyme. Hence, computer modeling and simulation have been shown to be very effective at dissecting all the components that determine the catalytic performance of enzymes. Nonetheless, the

challenge for theoretical chemistry continues to be the devising of efficient theoretical methods that reach chemical accuracy ($<5 \text{ kJ mol}^{-1}$) for real scale systems treated at the atomistic level.

Notably, every computational model has its limitations, with its accuracy relying on the size of the chemical model and the physical approximations applied to describe the chemical interactions. Nonetheless, a thoughtful choice of a model molecular system, in conjunction with highly accurate quantum chemical methods that properly describe the chemical interactions within the system, can give tremendous insight into mechanisms of catalysis. In particular, the case studies examined herein (NOS, AlkB and LpxC) show that the use of large enough active-site chemical models to study enzymatic catalysis can give qualitative and quantitative details of the catalytic mechanisms developed by these enzymes. More specifically, the interactions that govern the catalytic mechanisms of these enzymes are transition state stabilization, substrate destabilization, electrostatics and hydrogen bonding networks. The careful choice of anchor points allows one to keep the model compounds in appropriate positions with respect to the protein's frame of reference. At the same time, the active-site model is flexible enough to account for short-range dynamic behavior such as rearrangements of hydrogen bond networks and changes in the coordination patterns around a metal center (e.g., LpxC and AlkB). It is also worth emphasizing that the improvement in rate and specificity of a reaction when catalyzed by an enzyme must be measured with respect to the same reaction proceeding in aqueous solution. Accordingly, for instance, our active-site model calculations shed light on how LpxC modifies the pathway of hydrolysis of the amide bond in solution, and is able to halve the barrier of the rate-limiting step.

Furthermore, the catalytic mechanisms of redox enzymes tend to occur on several spin-state surfaces (e.g., NOS and AlkB). Although the chemical models are large enough to accurately describe the most likely catalytic mechanisms, they are still of a manageable size that enables us to see detailed electronic insights such as available through mapping of charge and spin densities along the course of a pathway. In particular, a full molecular orbital or valence bond description is not thorough enough to tackle the complex mechanisms of reactions consisting of steps in which both full ionic and full covalent interactions are simultaneously present in the chemically active system. Accordingly, the fragment orbital approach allows one to track the electron flow over the course of a pathway on each spin-state PES. This in turn provides insights that allow us to explain mechanistic differences between different spin-states for a given enzyme, as, for instance, in NOS and AlkB. Thus, the use of large active-site models in combination with quantum mechanical methods is an effective and valid approach for systematically exploring, with quite high accuracy, various enzymatic mechanisms.

Acknowledgments

We thank the Natural Sciences and Engineering Research Council of Canada (NSERC), Canadian Foundation for Innovation (CFI), the Ontario Innovation Trust

(OIT), the Ontario Graduate Scholarship (OGS) program and SHARCNET for financial support. We also acknowledge SHARCNET for additional computational resources. Finally, J.W.G. thanks his current and past group members who have made all of this research possible.

References

- Pauling, L. (1946) *Chem. & Eng. News*, **24**, 1375–1377.
- Pauling, L. (1948) *Nature*, **161**, 707–709.
- Bruice, T.C. (1976) *Annu. Rev. Biochem.*, **45**, 331–373.
- Jencks, W.P. (1986) *Catalysis in Chemistry and Enzymology*, Dover Publication, New York.
- Warshel, A. (1998) *J. Biol. Chem.*, **273**, 27035–27038.
- Villà, J., Strajbl, M., Glennon, T.M., Sham, Y.Y., Chu, Z.T., and Warshel, A. (2000) *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 11899–11904.
- Villà, J. and Warshel, A. (2001) *J. Phys. Chem. B*, **105**, 7887–7907.
- Crosby, J., Stone, R., and Lienhard, G.E. (1970) *J. Am. Chem. Soc.*, **92**, 2891–2900.
- Warshel, A., Åqvist, J., and Creighton, S. (1989) *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 5820–5824.
- Lee, J.K. and Houk, K.N. (1997) *Science*, **276**, 942–945.
- Shurki, A., Strajbl, M., Villà, J., and Warshel, A. (2002) *J. Am. Chem. Soc.*, **124**, 4097–4107.
- Devi-Kesavan, L.S. and Gao, J.L. (2003) *J. Am. Chem. Soc.*, **125**, 1532–1540.
- Page, M.I. and Jencks, W.P. (1971) *Proc. Natl. Acad. Sci. U.S.A.*, **68**, 1678–1683.
- Lau, E.Y. and Bruice, T.C. (1998) *J. Am. Chem. Soc.*, **120**, 12387–12394.
- Hur, S. and Bruice, T.C. (2003) *J. Am. Chem. Soc.*, **125**, 1472–1473.
- Kollman, P.A., Kuhn, B., and Perakyla, M. (2002) *J. Phys. Chem. B*, **106**, 1537–1542.
- Bruice, T.C. (2002) *Acc. Chem. Res.*, **35**, 139–148.
- Warshel, A. and Levitt, M. (1976) *J. Mol. Biol.*, **103**, 227–249.
- Warshel, A. (1978) *Proc. Natl. Acad. Sci. U.S.A.*, **75**, 5250–5254.
- Warshel, A. (1987) *Nature*, **330**, 15–16.
- Warshel, A. (1991) *Computer Modeling of Chemical Reactions in Enzymes and Solutions*, John Wiley & Sons Inc., New York.
- Warshel, A. and Papazyan, A. (1998) *Curr. Opin. Struct. Biol.*, **8**, 211–217.
- Schutz, C.N. and Warshel, A. (2001) *Proteins*, **44**, 400–417.
- Warshel, A. (2002) *Acc. Chem. Res.*, **35**, 385–395.
- Warshel, A. (2003) *Annu. Rev. Biophys. Biomol. Struct.*, **32**, 425–443.
- Warshel, A., Sharma, P.K., Kato, M., and Parson, W.W. (2006) *Biochim. Biophys. Acta*, **1764**, 1647–1676.
- Warshel, A., Sharma, P.K., Kato, M., Xiang, Y., Liu, H.B., and Olsson, M.H.M. (2006) *Chem. Rev.*, **106**, 3210–3235.
- Hammes, G.G. (2002) *Biochemistry*, **41**, 8221–8228.
- Eisenmesser, E.Z., Bosco, D.A., Akke, M., and Kern, D. (2002) *Science*, **295**, 1520–1523.
- Benkovic, S.J. and Hammes-Schiffer, S. (2003) *Science*, **301**, 1196–1202.
- Palmer, A.G. (2004) *Chem. Rev.*, **104**, 3623–3640.
- Eisenmesser, E.Z., Millet, O., Labeikovsky, W., Korzhnev, D.M., Wolf-Watz, M., Bosco, D.A., Skalicky, J.J., Kay, L.E., and Kern, D. (2005) *Nature*, **438**, 117–121.
- McElheny, D., Schnell, J.R., Lansing, J.C., Dyson, H.J., and Wright, P.E. (2005) *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 5032–5037.
- Smiley, R.D. and Hammes, G.G. (2006) *Chem. Rev.*, **106**, 3080–3094.
- Hammes-Schiffer, S. and Benkovic, S.J. (2006) *Annu. Rev. Biochem.*, **75**, 519–541.
- Liu, H.B. and Warshel, A. (2007) *Biochemistry*, **46**, 6011–6025.
- Henzler-Wildman, K.A., Thai, V., Lei, M., Ott, M., Wolf-Watz, M., Fenn, T.,

- Pozharski, E., Wilson, M.A., Petsko, G.A., Karplus, M., Hubner, C.G., and Kern, D. (2007) *Nature*, **450**, 838–844.
- 38 Watt, E.D., Shimada, H., Kovrigin, E.L., and Loria, J.P. (2007) *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 11981–11986.
- 39 Benkovic, S.J., Hammes, G.G., and Hammes-Schiffer, S. (2008) *Biochemistry*, **47**, 3317–3321.
- 40 Roca, M., Messer, B., Hilvert, D., and Warshel, A. (2008) *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 13877–13882.
- 41 Dafforn, A. and Koshland, D.E. (1971) *Proc. Natl. Acad. Sci. U.S.A.*, **68**, 2463–2467.
- 42 Bruice, T.C., Brown, A., and Harris, D.O. (1971) *Proc. Natl. Acad. Sci. U.S.A.*, **68**, 658–661.
- 43 Mesecar, A.D., Stoddard, B.L., and Koshland, D.E. (1997) *Science*, **277**, 202–206.
- 44 Scott, W.G. (2001) *J. Mol. Biol.*, **311**, 989–999.
- 45 Masgrau, L., Roujeinikova, A., Johannissen, L.O., Hothi, P., Basran, J., Ranaghan, K.E., Mulholland, A.J., Sutcliffe, M.J., Scrutton, N.S., and Leys, D. (2006) *Science*, **312**, 237–241.
- 46 Nagel, Z.D. and Klinman, J.P. (2006) *Chem. Rev.*, **106**, 3095–3118.
- 47 Garcia-Viloca, M., Gao, J., Karplus, M., and Truhlar, D.G. (2004) *Science*, **303**, 186–195.
- 48 Pu, J.Z., Gao, J.L., and Truhlar, D.G. (2006) *Chem. Rev.*, **106**, 3140–3169.
- 49 Olsson, M.H.M., Mavri, J., and Warshel, A. (2006) *Philos. Trans. R. Soc. London, Ser. B*, **361**, 1417–1432.
- 50 Hay, S. and Scrutton, N.S. (2008) *Photosynth. Res.*, **98**, 169–177.
- 51 Marcus, R.A. (2006) *J. Chem. Phys.*, **125**, 194504.
- 52 Marcus, R.A. (2006) *Philos. Trans. R. Soc. London, Ser. B*, **361**, 1445–1455.
- 53 Marcus, R.A. (2007) *J. Phys. Chem. B*, **111**, 6643–6654.
- 54 Heyes, D.J., Sakuma, M., de Visser, S.P., and Scrutton, N.S. (2009) *J. Biol. Chem.*, **284**, 3762–3767.
- 55 Olsson, M.H.M., Parson, W.W., and Warshel, A. (2006) *Chem. Rev.*, **106**, 1737–1756.
- 56 Frey, P.A., Whitt, S.A., and Tobin, J.B. (1994) *Science*, **264**, 1927–1930.
- 57 Warshel, A., Papazyan, A., and Kollman, P.A. (1995) *Science*, **269**, 102–104.
- 58 Tuckerman, M.E., Marx, D., Klein, M.L., and Parrinello, M. (1997) *Science*, **275**, 817–820.
- 59 Feierberg, I. and Åqvist, J. (2002) *Theor. Chem. Acc.*, **108**, 71–84.
- 60 Gerlt, J.A. and Gassman, P.G. (1993) *Biochemistry*, **32**, 11943–11952.
- 61 Gerlt, J.A. and Gassman, P.G. (1993) *J. Am. Chem. Soc.*, **115**, 11552–11568.
- 62 Yu, Y.B. (2003) *J. Phys. Chem. B*, **107**, 1721.
- 63 Borman, S. (2004) *Chem. & Eng. News*, **82**, 35–39.
- 64 Zhang, X.Y. and Houk, K.N. (2005) *Acc. Chem. Res.*, **38**, 379–385.
- 65 Braun-Sand, S., Olsson, M.H.M., and Warshel, A. (2005) *Adv. Phys. Org. Chem.*, **40**, 201–245.
- 66 Bearne, S.L. and Wolfenden, R. (1995) *J. Am. Chem. Soc.*, **117**, 9588–9589.
- 67 Radzicka, A. and Wolfenden, R. (1995) *Science*, **267**, 90–93.
- 68 Wolfenden, R., Lu, X.D., and Young, G. (1998) *J. Am. Chem. Soc.*, **120**, 6814–6815.
- 69 Wolfenden, R. and Snider, M.J. (2001) *Acc. Chem. Res.*, **34**, 938–945.
- 70 Miller, B.G. and Wolfenden, R. (2002) *Annu. Rev. Biochem.*, **71**, 847–885.
- 71 Callahan, B.P. and Wolfenden, R. (2003) *J. Am. Chem. Soc.*, **125**, 310–311.
- 72 Lad, C., Williams, N.H., and Wolfenden, R. (2003) *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 5607–5610.
- 73 Wolfenden, R. (2003) *Biophys. Chem.*, **105**, 559–572.
- 74 Sievers, A., Beringer, M., Rodnina, M.V., and Wolfenden, R. (2004) *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 7897–7901.
- 75 Snider, M.G., Temple, B.S., and Wolfenden, R. (2004) *J. Phys. Org. Chem.*, **17**, 586–591.
- 76 Callahan, B.P., Yuan, Y., and Wolfenden, R. (2005) *J. Am. Chem. Soc.*, **127**, 10828–10829.
- 77 Schroeder, G.K., Lad, C., Wyman, P., Williams, N.H., and Wolfenden, R. (2006) *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 4052–4055.
- 78 Wolfenden, R. (2006) *Chem. Rev.*, **106**, 3379–3396.
- 79 Schroeder, G.K. and Wolfenden, R. (2007) *Biochemistry*, **46**, 4037–4044.

- 80 Lewis, C.A. and Wolfenden, R. (2008) *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 17328–17333.
- 81 Wolfenden, R. and Yuan, Y. (2008) *J. Am. Chem. Soc.*, **130**, 7548–7549.
- 82 Nelson, D.L. and Cox, M.M. (2004) *Lehninger Principles of Biochemistry*, 4th edn, W.H. Freeman.
- 83 Houk, K.N., Leach, A.G., Kim, S.P., and Zhang, X.Y. (2003) *Angew. Chem. Int. Ed.*, **42**, 4872–4897.
- 84 Wolfenden, R., Snider, M., Ridgway, C., and Miller, B. (1999) *J. Am. Chem. Soc.*, **121**, 7419–7420.
- 85 Sievers, A., Beringer, M., Rodnina, M.V., and Wolfenden, R. (2004) *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 12397–12398.
- 86 Gao, J.L., Ma, S.H., Major, D.T., Nam, K., Pu, J.Z., and Truhlar, D.G. (2006) *Chem. Rev.*, **106**, 3188–3209.
- 87 Kuhn, B. and Kollman, P.A. (2000) *J. Am. Chem. Soc.*, **122**, 2586–2596.
- 88 Stanton, R.V., Perakyla, M., Bakowies, D., and Kollman, P.A. (1998) *J. Am. Chem. Soc.*, **120**, 3448–3457.
- 89 Marcus, R.A. and Sutin, N. (1985) *Biochim. Biophys. Acta*, **811**, 265–322.
- 90 Siddarth, P. and Marcus, R.A. (1993) *J. Phys. Chem.*, **97**, 6111–6114.
- 91 Moser, C.C., Page, C.C., Farid, R., and Dutton, P.L. (1995) *J. Bioenerg. Biomembr.*, **27**, 263–274.
- 92 Mulholland, A.J. (2005) *Drug Discov. Today*, **10**, 1393–1402.
- 93 Hu, H. and Yang, W.T. (2008) *Annu. Rev. Phys. Chem.*, **59**, 573–601.
- 94 Singh, U.C. and Kollman, P.A. (1986) *J. Comput. Chem.*, **7**, 718–730.
- 95 Field, M.J., Bash, P.A., and Karplus, M. (1990) *J. Comput. Chem.*, **11**, 700–733.
- 96 Maseras, F. and Morokuma, K. (1995) *J. Comput. Chem.*, **16**, 1170–1179.
- 97 Humbel, S., Sieber, S., and Morokuma, K. (1996) *J. Chem. Phys.*, **105**, 1959–1967.
- 98 Matsubara, T., Sieber, S., and Morokuma, K. (1996) *Int. J. Quantum Chem.*, **60**, 1101–1109.
- 99 Svensson, M., Humbel, S., Froese, R.D.J., Matsubara, T., Sieber, S., and Morokuma, K. (1996) *J. Phys. Chem.*, **100**, 19357–19363.
- 100 Svensson, M., Humbel, S., and Morokuma, K. (1996) *J. Chem. Phys.*, **105**, 3654–3661.
- 101 Dapprich, S., Komaromi, I., Byun, K.S., Morokuma, K., and Frisch, M.J. (1999) *J. Mol. Struct. (THEOCHEM)*, **461**, 1–21.
- 102 Vreven, T. and Morokuma, K. (2000) *J. Comput. Chem.*, **21**, 1419–1432.
- 103 Vreven, T., Mennucci, B., da Silva, C.O., Morokuma, K., and Tomasi, J. (2001) *J. Chem. Phys.*, **115**, 62–72.
- 104 Morokuma, K. (2002) *Philos. Trans. R. Soc. London, Ser. A*, **360**, 1149–1164.
- 105 Vreven, T., Morokuma, K., Farkas, O., Schlegel, H.B., and Frisch, M.J. (2003) *J. Comput. Chem.*, **24**, 760–769.
- 106 Torrent, M., Vreven, T., Musaev, D.G., Morokuma, K., Farkas, O., and Schlegel, H.B. (2002) *J. Am. Chem. Soc.*, **124**, 192–193.
- 107 Lundberg, M. and Morokuma, K. (2007) *J. Phys. Chem. B*, **111**, 9380–9389.
- 108 Åqvist, J. and Warshel, A. (1993) *Chem. Rev.*, **93**, 2523–2544.
- 109 Rosta, E., Klahn, M., and Warshel, A. (2006) *J. Phys. Chem. B*, **110**, 2934–2941.
- 110 Stanton, R.V., Peräkylä, M., Bakowies, D., and Kollman, P.A. (1998) *J. Am. Chem. Soc.*, **120**, 3448–3457.
- 111 Laio, A., Van de Vondel, J., and Rothlisberger, U. (2002) *J. Phys. Chem. B*, **106**, 7300–7307.
- 112 Carloni, P., Rothlisberger, U., and Parrinello, M. (2002) *Acc. Chem. Res.*, **35**, 455–464.
- 113 Rohrig, U.F., Guidoni, L., and Rothlisberger, U. (2005) *ChemPhysChem*, **6**, 1836–1847.
- 114 Siegbahn, P.E.M. and Blomberg, M.R.A. (1999) *Annu. Rev. Phys. Chem.*, **50**, 221–249.
- 115 Siegbahn, P.E.M. (2001) *J. Comput. Chem.*, **22**, 1634–1645.
- 116 Himmo, F. and Siegbahn, P.E.M. (2003) *Chem. Rev.*, **103**, 2421–2456.
- 117 Siegbahn, P.E.M. (2003) *Q. Rev. Biophys.*, **36**, 91–145.
- 118 Noodleman, L., Lovell, T., Han, W.G., Li, J., and Himmo, F. (2004) *Chem. Rev.*, **104**, 459–508.
- 119 Siegbahn, P.E.M. and Blomberg, M.R.A. (2005) *Philos. Trans. R. Soc. London, Ser. A*, **363**, 847–860.
- 120 Siegbahn, P.E.M. and Borowski, T. (2006) *Acc. Chem. Res.*, **39**, 729–738.
- 121 Blomberg, M.R.A. and Siegbahn, P.E.M. (2006) *Biochim. Biophys. Acta*, **1757**, 969–980.

- 122 Siegbahn, P.E.M. (2006) *J. Biol. Inorg. Chem.*, **11**, 695–701.
- 123 Bassan, A., Blomberg, M.R.A., Borowski, T., and Siegbahn, P.E.M. (2006) *J. Inorg. Biochem.*, **100**, 727–743.
- 124 Siegbahn, P.E.M., Tye, J.W., and Hall, M.B. (2007) *Chem. Rev.*, **107**, 4414–4435.
- 125 Lundberg, M., Blomberg, M.R.A., and Siegbahn, P.E.M. (2004) *Top. Curr. Chem.*, **238**, 79–112.
- 126 Gaudl, J.W. and Eriksson, L.A. (2000) *J. Am. Chem. Soc.*, **122**, 2035–2040.
- 127 Rankin, K.N., Gaudl, J.W., and Boyd, R.J. (2001) *J. Am. Chem. Soc.*, **123**, 2047–2052.
- 128 Robinet, J.J., Baciú, C., Cho, K.B., and Gaudl, J.W. (2007) *J. Phys. Chem. A*, **111**, 1981–1989.
- 129 Becke, A.D. (1993) *J. Chem. Phys.*, **98**, 5648–5652.
- 130 Becke, A.D. (1993) *J. Chem. Phys.*, **98**, 1372–1377.
- 131 Lee, C.T., Yang, W.T., and Parr, R.G. (1988) *Phys. Rev. B*, **37**, 785–789.
- 132 Frisch, M.J. *et al.* (2004) Gaussian 03, Gaussian, Inc., Wallingford CT.
- 133 Schrödinger, L.L.C. (1991–2003) Jaguar 5.5, Portland, OR.
- 134 Melius, C.F. and Goddard, W.A. (1974) *Phys. Rev. A*, **10**, 1528–1540.
- 135 Melius, C.F., Olafson, B.D., and Goddard, W.A. (1974) *Chem. Phys. Lett.*, **28**, 457–462.
- 136 Hay, P.J. and Wadt, W.R. (1985) *J. Chem. Phys.*, **82**, 270–283.
- 137 Harihara, P.C. and Pople, J.A. (1972) *Chem. Phys. Lett.*, **16**, 217–219.
- 138 Francl, M.M., Pietro, W.J., Hehre, W.J., Binkley, J.S., Gordon, M.S., Defrees, D.J., and Pople, J.A. (1982) *J. Chem. Phys.*, **77**, 3654–3665.
- 139 Cancès, E., Mennucci, B., and Tomasi, J. (1997) *J. Chem. Phys.*, **107**, 3032–3041.
- 140 Tomasi, J., Mennucci, B., and Cammi, R. (2005) *Chem. Rev.*, **105**, 2999–3093.
- 141 Tannor, D.J., Marten, B., Murphy, R., Friesner, R.A., Sitkoff, D., Nicholls, A., Ringnalda, M., Goddard, W.A., and Honig, B. (1994) *J. Am. Chem. Soc.*, **116**, 11875–11882.
- 142 Marten, B., Kim, K., Cortis, C., Friesner, R.A., Murphy, R.B., Ringnalda, M.N., Sitkoff, D., and Honig, B. (1996) *J. Phys. Chem.*, **100**, 11775–11788.
- 143 Wink, D.A. and Mitchell, J.B. (1998) *Free Radical Biol. Med.*, **25**, 434–456.
- 144 Griffith, O.W. and Stuehr, D.J. (1995) *Annu. Rev. Physiol.*, **57**, 707–736.
- 145 Alderton, W.K., Cooper, C.E., and Knowles, R.G. (2001) *Biochem. J.*, **357**, 593–615.
- 146 Rosen, G.M., Tsai, P., and Pou, S. (2002) *Chem. Rev.*, **102**, 1191–1199.
- 147 Huang, H., Hah, J.M., and Silverman, R.B. (2001) *J. Am. Chem. Soc.*, **123**, 2674–2676.
- 148 Cho, K.B. and Gaudl, J.W. (2004) *J. Am. Chem. Soc.*, **126**, 10267–10270.
- 149 Cho, K.B. and Gaudl, J.W. (2005) *J. Phys. Chem. B*, **109**, 23706–23714.
- 150 Robinet, J.J., Cho, K.B., and Gaudl, J.W. (2008) *J. Am. Chem. Soc.*, **130**, 3328–3334.
- 151 Duncan, T., Trewick, S.C., Koivisto, P., Bates, P.A., Lindahl, T., and Sedgwick, B. (2002) *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 16660–16665.
- 152 Koivisto, P., Duncan, T., Lindahl, T., and Sedgwick, B. (2003) *J. Biol. Chem.*, **278**, 44348–44354.
- 153 Mishina, Y., Yang, C.G., and He, C. (2005) *J. Am. Chem. Soc.*, **127**, 14594–14595.
- 154 Delaney, J.C., Smeester, L., Wong, C.Y., Frick, L.E., Taghizadeh, K., Wishnok, J.S., Drennan, C.L., Samson, L.D., and Essigmann, J.M. (2005) *Nat. Struct. Mol. Biol.*, **12**, 855–860.
- 155 Yang, C.-G., Yi, C., Duguid, E.M., Sullivan, C.T., Jian, X., Rice, P.A., and He, C. (2008) *Nature*, **452**, 961–965.
- 156 Yu, B., Edstrom, W.C., Benach, J., Hamuro, Y., Weber, P.C., Gibney, B.R., and Hunt, J.F. (2006) *Nature*, **439**, 879–884.
- 157 Trewick, S.C., Henshaw, T.F., Hausinger, R.P., Lindahl, T., and Sedgwick, B. (2002) *Nature*, **419**, 174–178.
- 158 Falnes, P.Ø., Johansen, R.F., and Seeberg, E. (2002) *Nature*, **419**, 178–182.
- 159 Liu, H., Llano, J., and Gaudl, J.W. (2009) *J. Phys. Chem. B*, **113**, 4887–4898.
- 160 Borowski, T., Bassan, A., and Siegbahn, P.E.M. (2004) *Chem.–Eur. J.*, **10**, 1031–1041.
- 161 Borowski, T., Bassan, A., and Siegbahn, P.E.M. (2004) *Biochemistry*, **43**, 12331–12342.
- 162 Bassan, A., Blomberg, M.R.A., and Siegbahn, P.E.M. (2003) *Chem.–Eur. J.*, **9**, 106–115.

- 163 Lundberg, M., Siegbahn, P.E.M., and Morokuma, K. (2008) *Biochemistry*, **47**, 1031–1042.
- 164 Raetz, C.R.H. and Whitfield, C. (2002) *Annu. Rev. Biochem.*, **71**, 635–700.
- 165 Raetz, C.R.H., Reynolds, C.M., Trent, M.S., and Bishop, R.E. (2007) *Annu. Rev. Biochem.*, **76**, 295–329.
- 166 Lipscomb, W.N. and Strater, N. (1996) *Chem. Rev.*, **96**, 2375–2433.
- 167 Hernick, M. and Fierke, C.A. (2005) *Arch. Biochem. Biophys.*, **433**, 71–84.
- 168 Whittington, D.A., Rusche, K.M., Shin, H., Fierke, C.A., and Christianson, D.W. (2003) *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 8146–8150.
- 169 Hernick, M., Gennadios, H.A., Whittington, D.A., Rusche, K.M., Christianson, D.W., and Fierke, C.A. (2005) *J. Biol. Chem.*, **280**, 16969–16978.
- 170 Robinet, J.J. and Gauld, J.W. (2008) *J. Phys. Chem. B*, **112**, 3462–3469.

Part Four

From Quantum Biochemistry to Quantum Pharmacology, Therapeutics, and Drug Design

24

Developing Quantum Topological Molecular Similarity (QTMS)

Paul L.A. Popelier

24.1

Introduction

In 2006, a review [1] appeared on target-related applications of first principles quantum chemical methods in drug design. In it, an impressive table of case studies proves how computational chemistry tools enable the characterization of structure, dynamics and energetics of drug–ligand interactions. These studies reaffirm the familiar paradigm of drug design that pharmacological activity of the ligand is ultimately due to the spatial arrangement and electronic nature of its atoms. It is an exciting prospect that whole enzymes can be followed in real time via molecular dynamics simulations or that structural details of crucial enzyme fragments can be gleaned at increasingly high levels of theory. In the beginning of its introduction the review fleetingly mentions quantitative structure–activity relationships (QSAR), including 3D QSAR, featuring descriptors that were mainly obtained at the semi-empirical level [2].

Classical (2D) QSAR originates in the work of Fujita and Hansch [3], who relentlessly drove this approach, and have accumulated [4, 5] thousands of QSARs over the last 45 years. Hansch's approach is a bold extension of the linear-free energy relations (LFERs) [6] that were discovered by Hammett [7] almost three decades before Hansch started. Hammett's work is a cornerstone of traditional physical organic chemistry. Hammett observed the truly remarkable fact that the pK_a of benzoic acid and phenylacetic acid in aqueous solution is linearly coupled and merely dependent on the substituent (and a proportionality constant fixed by the solvent and temperature). Such relationships turned out to be more universal, which is why Hammett's equation can be used to estimate the equilibrium constants with different ring substituents with known σ constants [8]. The fact that LFERs exist means that one can focus on intrinsic features of an isolated molecule, in the gas phase, and relate those features to a complex phenomenon, such as an acid–base equilibrium in a solvent. The latter is a high-level physical organic property whose variation can be traced back to structural features of the molecules in question. This is both a strength and a weakness of QSAR. On one hand, QSAR aspires to predict the most complex

(“high-level”) properties using only structural information encoded in descriptors of isolated molecules. If this works for the right reasons then the strength lies in the sufficiency of the information locked in the isolated molecules (or even parts thereof). The power of the QSAR is then, in principle, its capacity to eliminate the need for actual simulations and the explicit representation of the environment of the molecule. On the other hand, however, QSAR can end up as a futile fitting exercise without true predictive power. This is the weakness of QSAR. The hilarious example of the curiously high correlation between the number of brooding storks and the number of newborn babies in what was then West Germany between 1965 and 1980 serves as a stark reminder of QSARs potential pitfalls [9]. Clearly, confusing correlation and causality is a dangerous affair. There is a higher probability of accidental correlation the larger the number of independent variables absorbed in the model. Through a well-known and careful study [10], Topliss and Edwards demonstrated numerically how dramatic this effect can be. Indeed, the popular correlation coefficient r^2 can be rightly discredited due to its sensitivity to chance correlation. In summary, the scientific community is strongly polarized about the true merit of QSAR as an activity. Even industrial researchers are torn by the lure of QSARs power and simplicity, on one hand, and the ineffectuality of its fitting, on the other.

It is against this background that the narrative of quantum topological molecular similarity (QTMS) will be given. QTMS is a working model that uses molecular descriptors, once new to the realm of QSAR, to capture the so-called electronic effects governing a physicochemical property or a biological activity. The abbreviation QTMS contains the letters Q and T, which also occurs in QCT (quantum chemical topology, which has its roots in “quantum theory of atoms in molecules” [11, 12] and its extensions that are based on the central idea of a gradient vector field [13]). A gradient vector field is an infinite set of gradient paths, which are paths of steepest ascent through a three-dimensional function (in our case). Using integration techniques [14], they can be traced from a triplet of differential equations. Gradient paths constitute the hallmark of QCT; they are the central feature that defines the approach. They are traced by following the gradient vector in space, point by point. Ultimately, the gradient vector is all that is needed to uncover the complex patterns and shapes lurking in the function of interest. A non-exhaustive list of 3D functions of interest for which the topology has been studied includes: the electrostatic potential [15], the virial field [16], magnetically induced molecular current distributions [17], the electron localization function (ELF) [18] and the intracule density [19]. The energy partitioning studies of the Oviedo group, who developed the interacting quantum atoms (IQAs) approach [20], also resides under the QCT umbrella. A detailed justification for the name QCT was given in a footnote of Reference [21], in the Appendix of Reference [22] and in Section 2 of Reference [23]. Alternative methods of interpretative quantum chemistry [24–30] do not share the central concept of the gradient vector field. This crucial difference draws together the topological analysis of the various functions under the heading of QCT, which is thus distinct from non-QCT methods.

As gradient paths unfold they display a topological pattern that conveys information of the quantum system (i.e., molecule) at hand. For example, there are special

points in space, called critical points, where the gradient vanishes. It is commonplace to evaluate properties at these critical points and use them as a compact “quantum fingerprint” of the molecule. Before we make this more concrete we recount the story of QTMS from where it started.

24.2

Anchoring in Physical Organic Chemistry

In 1980, Carbó *et al.* wrote an influential paper [31] entitled “How similar is a molecule to another? An electron density measure of similarity between two molecular structures.” This article, which nearly 30 years later fetched more than 450 citations, is barely five pages long and has only eleven references. Its impact can be largely understood from the absence of any answer to its title question at the time of publication, combined with the general importance of this question. The authors could only think of the crude attempt by Amoore, made ten years earlier, in the context of the molecular basis of odor [32], a notoriously difficult area for the formulation of quantitative models.

The basic *ansatz* of their work was simple. Carbó *et al.* proposed a “general matching measure” ϵ_{AB} , which is defined in Equation 24.1:

$$\epsilon_{AB} = \int_V dV |q_A - q_B|^2 \quad (24.1)$$

where q_A and q_B are the electron densities of the respective molecules being compared. The geometries of the two molecules remain frozen in the calculation of the measure ϵ_{AB} . Hence, since molecules A and B are rigid bodies, ϵ_{AB} depends on six parameters, that is, three rotation angles and three translation vector components. These parameters are varied to give a minimum of ϵ_{AB} for each pair of compared molecules. The basic premise of this approach is that molecular properties are ultimately due to their electron density, regardless of their high complexity, as found in biological activity for example.

It is trivial to prove that finding a local minimum in ϵ_{AB} is equivalent to finding a local maximum in the volume integral:

$$\int_V dV q_A q_B$$

However, after some preliminary computation Carbó *et al.* introduced a new matching measure, denoted r_{AB} , which filters out the molecular sizes via normalization. The quantity r_{AB} resembles a correlation index already in use in pattern recognition. This brought the approach in line with Amoore’s, who had applied pattern recognition techniques, but on macroscopic molecular models. The novelty of the work of Carbó *et al.* was to incorporate *ab initio* electron densities (being it semiempirical) into molecular similarity analysis. This contribution was surely driven by the physicalization of biological sciences, which is also the driving force

for developing QTMS, as explained below. The paper presented a correlation matrix for the ten pheromone analogues that Amoore had measured [33] the activities of. Although the agreement was not as convincing as one could hope for, the feasibility of amalgamating *ab initio* electron densities with (elementary) biological pattern recognition was demonstrated. However, the computational cost necessary to align the two molecules and then to compute the field-based similarity prevented it becoming a popular method in the 1980s. This situation changed when Good *et al.* [34] described an approach based on the use of Gaussian approximations to the exact molecular electron density that substantially increased the computational speed. The citation profile of the paper of Carbó *et al.* reached its peak in 1998 (and after a period of decline seems to be enjoying a revival in the last few years).

Other than the problem of excessive CPU time requirements, which admittedly has been alleviated, there is the issue of multiple local maxima in r_{AB} . The superposition of molecular electron densities unfortunately leaves the question of whether the maximum found is the global one. Furthermore, there is the well-known problem of the dominance of electron density near the nuclei. These core densities show up as huge peaks against a background of valence electron densities that are typically orders of magnitude smaller. It is easy to imagine that small spatial realignments of these peaks lead to relatively large changes [35] in r_{AB} . This problem can be circumvented by using electrostatic potentials [36], valence electrons [37] or momentum densities [38], yielding overlap between molecules that is less sensitive to nuclear positions. Finally, r_{AB} is a single number that hides which molecular fragments contribute more to the overall similarity than others. The volume integral over all space at the heart of r_{AB} obliterates any local (i.e., spatially resolved) information that may be desirable to know.

The main question that triggered the development of QTMS is whether one needs the full electron density of a molecule to quantify its similarity with another molecule. We argue that one does *not*. A chapter in Dean's book [39] on molecular similarity in drug design proposed for the first time the idea of bond critical point (BCP) space. A BCP is one of two possible types of saddle point in ordinary 3D space. As a saddle point it is a stationary point, and thus the gradient of the electron density vanishes at its position. The hallmark of this saddle point is that the electron density is a minimum in one direction (connecting the nuclei) and a maximum in the two remaining orthogonal directions. Locating these BCPs is a fast and robust process [40]. Figure 24.1 demonstrates the appearance of BCPs in phenol.

At those points, evaluating functions such as the electron density, its Laplacian or a kinetic energy density is straightforward. A BCP can thus be characterized by a number of BCP properties, each property being a function evaluated at its 3D position. The basic idea of the BCP space is that of an abstract multidimensional (hyper)space in which each dimension represents a BCP property. Let us look at a simple example. Using the program MORPHY98 [41] we can also sample the three curvatures of the electron density at a given BCP. They are typically denoted by λ_1 , λ_2 and λ_3 and are actually the eigenvalues of the Hessian of the electron density. When interpreted as local curvatures, one eigenvalue must be positive and the two other

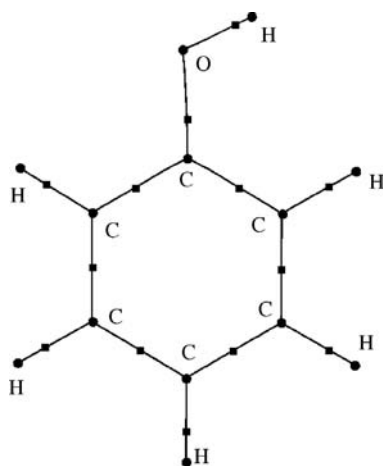


Figure 24.1 Full set of 13 BCPs (represented as filled squares) appearing between all bonded nuclei in phenol.

negative, to recover the type of saddle point that is a BCP. The ellipticity, denoted by ϵ_b , is defined as $\lambda_1/\lambda_2 - 1$. It can be regarded as a simple shape descriptor of electronic structure, for instance expressing the π character of a CC bond, when it has a bond order between one and two. Figure 24.2 shows a concrete example of a molecule

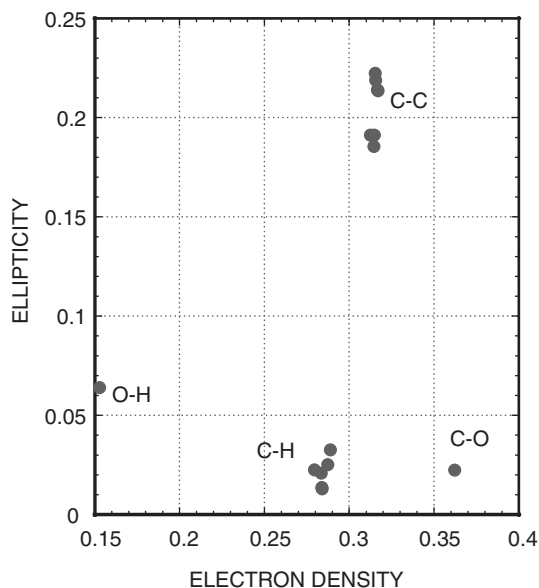


Figure 24.2 Representation of phenol in a two-dimensional BCP space. Ellipticity is plotted versus electron density evaluated at each BCP, in atomic units.

being represented in BCP space by a series of points. The thirteen bonds in phenol are characterized in a 2D BCP space, with one dimension of electron density and one of ellipticity.

As expected, the CC bonds all have a high ellipticity, with a fine structure of six non-coinciding points in BCP space. Perhaps not surprisingly either, the CO bond shows the highest electron density. The BCPs all appear in distinct regions of BCP space. Overall, similar types of bonds are in similar areas of BCP space, in which molecules appear as compact fingerprints. One can ask how unique a particular position in BCP space actually is. Can a given point in BCP space be associated with *only one bond* in one molecule (in one conformation)? The answer is most likely yes. There are many bonds in different molecules that have the same electron density. In fact, by altering the nuclear configuration the electron densities of different bonds can be continuously tuned to exactly match each other. Each time a substituent is added to phenol's phenyl ring the pattern of all of phenol's BCPs will change. Some points will be affected more than others. Figure 24.2 only shows a two-dimensional BCP space but one can use as many BCP properties as possible. Also, any combination of BCP properties could have been used in Figure 24.2. The drug haloperidol featured in Dean's book as the first example of a molecule represented in BCP space. The 51 bonds appeared in a 3D BCP space, with dimensions of electron density, its Laplacian and the ellipticity. Five BCP points, well separated as singletons representing unique bonds, appeared together with five tight clusters of BCP points. The $C_{\text{arom}}-C_{\text{arom}}$ cluster consisted of two subclusters, the smaller one representing the two pairs of phenyl CC bonds adjacent to the CF or CCl bond. These four bonds show a somewhat higher ellipticity, which is due to the fact that halogens are π donors. The ellipticity even detects that fluorine is a stronger π donor than chlorine.

This detailed but compact characterization encouraged the exploration of BCP space in case studies where the so-called Carbó (molecular similarity) index r_{AB} has been calculated. Lee and Smithline [42] computed the similarity indices of a small set of substituted benzenes, containing benzene, aniline, nitrobenzene, *p*-nitroaniline, *m*-nitroaniline and *o*-nitroaniline. The molecules were ranked according to their similarity to benzene, where aniline turned out to be the most similar, followed by nitrobenzene. An analysis based on distances in BCP space ranked the molecules differently. The fact that no experiment could arbitrate this disagreement spurred the application [43] of BCP space in a case where experiment could not. Attention turned to one of the oldest QSARs, namely Hammett's substituted benzoic acids. Hammett introduced a substituent constant σ as the logarithm of the ratio of the measured ionization constant of a substituted benzoic acid (substituent S) to that of benzoic acid itself in water at 25 °C, as summarized in Equation 24.2:

$$\sigma = \log \frac{K_S}{K_H} = pK_{a,H} - pK_{a,S} \quad (24.2)$$

If a similarity index in BCP space could somehow be related to σ then this would experimentally validate the concept of BCP space. A simple measure of similarity is

Euclidean distance. The distance d_{ij} between two BCPs i and j in 3D BCP space is defined as follows:

$$d_{ij} = [(\rho_i - \rho_j)^2 + (\nabla^2 \rho_i - \nabla^2 \rho_j)^2 + (\varepsilon_i - \varepsilon_j)^2]^{1/2} \quad (24.3)$$

where ρ stands for the electron density, $\nabla^2 \rho$ for its Laplacian and ε for ellipticity, at BCP i or j . The three components constituting this distance have different dimensions, which is why the “raw” distance defined in Equation 24.3 should be modified. In most accounts on clustering, for example, standardization of the variables to zero mean and unit variance is recommended, using the standard deviation from the complete set of entities [44]. This means that a variable x is replaced by $(x - \mu)/\sigma$, where x is any of the three BCP descriptors in Equation 24.3. The distance $d(A, B)$ between two molecules A and B is then defined as a sum of these BCP distances via Equation 24.4:

$$d(A, B) = \sum_{i \in A} \sum_{j \in B} d_{ij} \quad (24.4)$$

Again, this is not the only possibility but it is a natural one and one that worked, as shown below. The lower the value $d(A, B)$, the more similar the two molecules are. Equation 24.4 raises an important question: which BCPs of molecule A should be compared to which BCPs of molecule B? The set symbol $\in$ leaves that question open but answering it led to an important feature of the QTMS method, which is that it can highlight the “locus of action” in the molecule.

One possible answer is to compare each BCP in A with each BCP in B. This “total” distance is the most complete since it compares each bond with another, even if they have nothing in common chemically. For example, one could compare the distance in BCP space between water and methane, where a CH bond would be contrasted with an OH bond. Figure 24.2 suggests that such bonds occur in very different corners of BCP space. Although such a “total” distance provides a valid measure between two entirely different molecules, it turns out to be an inadequate distance to gauge the similarity between a set of congeneric molecules typical for QSAR. It is straightforward to include only the distances between two *corresponding* BCPs in molecule A and molecule B. The substituted benzoic acids have the phenyl ring and the carboxyl group in common. Of course there is the bias of a priori matching of BCPs (i.e., bonds) but it is a perfectly natural mode of operation. The para-substituted benzoic acids were looked at first.

There are some discrepancies in the experimental σ values depending on the cited source, but all figures agree that the para substituents should be ranked as follows: $\text{NH}_2 < \text{OCH}_3 < \text{CH}_3 < \text{H} < \text{F} < \text{Cl} < \text{CN} < \text{NO}_2$. All *ab initio* wavefunctions were obtained at the B3LYP/6-311 + G**//B3LYP-6-311 + G** level of theory using the program GAUSSIAN94. The topological analysis of the electron distribution was performed using MORPHY98. The distance between all benzoic acids was calculated using Equation 24.4, yielding a similarity matrix. A natural choice for the reference molecule is the first or last member of a sequence, such as the NH_2 or the

NO_2 substituent. Even if this substituent is not known a priori to be the edge of a sequence, it will emerge automatically from inspecting the distance matrix. The substituent sequences were generated, with respect to NH_2 , for several BCP subsets. In other words, one varies the decision which BCPs the set symbol $\in$ in Equation 24.4 selects. It turned out that the experimental substituent sequence is *only* reproduced if the similarity measure (i.e., distance in BCP space) is restricted to BCP contributions from the COOH group. Any inclusion of BCPs from the phenyl group seriously disrupted the sequence. Consequently, the budding QTMS method pointed out the “centre of action” for this well established QSAR. This study seems to have inspired the Carbó group to link their quantum molecular similarity (QMS) [45] method with QSAR. In their study [46] that appeared two years later than the QTMS study [43] currently discussed, substituted benzoic acids were investigated. In addition, the ability of QTMS to highlight a centre of action common to a class of congeneric molecules seems to have awakened this research group’s interest to achieve the same with their method. The partitioning technique they used to define the electron density of molecular fragments is the same as that behind the Mulliken population analysis. This technique simply assigns all basis functions that are centered on any nucleus in the fragment to that fragment. The assignment of the basis functions also assigns the electron density (and hence the derived properties) that the basis functions contribute to, to a given fragment. A drawback of this partitioning is that it evaporates if the electron density is described by functions that have no centers such as plane waves. A diffuse (Gaussian) function is still compatible with this Mulliken type partitioning but spoils it already, basically for the same reason: a tenuous connection between the basis function’s centre and the spatial localization of its contribution. The QMS study of Carbó and coworkers [46] concluded that the best QSAR models are indeed generated with the COOH fragment or any of its sub-fragments. However, the situation is slightly more complex because in addition to “expected” correlations with the QS-SM related to the COOH fragment, there were also a considerable number of QSAR models whose theoretical descriptors seem to be in no relation to this group. This observation could be related to the so-called “contaminations,” discussed below, that harass QTMSs uncovering of the “pure” centre of action. For completeness and relevance in the discussion further on in this chapter we mention now that the electron densities used in the study of Carbó and coworkers belonged to fully optimized molecular geometries, performed by the Hartree–Fock method using the 3-21G* basis set.

To further test QTMS, the wavefunctions of five more para-substituted benzoic acids were generated: COCH_3 , CHO, phenyl, OH and O^- . Until 1960 the experimental σ value for OH and O^- was unavailable because of experimental difficulties. QTMS correctly predicts that this substituent is bracketed by OCH_3 and CH_3 . The case of O^- is interesting because it offers for the first time the possibility of extrapolation rather than interpolation. Indeed, the experimental σ value puts it “left of NH_2 ,” outside the $[\text{NH}_2, \text{NO}_2]$ bracket. The correct experimental sequence from the point of view of the O^- substituent was recovered because the distance between O^- and NH_2 is the smallest of all distances and increases monotonically

through the sequence. This is how QTMS correctly deduces that O^- is even more electron donating than NH_2 .

In this early QTMS study the meta-substituted benzoic acids were looked at next. The substituents OH and OCH_3 have deliberately been included because they appear at the NO_2 end of the range instead of at the NH_2 end in the case of para-benzoic acids. This is important to prove that QTMS is reliable in predicting the different behavior of para compared to meta, as indeed it does. Again one retrieves the main result that the theoretical sequence generated by the centre of action $COOH$ perfectly matches the experimental sequence. Just as in the para case, the sequence based just on the $C=O$ BCP also makes the theoretical and experimental sequence match but the $O-H$ BCP on its own fails to do so. To finalize the analysis, a simple linear regression analysis was carried out on the eight original para-substituted benzoic acids against the proposed similarity distance $d(NH_2, S)$, where S is a substituent. The distance is computed with respect to the NH_2 substituent. The Pearson correlation coefficient for this particular fit was 0.993.

Despite this success several issues remained open at that stage. Five questions called for more work: What is the actual dimension of BCP space? Or, more precisely, which BCP properties contribute to the best possible reproduction of an experimental sequence? A related issue is whether the BCP properties are actually independent. A second question is how reliable BCP space really is on a practical level, in particular with regard to large systems, and if it works for less straightforward QSARs. The third issue is of a technical nature and concerns the level of theory (which had not been varied in the previous case study). The fourth question is whether the Euclidean distance (including the standardization of the variables) is the best similarity measure. The final question is how conformational changes appear in BCP space and to what extent they influence distances in BCP space.

The case study [43] concluded that quantum similarity measures had been proposed before under the hypothesis that molecular properties can ultimately be reduced to the electron distribution. It turned out that they were unnecessarily cumbersome and biased by chemically unimportant regions. Secondly, QTMS revealed that the experimental activity sequence will only be reproduced if the distance measure is confined to contributions from the BCPs from the common center of action of the molecules.

The use of BCP properties for molecular similarity work has some obvious advantages over the use of the full density. The process of obtaining the density from *ab initio* calculations is the same in both instances, but the subsequent analysis is quite different. The location of BCPs and the evaluation of their properties are computationally inexpensive. For most QSARs there is no problem associated with molecular alignment. Admittedly this is only so because molecules in a congeneric series have large fragments in common and hence a BCP in one molecule can easily be mapped onto another. Moreover, as BCPs occur at the minimum in the electron density broadly along the internuclear axis, the electron density in this region is not overshadowed by the core electron densities near the nuclei. They appear in the areas of the molecule at which chemical activity is more interesting. This is not to say that BCP properties are not affected by the core electron densities. Indeed, in

semiempirical calculations where the core electron densities are not explicitly modeled, BCPs do not appear. Their properties depend on the accuracy of the whole molecular electron density.

24.3

Equilibrium Bond Lengths: “Threat” or “Opportunity”?

So far we did not mention, let alone emphasize, that QTMS and BCP space operate on optimized molecular geometries. There is unpublished evidence that, without geometry optimization, the QSAR models collapse. The equilibrium bond lengths, denoted R_e , can be used as descriptors themselves in QTMS. Later, evidence was accumulated on various QSARs that bond lengths alone could generate models with respectable correlation coefficients. A natural question to then ask is:

Do the BCP properties such as q or a type of kinetic energy density $K(\mathbf{r})$ [47] actually add any information or value over and above bond lengths?

A subsequent QTMS study [48] set off to address this question by investigating possibly linear relationships between BCP properties and bond lengths. For that purpose a data set of 57 molecules was designed, based on single conformations of the 20 natural amino acids and smaller derived molecules, yielding 731 BCPs. The existence of *local* linear relationships was confirmed on the condition that the bonds varied little in their chemical surroundings. Such relationships break down completely for larger subsets of BCPs encompassing a wider variety of bonds. The patterns observed in the global picture showed so little correlation that one may safely conclude that BCP properties cannot be trivially recovered or even predicted by knowledge of bond length alone. Within the limits provided by the C–O and C–C data sets the kinetic energy densities evaluated at BCPs do appear to obey highly linear relationships with R_e . This study examined five different levels of theory and took into account the influence of basis set variation to BCP properties.

The question at the heart of this section cannot be answered by an unqualified yes or no, when asked at the stage of an actual QTMS application. The study described above clearly confirms the added value of BCP properties beyond mere bond lengths. The issue is that this value only surfaces when widely differing bonds are compared, and such comparisons are absent in a QSAR of congeneric molecules. In any event, it is useful to add the bond length to BCP space, even it is perhaps artificial to do so. If bond lengths alone were sufficient then there would be no need for BCP space, let alone QCT. Hence the word “threat” in the section title, where the quotes are of course meant to be ironic. Equally, if bond lengths alone are sufficient this is an “opportunity” to exploit this at a larger scale, since it simplifies the analysis and the QSAR. It is perhaps curious that the power of bond lengths as a QSAR descriptor was discovered in this roundabout manner, that is, via the topology of q . To the best of our knowledge it has not been discovered before, despite the availability of optimized bond lengths.

24.4

Introducing Chemometrics: Going Beyond r^2

In the next paper [49] QTMS was incorporated in a firm statistical framework currently employed in modern QSAR, making use of the partial least-squares (PLS) procedure [50–52]. PLS is a regression technique that makes use of quantities like principal components (PCs) [53] derived from the set of independent variables, or BCP descriptors in this case. The PCs in the PLS regression are called latent variables (LVs). The dependent variable (an activity or physical property of interest) is then expressed as a linear combination of the LVs, in the PLS equation. Each LV is a linear combination of the independent variables. The PLS algorithm is an iterative procedure that combines the step of PC analysis (PCA) with the regression step. The first LV will explain most of the variance in the independent set, the second LV the next largest amount and so on.

The early results on *para*-benzoic acids presented above were shown to survive the more rigorous statistical treatment. Furthermore, it was shown that the range of applications of QTMS extends to other carboxylic acid systems, such as *para*-phenylacetic acid, 4-X-bicyclo[2.2.2]octane-1-carboxylic acids and polysubstituted benzoic acids. The BCP descriptors consisted of Q , $\nabla^2 Q$, λ_1 , λ_2 , λ_3 , ϵ , K and G . These quantities were packed together into eight-dimensional vectors, each one serving as a chemical descriptor for a given bond. Thus the whole molecular set is described by $8n$ variables, where n is the number of bonds present in the common molecular skeleton. These variables (“X-variable,” independent) are regressed against the property of interest (“Y-variable,” dependent) using PLS. One could object to using λ_1 , λ_2 and λ_3 together with $\nabla^2 Q$ since the latter is the sum of the former. However, PLS is ideally suited for this form of multivariate analysis as it is designed to tackle data with many collinear variables. PLS also copes well with noisy variables. The quality of the PLS equation can be judged from the correlation coefficient (r^2) and the cross-validated correlation coefficient [53], denoted as q^2 . Randomization of the response variables is also carried out to assess the likelihood of the correlation occurring by chance. If good correlations are obtained with incorrect data, the initial regression may be due to chance factors alone. A full discussion follows below (Figure 24.4). Assessment of the QSARs validity is essential before any results can be interpreted in any physical sense.

PLS highlights those variables that can “explain” the property of interest. We do not have to choose where in the molecule to look or which properties to look with. Hence, a priori knowledge of a centre of action for any given activity is not necessary [54]. Below we discuss a case study demonstrating how QTMS uses PLS to highlight the centre of action.

To simplify the interpretation of the QSAR, we group all variables that describe the same region of space [55]. These eight descriptors are reduced to a single variable by PCA, which is carried out by the program SPSS [56]. The representation of molecules in BCP space is now performed by means of BCP descriptors forming the first PC, evaluated for each BCP. In the same way as before, PLS highlights the important variables for each activity or property examined. These variables can now

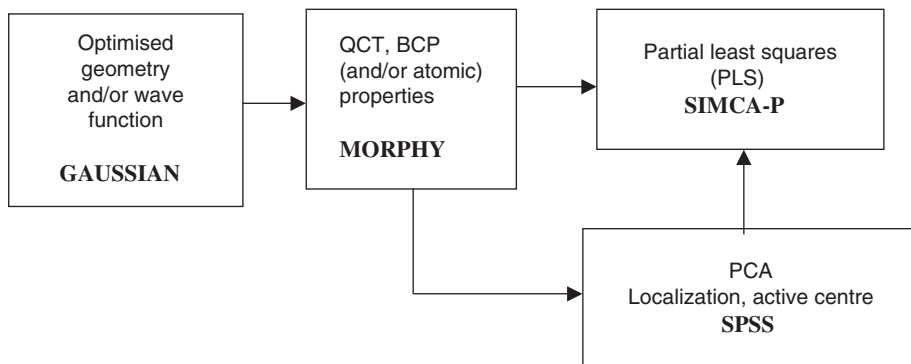


Figure 24.3 Chart representing the main computational modules involved in a QTMS analysis. Bold text represents the names of the programs used.

be associated directly with a region of electron density that is recognized as a bond. Figure 24.3 summarizes and illustrates this whole procedure, mentioning the software packages used.

Typically, all the ("raw") BCP descriptors available are used to carry out the regression for *prediction* purposes. A reduced set of variables arising from PCA of localized BCP properties should then be used to *interpret* more fully the PLS analysis. The variables identified in the PLS analysis can be regarded as intuitively reasonable and easily interpretable.

Validation of a QSAR is essential if there is to be any meaningful interpretation of its results. Figure 24.4 shows the values of r^2 and q^2 that arise if the same PLS analysis is carried out but after the response variables were randomized. In other words, each time the analysis is carried out (ten times in this case) each molecule is randomly assigned a value from the pool of Y variables. Any good correlations obtained would then be due to random factors or chance. One can be certain of the

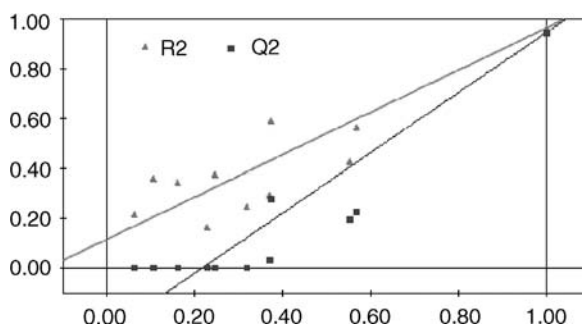


Figure 24.4 Plot of r^2 and q^2 obtained following randomization of the response variable (σ). The y-axis marks the value of the correlation coefficient (r^2 or q^2) and the x-axis marks the correlation of the original set of response variables with its permutation. Each point arises from a new permutation (ten in total). Results are for para-substituted benzoic acids.

validity of the PLS results if the values of the correlation coefficients recovered due to the permutation of Y variables are consistently below those obtained by the real analysis. It is clear from Figure 24.4 that any conclusion made from the QSAR for para-substituted benzoic acids is founded on genuine correlations with the properties of the electron density. Similar validations were also noted for the meta-substituted benzoic acids.

PLS analysis was performed by the SIMCA-P package [57]. All *ab initio* calculations were carried out at the following levels of theory: HF/3-21G(d)//HF/3-21G(d), HF/6-31G(d)//HF/6-31G(d), B3LYP/6-311 + G(2d,p)//HF/6-31G(d) and B3LYP/6-311 + G(2d,p)//B3LYP/6-311 + G(2d,p). As the trends in BCP properties are preserved between basis sets [48] and the results of the PLS analysis are essentially basis set independent we only report those carried out at the highest level of theory.

This is the place to discuss briefly how 1-1 mapping of BCPs between molecules in a less congeneric series can be achieved. The requirement of a maximally common skeleton, such as in *para*-benzoic acids, was relaxed in a further contribution [58]. Bonds between different molecules from a single congeneric set can be unambiguously mapped onto each other. This is because bonds are marked by their position in the molecule and by the elements they connect (e.g., C–Cl connects carbon and chlorine). If the latter requirement is relaxed then *para*- and *meta*-benzoic acids can be put in the same set. The fact that a C–H bond is then matching a C–X twice (once for C–X in the *para* position and once for C–X in the *meta* position) does not disrupt the QTMS analysis. Moreover, with a small modification, the *p*-phenylacetic acids can be added to this set [58]. This is also the first study where QCT atomic properties feature. Only a model with a single LV passed the randomization validation test, yielding a very high q^2 value.

Before we discuss an example of a QTMS analysis pointing at different centers of action (Section 24.5) depending on the Y variable fed to PLS, we briefly mention a simple QTMS application to carbon-13 NMR shifts, reported for a series of substituted benzonitrile compounds [59]. To explore whether BCP space can recover NMR data we formulate a QSAR with the chemical shift found for the carbon in $C \equiv N$. A set of eight *para* and *meta*-substituted compounds were geometry optimized only at the HF/3-21G(d) level. The cyano bond was recovered as the centre of action and the r^2 coefficient turned out to be 0.99 with q^2 still at a respectable 0.88.

24.5

A Hopping Center of Action

A short series of *para*-substituted phenols has been studied by Damborsky and Schultz [60]. They determined the biodegradability and toxicity of the phenols (Y variables) experimentally. The values for toxicity are the 50% growth inhibitory concentration (IGC_{50}), while the biodegradability is the second-order rate constant of the compound's oxidation (k_b). Among the descriptors used were pK_a , molar mass (M) and the logarithm of the octanol–water partitioning coefficient, $\log K_{ow}$ (X variables) (Table 24.1). The lipophilicity and acidity information was obtained

Table 24.1 Substituents and properties of the substituted phenols.

Substituents	Log K_{ow}	M (Da)	pK_a	Log(1/IC $_{50}$) (mol L $^{-1}$)	Log k_b (L org $^{-1}$ h $^{-1}$)
H	1.48	94	9.92	−0.241	−11.16
Br	2.63	173	9.45	0.500	−11.80
CH $_3$	2.12	108	10.1	−0.162	−11.33
Cl	2.48	129	9.38	0.402	−11.77
CN	1.60	119	7.96	0.516	−13.82
NO $_2$	1.85	139	7.15	1.420	−13.00
CHO	1.57	122	10.2	−0.143	−12.70
COCH $_3$	1.45	136	8.05	−0.093	−12.51

from the literature [61, 62]. Figure 24.5 shows the molecular skeleton for the series of congeneric molecules.

We tested the ability of the BCP variables (X variables) to reproduce each of the five properties in Table 24.1, now all regarded as Y variables. Table 24.2 gives the correlation coefficients obtained for all the QSARs. We show that BCP space reliably reproduces all properties except molar mass. The lack of any reasonable fit with molar mass demonstrates that BCP properties cannot function as descriptors for this quantity, which does correlate well with steric hindrance [60]. More unpublished evidence suggests that BCP properties cannot capture steric effects. In fact, they are known to perform best for electronic effects, based on later work. This is why the good correlation with log K_{ow} (Table 24.2) is surprising and needs further investigation.

An important question is now how the centre of action changes as the Y variable is varied. The program SIMCA-P prescribes a criterion for the significance of an LV, that is, if $q^2 < 0.097$ then the LV is not significant and no more LVs are computed. The PLS regression is then deemed complete. We also use the variable importance in the projection (VIP) values. The VIP gives the relative importance of each independent variable ("X") in the regression. Hence factors that contribute substantially to the fit have high VIP scores.

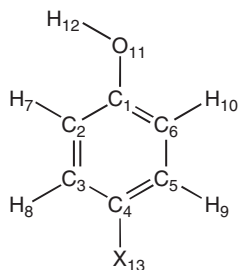
**Figure 24.5** Molecular skeleton of the para-substituted phenol molecules; X denotes the substituent.

Table 24.2 Correlations obtained with BCP variables when fitted to the properties below.

Variables	Log K_{ow}	M (Da)	pK_a	Log(1/IGC ₅₀) (mol L ⁻¹)	Log k_b (L org ⁻¹ h ⁻¹)
r^2	0.98	0.40	0.98	0.94	0.99
All variables					
q^2	0.89	0.06	0.95	0.81	0.90
All variables					
r^2 – 1st PC	0.96	0.39	0.98	0.92	0.98
q^2 – 1st PC	0.82	0.07	0.95	0.79	0.84

Figure 24.6 shows the VIP plots for each respective QSAR. Different regions of the molecular electron density are being highlighted each time. Hence without any a priori knowledge of the systems or mechanisms of action we are able to reproduce the properties of interest. The centre of action for pK_a turns out to be bond O11–H12, closely followed by bond C1–O11. The former bond is indeed the one that breaks or forms in the (de)protonation of the acid–base equilibrium (between the forward and backward reaction). Ideally, the VIP plot drops off very quickly below values of unity, as it does for the log k_b plot. Unfortunately, in the pK_a plot there is still much influence from C5–H9 and C6–H10. These spurious bonds have so far been dismissed as “contaminations.” The log K_{ow} plot highlights the bond that attaches the substituent X to the phenyl group (C4–X13). The bonds with the next highest VIP values are near the bond C4–X13 but with much reduced VIP influence. The partitioning coefficient is not a property that involves bond breaking or making, so it is perhaps not surprising that the only bond clearly highlighted (C4–X13) is not in the common skeleton. In the log k_b plot the C2=C3 bond features strongly. It needs to be verified if the oxygenase enzyme in charge of this biodegradation interacts strongly with this bond while oxygenating the phenols. In the log(1/IGC₅₀) plot the highlighted bonds are C3–H8 and C2–H7, which again awaits an experimental verification. In fact it is promising that these four properties, which are not closely related (as proven by their correlation matrix) can all be modeled by the same group of descriptors, which all arose from only 13 selected points in the electron densities.

It is a working hypothesis of QTMS that the centre of action of a molecule constitutes the BCPs associated with the highest VIPs. However, the highest VIPs may contain other parts of the molecule that cannot be readily associated with the mode of activity. This also depends on the level of theory, as is made clear by color-coded molecular diagrams (“rainbow plots”), assigning a color to predetermined brackets of VIP values [49]. At the moment we have no rigorous criterion to isolate the centre of action from the VIP plot. Nevertheless, based on our experience, we believe that QTMS has sufficient suggestive power to highlight the part of the molecule responsible for the measured response. This part is invariably spatially close to where the change due to the (re)action occurs. In some cases this highlighting is rather sharp (e.g., tri-substituted carboxylic acids [49]) in others it is more diffuse

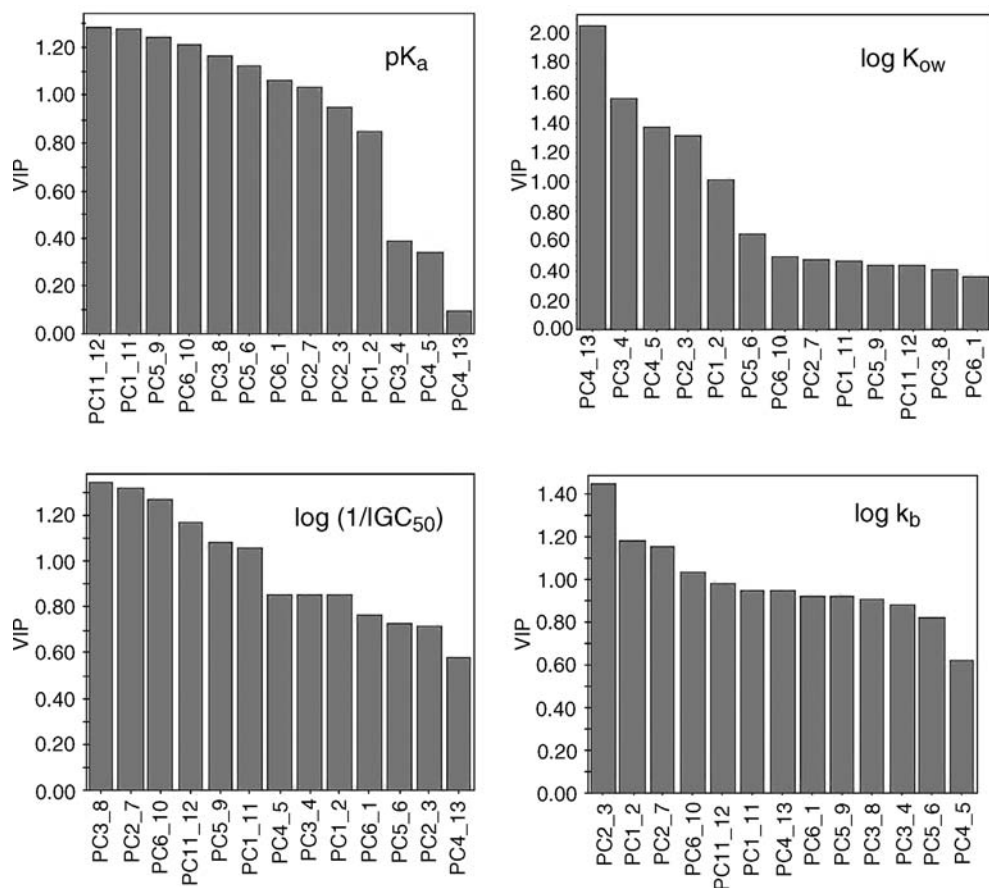


Figure 24.6 VIP plots of four different properties against principal components, for bonds appearing in Figure 24.5, constructed from BCP properties.

with some contamination. Unpublished results suggest that contaminations are not an artifact of PLS. A dedicated study targeting this problem with machine learning is warranted.

24.6

A Leap

Armed with the confidence gained from the QTMS work on the classic Hammett QSARs, an excursion [63] into medicinal chemistry seemed appropriate. The α,β -unsaturated ketone (*E*)-1-(4-hydroxyphenyl)but-1-en-3-one has been used in traditional Chinese medicine as an antitumor agent. It was found to possess moderate antitumor activity and hence a series of substituted analogues was synthesized by Ducki *et al.* [64]. They calculated the IC₅₀ concentration with reference to a standard

growth curve and represented the concentration required to cause a 50% decrease in cell growth after five days of incubation. A PLS analysis on 17 substituted ketones, calculated at the B3LYP/6-311G(2d,p)//HF-6-31G(d) level, yielded a validated regression with $r^2 = 0.91$ and $q^2 = 0.86$. QTMS highlighted a region in the molecule that made up the centre of action of a Michael addition surmised to be responsible for the mode of activity. It is rewarding that QTMS independently confirmed this hypothesis. An analysis of Mulliken charges, however, provided a QSAR with little predictability and no comprehensible insight into the mode of action.

Given the surging prominence of QSARs in environmental toxicology, QTMS was applied to a well-known set of molecules known as polychlorinated dibenzo-*p*-dioxins (PCDDs) [58]. It was feasible to examine PCDDs by QTMS because of their modest size and molecular rigidity. Indeed, since not much work had been carried out on the influence of conformational flexibility on QCT descriptors, lack of rigidity would pose a practical problem. PCDDs produce a wide span of toxic effects, most of which involve binding to the aromatic hydrocarbon (Ah) receptor, whose structure is unknown. QTMS was applied to predict three different activities (pEC_{50}) of a set of 13 ecologically relevant PCDDs. Overall, the pEC_{50} (bind) response produced the most predictive models, which is in line with the results of an earlier CoMFA study. The three measured pEC_{50} activities are thought to be mediated by a common (Ah or dioxin) receptor mechanism of action. Since the activity refers to the ability to bind to a receptor, the center of action is not involved in bond breaking as in the case of carboxylic acids. The centre of action is found to be concentrated near the lateral C–Cl bonds. Our main conclusion about the importance of the lateral side overlaps with the deductions made by Bonati *et al.* in their electrostatic (MEP) study [65] of the enzyme–substrate recognition step. It appears that a deeper QTMS analysis is needed here, perhaps with QCT atomic multipole moments, which generate the corresponding atomic electrostatic potential [66].

Subsequently, QTMS was applied to the standard “steroid set” that Cramer *et al.* [67] used to launch their CoMFA analysis. They introduced multiple errors in topological coding and stereochemistry, which were corrected by Wagener *et al.* [68]. Here AM1 and HF/3-21G* bond lengths were used in conjunction with PLS and a genetic algorithm (GA) to predict the corticosteroid-binding globulin (CBG) binding activity. Once the initial PLS model incorporating all descriptors is obtained, variable selection is performed using the GA to select the optimum number of descriptors for use in subsequent PLS analysis. A new PLS model is then obtained using the descriptors selected by the GA. Variable selection is not employed in the original analysis to compare the bonds selected by the GA with the bonds that are allocated high VIP scores in the model obtained using all the descriptors. Good r^2 and q^2 values are obtained and the notorious steroid “number 31,” called 2 α -methyl-9 α -fluorocortisol, is not found to be an outlier. It is not clear to what extent the *precise* knowledge of bond length is actually required. For example, the C₃=O₁₈ bond length, which has a high VIP value, barely varies between steroids 4, 7 and 8. The large variation between double CO bonds on one hand and single CO bonds on the other is most likely the actual reason for the correct prediction of activity. This test case involves only binding rather than a reaction, so it is not obvious why an “interior

property” such as bond length or BCP property would actually work. Using convenient color-coded plots the regions of the molecule considered important for binding activity were delineated. The regions obtained correspond well with previous binding site specificity studies.

Nine nitrofur derivatives were also analyzed [67] given their importance in medicinal chemistry, principally being used as antibacterial and anticancer agents. The compounds also possess mutagenic and carcinogenic activities. For *Staphylococcus aureus* *S. aureus* activity a model (with two LVs) was generated from the AM1 bond lengths with $r^2 = 0.92$ and $q^2 = 0.78$, and another model for (bacterium) *C. crescentus*, with $r^2 = 0.96$ and $q^2 = 0.90$, both passing the randomization test. The rainbow plot showed a dramatic difference between the center of action site for the two types of bacteria, but experimentalists have not been able to confirm or deny this (despite an invitation to do so).

QTMS work [69] on halogenated hydroxyfuranone (mutagen-X) derivatives and the mutagenic activity of triazenes led to an interesting result in the case of the latter. Details of the hydroxylation mechanism are still obscure but the two most likely alternatives have been proposed [70]. Path A involves transfer of one electron from the triazene to the enzyme cytochrome P-450. The resulting radical cation then loses a proton to give a triazene-substituted methyl radical. In the alternative, path B, this radical is formed in one step by direct abstraction of a hydrogen atom. Previously published QSAR equations [70] were unable to distinguish these two paths and suggest a preferred route. Based on analysis of the VIP plot, QTMS can decide between the two possible pathways and suggests that pathway A is favored over pathway B.

The next QTMS study [71] tackled a remarkable and unusual set of ortho alkyl substituted phenols, known for their cytotoxicity and previously investigated by the Hansch group [72]. The QTMS results do not support their proposal that a steric factor is important in the determination of the cytotoxicity. In fact, the QTMS results suggest no steric contribution whatsoever. QTMS descriptors capture electronic effects only but can be combined with externally provided descriptors of any nature (e.g., $\log P$).

Seven datasets of medicinal interest were investigated [73]. They are the dissociation constants (pK_a) for a set of substituted imidazolines, the pK_a of imidazoles, the ability of a set of indole derivatives to displace [^3H]flunitrazepam from binding to bovine cortical membranes, the influenza inhibition constants for a set of benzimidazoles, the interaction constants for a set of amides and the enzyme liver alcohol dehydrogenase, the natriuretic activity of sulfonamide carbonic anhydrase inhibitors and the toxicity of a series of benzyl alcohols. A PLS analysis in conjunction with a GA delivered excellent models. They are also able to highlight the active site of the ligand or the molecule whose structure determines the activity.

Six out of these seven sets were revisited more recently [74] by using the multiway data analysis method called molecular maps (MOLMAPs). A three-dimensional array of quantum topological molecular similarity descriptors is transferred into new two-dimensional parameters using Kohonen networks, followed by PLS. Overall, the results showed better statistical results compared with simple unfolding.

Furthermore, variables important in projection plots confirmed previous findings about active centers and even in some cases showed more accurate results. Self-organized maps (or Kohonen nets) had featured earlier in the context of QTMS [75] in an attempt to confront the notoriously difficult QSAR of 1,4-dihydropyridine calcium channel blockers.

The latest QTMS studies focused on *external* predictability, the holy grail of QSAR. A study [76] on the hepatotoxicity of phenols at five different levels of theory showed that the external predictability of the best models at the higher levels of theory is higher than that at the lower levels. Moreover, the best QTMS models are better in external predictability than the PLS models using pK_a and Hammett σ^+ along with $\log P$. The current study implies the advantage of quantum chemically derived descriptors over physicochemical (experimentally derived or tabular) electronic descriptors in QSAR studies. In view of the widespread industrial use of nitroaromatics and their consequent ecotoxicological hazard potential, QTMS again explored [77] predictive models for their toxicity to *S. cerevisiae*. QTMS descriptors were employed to complement the deficiency of E_{LUMO} in setting up predictive QSAR models from the view point of external validation. The nitro group was identified as the center of action.

Finally, returning to physical organic chemistry, QTMS also corroborated [78] that a wider class of Hammett constants can also be replaced by QCT descriptors than thought based on the early work on carboxylic acids. It emerged that they can also substitute σ^+ and σ^- by studying para-substituted phenols, substituted toluenes and bromophenethylamines. Secondly, QTMS also constructed [79] successful QSARs for the pK_a of carboxylic acids, anilines and phenols. A more recent QTMS study [80] delivered externally validated models for pK_a prediction in seven different solvents. Good predictive models were developed in all solvent systems except isopropanol. Considering all seven solvent systems, distance descriptors give consistently good results whereas ellipticity descriptors are of less importance. Moreover, VIP plots for the best models highlight the importance of the bond connecting the phenolic oxygen to the aromatic ring. Thirdly, a QTMS study [81] showed that optimized bond lengths and BCP descriptors perform well in predicting the rate hydrolysis of a set of 40 esters. This work is relevant for environmental exposure and risk analysis. Models for three subsets, each having a different common skeleton, point toward the three central ester bonds as the active center. A new model that just includes those three bonds links the Laplacian of the electron density at the C=O BCP to the formal reaction mechanism of base-promoted ester hydrolysis. This successful application of the QTMS method demonstrates that there is no need to measure infrared frequencies, which itself has been introduced as a faster method to avoid the measurement of rate constants.

24.7

A Couple of General Reflections

Molecular quantum similarity measures (MQSMs) have also been used to predict $\log P$ values. In this work [82], the full set of 58 molecules investigated was cut into

eight subsets, each featuring a functional group. The set of (only) five acetic acid esters (CH_3COOR), for example, differed in the length of the *R* aliphatic chain. In each subset, $\log P$ increased by an approximately constant increment. So did the MQS measure. Therefore it is not surprising that a very high correlation coefficient is obtained for the linear equation governing $\log P$ as a function of the MQSM. One would also obtain a very high correlation if one counted the methylene groups in the growing aliphatic chains in each subset. This is an example of the view, held by some industrial scientists, that there are “expensive ways of counting carbon.”

In a further effort to speed up the computation of quantum molecular similarity measures (QMSMs) (an abbreviation coexisting with MQSM, carelessly introduced by the same research group) the atomic shell approximation (ASA) was proposed [83]. Essentially, the molecular electron density is expressed as a linear combination of Gaussian *s* functions. This overcomes the highly time-consuming evaluation of many center integrals. This approximation by spherically symmetric functions can lead to negative expansion coefficients, which is avoided by ASA, however. A follow-on paper [84] then proposed to apply the ASA approximation to promolecular electron densities, which are a mere superposition of atomic densities. They are popular in high-resolution crystallography [85] where they are subtracted from measured molecular electron densities to reveal chemical features. Superficially, promolecular and molecular electron densities look very similar. The difference between them, however, is exactly what makes a molecular electron density more than a sheer set of superimposed atoms. Chemical features express themselves only as a “ripple” on top of the promolecular electron density. If the latter covered chemical information it would be futile to spend CPU hour after CPU hour solving the molecular Schrödinger equation. So, surely, the “ripple” is very important, and is the “only” quantity that matters if one wants to truly characterize a molecule. It is then curious to read [84] that promolecular electronic densities describe molecular densities with sufficient accuracy for MQSM. Does this not infer that MQSM is too blunt an instrument to detect (and connect to) the chemically important features of a set of molecules?

24.8

Conclusions

We have shown how new QSAR descriptors can be easily and efficiently extracted from geometry optimized wavefunctions. The topology of the electron density is the guide, through the gradient vector field, which lies at the heart of quantum chemical topology (QCT). A method called quantum topological molecular similarity (QTMS) uses QCT descriptors in combination with a machine learning method, such as genetic algorithms and including partial least-squares. Once inspired by the quantum molecular similarity measures (QMSMs) approach of Carbó and coworkers, we have proven, however, that QTMS avoids this method’s disadvantages. The use of simple equilibrium bond lengths as descriptors to capture electronic effects is an attractive proposal still waiting to be applied on a routine scale. The independent

ability of QTMS to suggest the center of action in a set of congeneric molecules is an asset to experimental studies. QTMS can confirm experimental hypotheses or suggest preferred pathways. In a recent review [86] on the description of electron delocalization via the analysis of molecular fields, QCT features heavily. It is hoped that the wealth of indices and case studies in which they helped increase insight can also benefit and extend QTMS. However, one should remain vigilant against the snag that “there are expensive ways of counting carbons.” Extracting descriptors from quality *ab initio* levels of theory is appealing since approaching an underlying quantum reality must surely lead to true predictive power. This presumed veracity should be carefully balanced against computational parsimony, however. An important question that keeps nagging QSAR is whether a descriptor works for the right (physical or quantum chemical) reason. If *ab initio* equilibrium bond lengths are a genuine and independent source of information, then they are better than *ad hoc* measures of dubious theoretical background. But if counting carbons is fundamentally justified then one should not spend CPU days superimposing high level electron densities.

Acknowledgments

Mr Alex Harding is gratefully acknowledged for his critical reading of the manuscript and helpful comments.

References

- 1 Cavalli, A., Carloni, P., and Recanatini, M. (2006) *Chem. Rev.*, **106**, 3497–3519.
- 2 Karelson, M., Lobanov, V.S., and Katritzky, A.R. (1996) *Chem. Rev.*, **96**, 1027–1043.
- 3 Hansch, C. and Fujita, T. (1964) *J. Am. Chem. Soc.*, **86**, 1616–1626.
- 4 Hansch, C. and Leo, A. (1995), in Exploring QSAR: fundamentals and applications in chemistry and biology, *Computer Applications In Chemistry*, vol. 1, 1st edn, The American Chemical Society, Washington DC.
- 5 Hansch, C., Leo, A., and Taft, R.W. (1991) *Chem. Rev.*, **91**, 165–195.
- 6 Miller, B. (1998) *Advanced Organic Chemistry. Reactions and Mechanisms*, Prentice-Hall, New Jersey, USA.
- 7 Johnson, C.D. (1973) *The Hammett Equation*, Cambridge University Press, GB.
- 8 Jaffe, H.H. (1953) *Chem. Rev.*, **53**, 191.
- 9 Doweyko, A.M. (2008) *J. Comput. Aided Mol. Des.*, **22**, 81–89.
- 10 Topliss, J.G. and Edwards, R.P. (1979) *J. Med. Chem.*, **22**, 1238–1244.
- 11 Bader, R.F.W. (1990) *Atoms in Molecules. A Quantum Theory*, Oxford University Press, Oxford, GB.
- 12 Popelier, P.L.A. (2000) *Atoms in Molecules; An Introduction*, Pearson Education, London.
- 13 Popelier, P.L.A. and Bremond, E.A.G. (2009) *Int. J. Quantum Chem.*, **109**, 2542–2553.
- 14 Popelier, P.L.A. (1994) *Theor. Chim. Acta*, **87**, 465–476.
- 15 Gadre, S.R., Kulkarni, S.A., and Shrivastava, I.H. (1992) *J. Chem. Phys.*, **96**, 5253–5261.
- 16 Keith, T.A., Bader, R.F.W., and Aray, Y. (1996) *Int. J. Quantum Chem.*, **57**, 183–198.
- 17 Keith, T.A. and Bader, R.F.W. (1993) *J. Chem. Phys.*, **99**, 3669–3682.
- 18 Silvi, B. and Savin, A. (1994) *Nature*, **371**, 683–686.
- 19 Cioslowski, J. and Liu, G.H. (1999) *J. Chem. Phys.*, **110**, 1882–1887.

- 20 Blanco, M.A., Pendás, A.M., and Francisco, E. (2005) *J. Chem. Theor. Comput.*, **1**, 1096–1109.
- 21 Popelier, P.L.A. and Aicken, F.M. (2003) *ChemPhysChem*, **4**, 824–829.
- 22 Devereux, M., Popelier, P.L.A., and McLay, I.M. (2009) *J. Chem. Inf. Model*, **49**, 1497–1513.
- 23 Popelier, P.L.A. (2005) *Structure and Bonding. Intermolecular Forces and Clusters*, vol. 115 (ed. D.J. Wales), Springer, Heidelberg, Germany, pp. 1–56.
- 24 Webster, B. (1990) *Chemical Bonding Theory*, Blackwell, Oxford, Great Britain.
- 25 Reed, A.E., Curtiss, L.A., and Weinhold, F. (1988) *Chem. Rev.*, **88**, 899–926.
- 26 Stone, A.J. (1981) *Chem. Phys. Lett.*, **83**, 233–239.
- 27 Mulliken, R.S. (1955) *J. Chem. Phys.*, **23**, 1833–1940.
- 28 Pearson, R.G. (2007) *Int. J. Quantum Chem.*, **108**, 821–826.
- 29 Kovacs, A., Esterhuysen, C., and Frenking, G. (2005) *Chem.–Eur. J.*, **11**, 1813–1825.
- 30 McWeeny, R. (1992) *Methods of Molecular Quantum Mechanics*, 2nd edn, Academic Press, San Diego, USA.
- 31 Carbó, R., Leyda, L., and Arnau, M. (1980) *Int. J. Quantum Chem.*, **17**, 1185–1189.
- 32 Amoores, J.E. (1970) *Molecular Basis of Odor*, Charles C Thomas, Springfield, Illinois, USA.
- 33 Amoores, J.E., Palmieri, G., Wanke, E., and Blum, M.S. (1969) *Science*, **165**, 1266.
- 34 Good, A.C., Hodgkin, E.E., and Richards, W.G. (1992) *J. Chem. Inf. Comput. Sci.*, **32**, 188–191.
- 35 Richard, A.M. and Rabinowitz, J.R. (1987) *Int. J. Quantum Chem.*, **31**, 309–323.
- 36 Burt, C., Richards, W.G., and Huxley, P. (1990) *J. Comput. Chem.*, **11**, 1139.
- 37 Bowen-Jenkins, P.E. and Richards, W.G. (1986) *J. Chem. Soc., Chem. Commun.*, 133–135.
- 38 Cooper, D.L., Mort, K.A., Allan, N.L., Kinchington, D., and McGuigan, C. (1993) *J. Am. Chem. Soc.*, **115**, 12615–12616.
- 39 Popelier, P.L.A. (1995) *Molecular Similarity in Drug Design* (ed. P.M. Dean), Chapman & Hall, London, pp. 215–240.
- 40 Popelier, P.L.A. (1994) *Chem. Phys. Lett.*, **228**, 160–164.
- 41 Popelier, P.L.A. (1996) *Comput. Phys. Commun.*, **93**, 212–240.
- 42 Lee, C. and Smithline, S. (1994) *J. Phys. Chem.*, **98**, 1135–1138.
- 43 Popelier, P.L.A. (1999) *J. Phys. Chem. A*, **103**, 2883–2890.
- 44 Everitt, B. (1980) *Cluster Analysis*, 2nd edn, Halsted Press, London, GB.
- 45 Besalu, E., Girones, X., Amat, L., and Carbó-Dorca, R. (2002) *Acc. Chem. Res.*, **35**, 289–295.
- 46 Amat, L., Besalu, E., Carbó-Dorca, R., and Ponec, R. (2001) *J. Chem. Inf. Comput. Sci.*, **41**, 978–991.
- 47 Bader, R.F.W. and Preston, H.J.T. (1969) *Int. J. Quantum Chem.*, **3**, 327–347.
- 48 O'Brien, S.E. and Popelier, P.L.A. (1999) *Can. J. Chem.*, **77**, 28–36.
- 49 O'Brien, S.E. and Popelier, P.L.A. (2001) *J. Chem. Inf. Comput. Sci.*, **41**, 764–775.
- 50 Wold, S. (1993) *Technometrics*, **35**, 136–139.
- 51 Wold, S., Sjostrom, M., and Eriksson, L. (1998) in *Encyclopedia of Computational Chemistry*, vol. 3 (ed. P.v.R. Schleyer), John Wiley & Sons, Ltd, Chichester, GB, pp. 2006–2021.
- 52 Geladi, P. and Kowalski, B.R. (1986) *Anal. Chim. Acta*, **185**, 1–17.
- 53 Livingstone, D.J. (1995) *Data Analysis for Chemists*, 1st edn, Oxford University Press, GB.
- 54 Ponec, R., Amat, L., and Carbó-Dorca, R. (1999) *J. Comput.-Aided Mol. Des.*, **13**, 259–270.
- 55 Wold, S., Kettaneh, N., and Tjessem, K. (1996) *J. Chemometrics*, **10**, 463–482.
- 56 SPSS Inc. (2000) SPSS version 10.0.7, <http://spss.com>, Chicago, USA.
- 57 UMETRICS (1998) info@umetrics.com: www.umetrics.com, Umeå, Sweden.
- 58 Popelier, P.L.A., Chaudry, U.A., and Smith, P.J. (2002) *J. Chem. Soc., Perkin Trans. 2*, 1231–1237.
- 59 Exner, O. and Budesinsky, M. (1989) *Magn. Reson. Chem.*, **27**, 27–36.
- 60 Damborsky, J. and Schultz, T.W. (1997) *Chemosphere*, **34**, 429–446.
- 61 Perrin, D.D., Dempsey, B., and Serjeant, E.P. (1981) *pKa Prediction for Organic Acids and Bases*, Chapman & Hall, London, GB.
- 62 Dean, J.A. (1992) *Lange's Handbook of Chemistry*, 14th edn, McGraw-Hill, New York.
- 63 O'Brien, S.E. and Popelier, P.L.A. (2002) *J. Chem. Soc., Perkin Trans. 2*, 478–483.

- 64 Ducki, S., Hadfield, J.A., Hepworth, L.A., Lawrence, N.J., Liu, C.Y., and McGown, A.T. (1997) *Bioorg. Med. Chem. Lett.*, **7**, 3091–3094.
- 65 Bonati, L., Fraschini, E., Lasagni, M., and Pitea, D. (1994) *J. Mol. Struct. (THEOCHEM)*, **303**, 43–54.
- 66 Kosov, D.S. and Popelier, P.L.A. (2000) *J. Chem. Phys.*, **113**, 3969–3974.
- 67 Smith, P.J. and Popelier, P.L.A. (2004) *J. Comput. -Aided Mol. Des.*, **18**, 135–143.
- 68 Wagener, M., Sadowski, J., and Gasteiger, J. (1995) *J. Am. Chem. Soc.*, **117**, 7769–7775.
- 69 Popelier, P.L.A., Chaudry, U.A., and Smith, P.J. (2004) *J. Comput. -Aided Mol. Des.*, **18**, 709–718.
- 70 Shusterman, A.J., Debnath, A.K., Hansch, C., Gregory, W.H., Frank, R.F., Greene, A.C., and Watkins, S.F. (1989) *Mol. Pharm.*, **36**, 939–944.
- 71 Loader, R.J., Singh, N.K., O'Malley, P.J., and Popelier, P.L.A. (2006) *Bioorg. Med. Chem. Lett.*, **16**, 1249–1254.
- 72 Selassie, C.D., Verma, R.P., Kapur, S., Shusterman, A.J., and Hansch, C. (2002) *J. Chem. Soc., Perkin 2*, 1112–1117.
- 73 Popelier, P.L.A. and Smith, P.J. (2006) *Eur. J. Med. Chem.*, **41**, 862–873.
- 74 Hemmateenejad, B., Mehdipour, A.R., and Popelier, P.L.A. (2008) *Chem. Biol. Drug Des.*, **72**, 551–563.
- 75 Chaudry, U.A., Singh, N.K., and Popelier, P.L.A. (2007) in *Theoretical Aspects of Chemical Reactivity* (ed. A. Toro-Labbé), Elsevier, Netherlands, Chapter 15, pp. 301–317.
- 76 Roy, K. and Popelier, P.L.A. (2008) *Bioorg. Med. Chem. Lett.*, **18**, 2604–2609.
- 77 Roy, K. and Popelier, P.L.A. (2008) *QSAR Comb. Sci.*, **27**, 1006–1012.
- 78 Smith, P.J. and Popelier, P.L.A. (2005) *Org. Biomol. Chem.*, **3**, 3399–3407.
- 79 Chaudry, U.A. and Popelier, P.L.A. (2004) *J. Org. Chem.*, **69**, 233–241.
- 80 Roy, K. and Popelier, P.L.A. (2009) *J. Phys. Org. Chem.*, **22**, 186–196.
- 81 Chaudry, U.A. and Popelier, P.L.A. (2003) *J. Phys. Chem. A*, **107**, 4578–4582.
- 82 Amat, L., Carbó-Dorca, R., and Ponec, R. (1998) *J. Comput. Chem.*, **19**, 1575–1583.
- 83 Constans, P. and Carbo, R. (1995) *J. Chem. Inf. Comput. Sci.*, **35**, 1046–1053.
- 84 Amat, L. and Carbo-Dorca, R. (1997) *J. Comput. Chem.*, **18**, 2023–2039.
- 85 Koritsanszky, T.S. and Coppens, P. (2001) *Chem. Rev.*, 1583–1627.
- 86 Merino, G., Vela, A., and Heine, T. (2005) *Chem. Rev.*, **105**, 3812–3841.

25

Quantum-Chemical Descriptors in QSAR/QSPR Modeling: Achievements, Perspectives and Trends*Anna V. Gubskaya*

25.1

Introduction

Quantitative structure–activity or property relationships (QSAR/QSPR) modeling has been used in the fields of medicinal (drug-design, toxicology), industrial, agricultural and environmental chemistry for more than 40 years. During the last decade it has also been successfully applied in biochemistry, molecular biology and material science. The concept of QSAR is based on the postulate that the structure of a molecule represented by selected molecular characteristics (descriptors) can be correlated to its biologic activity. Once such a correlation is established for the compounds with known biologic activity, it becomes possible using specified computational protocol to predict biologic activities for new or untested chemicals of the same class. The growing popularity of modern QSAR can be attributed to its ability to select compounds with desirable biologic response from combinatorial libraries containing thousands of molecules *in silico*, that is, without synthesis as well as time and labor-consuming screening.

Several excellent recent reviews published on QSAR, in addition to offering a historical overview, describe the most prominent trends in this field [1–3]. One of these trends is associated with the rapidly increasing amount of molecular descriptors represented by quantum-chemical or by various classical parameters [e.g., constitutional, topological, connectivity, weighted holistic invariant molecular descriptors (WHIM), etc.] that were designed and tested as potential variables for QSAR modeling. To date, the work of Todeschini *et al.* [4] is the most comprehensive source of QSAR descriptors. Quantum-chemical parameters represent a special class of molecular properties. They can be obtained from sophisticated *ab initio* calculations or by means of relatively inexpensive semiempirical methods, but in either case such calculations require more time and effort than those for one, two or three-dimensional classical parameters, which can be computed from molecular structures within a few minutes. However, in contrast to most classical descriptors, quantum chemical parameters are capable of expressing all the electronic and

geometric properties of the molecules being analyzed as well as their interactions. Therefore, in some cases the interpretation of quantum-chemical descriptors can provide much deeper insights into the nature of biologic or physicochemical mechanisms under consideration than that of classical descriptors.

To the best of the author's knowledge, the latest comprehensive review on quantum-chemical descriptors and their applications in QSAR studies was published by Karelson *et al.* [5] in 1996. Since then, significant progress has been achieved in computational hardware, the development of quantum mechanical methodologies such as density functional theory (DFT) [6, 7], the concepts of molecular quantum similarity measure and quantum topological molecular similarity [8] as well as the design of corresponding algorithms. Advances in combinatorial chemistry have given a new start to the use of machine-learning methods for the selection of the most relevant to specific bioactivity descriptors and to the development of more sophisticated QSAR models. The increasing complexity of investigations in life science-related fields has facilitated the process of generating and testing new molecular (including quantum-chemical) descriptors. As most biologic processes take place in aqueous media, the advantages of descriptors calculated by means of quantum-chemical approaches that account for specific and non-specific solvation effects are of prime importance. The present chapter focuses on QSAR/QSPR studies in biologic sciences carried out during past decade or so. It will cover the recent applications of quantum descriptors and the new conceptual and methodological trends associated with their use. The capability of quantum descriptors in predicting biologic activities and biologically important properties will be demonstrated.

25.2

Quantum-Chemical Methods and Descriptors

25.2.1

Quantum-Chemical Methods

Calculations of electronic properties of a molecule that have potential value to QSAR studies can be performed by various quantum-mechanical methods. These methods, represented by two major groups, *ab initio* and semiempirical methods, have been further classified and their methodological details, corresponding approximations as well as advantages of utilization have been described by Karelson *et al.* [5]. The authors mention no applications of density functional theory [6, 7] in QSAR studies, while in about one half of 80 original research articles reviewed in the present work DFT formalism was used to obtain descriptors for highly predictive QSAR/QSPR models. Among semiempirical methods, Austin model 1 (AM1) [9], modified neglect of differential overlap (MNDO) [10] and parametric model 3 (PM3) [11], known as evolution of MNDO parameterization, were chosen in 16, 4 and 13% of cases, respectively. Approximately 20% of the studies reviewed here were devoted to the development and applications of quantum similarity approaches. In several cases *ab initio* methods [12], namely, Hartree-Fock (12%) and Møller-Plesset theory of

second order (MP2) (3%), were used to calculate quantum descriptors. Some of these studies reconfirmed the conclusion made by Karelson *et al.* [5] that electronic descriptors as well as optimized geometrical parameters obtained from AM1 and PM3 calculations are more satisfactory than those from *ab initio* calculations carried out with insufficiently large basis set [13].

The development of DFT accelerated the utilization of electronic structure theory in determining molecular properties for biologically significant molecules. The DFT method belongs to the group of *ab initio* methods that allow calculations of quantum-chemical descriptors at a reasonable cost and with higher accuracy than that of semiempirical methods. QSAR models generated using quantum descriptors obtained by DFT were found to be more predictive than the models incorporating descriptors calculated by AM1 [14, 15] or PM3 [15] methods. The fact that DFT accounts for dynamic correlation effects makes it an attractive alternative to the Hartree–Fock (HF) method as well as the much more CPU-demanding post-HF methods: MP theory, coupled-cluster theory and configuration interaction approach [5]. In the commonly applied Kohn–Sham DFT formalism the functional of electron density, in addition to classical (kinetic and electrostatic) energy terms, includes contribution from exchange–correlation energy. Becke’s three-parameter hybrid exchange functional and the Lee–Yang–Parr correlation functional (B3LYP) [16–18] is probably the most popular hybrid density functional used in QSAR-related DFT calculations at the time of writing. Omitting electron correlations in HF theory lowers the accuracy of computations in comparison with those carried out using DFT: statistical parameters of QSAR models obtained with descriptors computed at B3LYP/LANL2DZ and at HF/LANL2DZ levels showed the obvious advantage of DFT-based descriptors [19]. Several QSPR studies in material science demonstrated successful utilization of thermochemistry data calculated at the B3LYP/6-31G(d) level in predicting physicochemical properties of polymers [20–22]. Since density functional does not account for dispersion energy [23], its applicability might be limited in cases where contribution from dispersion interactions is expected or known to be significant.

Among all quantum-mechanical methods used for calculation of QSAR descriptors the quantum methodologies dealing with the principle of molecular similarity represent efficient tools for solving various chemistry-related problems. The molecular quantum similarity measures (MQSMs) approach developed by Carbó-Dorca and co-authors [24–27] establishes a quantitative measure of resemblance between two molecules based on their first-order density functions (DFs), constructed in a specific internal energy state. It was also possible to include one of extended DFs, namely kinetic energy DF, into MQSM formalism and to use it to correlate the antimalarial activity of two series of compounds [28]. In the framework of MQSM approach [27] a quantitative molecular similarity measure between two molecules (or molecular fragments [29]), A and B, described by density functions $\varrho_A(\mathbf{r})$ and $\varrho_B(\mathbf{r})$ is expressed as a direct volume integral:

$$Z_{AB}(\Omega) = \iint \varrho_A(\mathbf{r}_1) \Omega(\mathbf{r}_1, \mathbf{r}_2) \varrho_B(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (25.1)$$

where $\Omega(\mathbf{r}_1, \mathbf{r}_2)$ is a positive definite two-electron operator that acts as a weighting factor. The choice of this operator determines the type of QSM; for example, the identification of Ω with the Dirac delta function or with the Coulomb repulsion term produces overlap-like or Coulomb-like similarity measures [29], respectively. A quantitative measure is transformed into an absolute magnitude by means of the Carbó index given as:

$$C_{AB} = Z_{AB}(Z_{AA}Z_{BB})^{-1/2}. \quad (25.2)$$

The closer to unity the Carbó index is the more similar are the two compounds subject to comparison. All quantitative measures computed for molecular pairs can be transformed into Carbó indexes and represented as a matrix form whose columns are used as MQSM molecular descriptors [24, 26, 28]. Computations of Z_{AB} integral can be very time-consuming and thus the authors proposed the so-called atomic shell and promolecular approximations [30, 31] to reduce the MQSM computational costs without a significant loss of accuracy.

MQSM originally introduced by Carbó-Dorca provided a fundamental framework for the quantum-mechanical description of molecular similarity and stimulated the development of various computational methods and theories focused on this problem. Mezey and co-authors [32–35] have made a substantial contribution to shape similarity analysis, specifically to the shape analysis of electron density, $q(\mathbf{r})$. The authors introduced the concepts of density domain and molecular isodensity contour surface and proposed two families of shape analysis techniques: the shape group methods [32] and T-hull method with its extensions [33]. Mezey also reviewed Münch and Reiss's proof that a unique mapping exists between the electron density of a subsystem and that of the entire system [35, 36]. The principles and methods of the molecular shape analysis were reflected in the new the molecular electron density “lego” assembler (MEDLA) [34] method for rapid “*ab initio* quality” computation of shapes for large molecules. The MEDLA approach is based on a simple electron density fragment additivity principle. The fragment density matrices (along with the basis set information) are stored in the MEDLA data bank and are used as fuzzy building blocks for the construction of electron densities for target molecules of any size. In the QShAR (here “Sh” stands for “shape”) study of toxicological risk assessment of polycyclic aromatic hydrocarbons (PAHs) to *L. gibba*, utilization of similarity measures calculated by the MEDLA method provided excellent correlation coefficients [7, 34]. It is somewhat surprising that this interesting method had not yet been tested in a wider variety of cases. A conceptually similar (in terms of additivity and transferability) but methodologically alternative approach for computing the properties of large molecules not amenable to direct computation was presented by Matta [37]. This author described a real space approach to reconstructing the electron density of a large molecule from electron density fragments extracted from molds [37], using the quantum theory of atoms in molecules (QTAIM) developed by Bader [38] and members of his school [39–41].

Bader's QTAIM [38, 42] represents another promising approach to molecular similarity. Popelier and coworker [43–45] have proposed and extensively described

applications of a method called quantum topological molecular similarity (QTMS) – reviewed by Popelier in Chapter 24. Herein is provided only a brief description of it, supported by the most recent references. The AIM theory takes advantage of the topology of the electron density distribution in a molecule. The QTMS method probes the electron density of a molecule at critical points, specifically the points in 3D space where the gradient of electron density vanishes ($\nabla \rho = 0$). The most chemically informative among the four types of critical points is the bond critical point (BCP), which is located on the intersection of the interatomic surface between bonded nuclei and the bond path linking these two nuclei. The BCP represents a quantum-chemical “signature” of a bond and can be identified by evaluating the Hessian matrix of the electron density at the location of the BCP. This second derivative matrix is composed of three eigenvalues ($\lambda_1, \lambda_2, \lambda_3$) and three associated eigenvectors. The first two eigenvalues represent the local curvatures that are perpendicular to the bond path and they must be negative, while the last eigenvalue is positive and corresponds to the curvature along the bond path. The properties evaluated at BCP are used as descriptors, that is, as measures of quantum topological similarity [46–53].

To date the QTAIM provides an essential link between the rigorous theory of quantum mechanics and the processes usually described by organic chemistry, that is, those that involve atoms as parts of larger molecular fragments. One of the applications of AIM approach in QSAR, the transferable atom equivalent method (TAE/RECON) [54], utilizes atomic contributions to generate electron-density-derived descriptors that approximate regular descriptors available through *ab initio* calculations. Mazza *et al.* [54] have employed this approach to model protein retention in ion-exchange systems.

25.2.2

Quantum-Chemical Descriptors: Classification, Updates

The choice of descriptors is crucial for obtaining a highly predictive QSAR model, whether they are chosen to be classical, electronic, that is, derived from quantum-mechanical calculations, or experimentally measured quantities or selected representatives of different types. For instance, to predict anticancer activity of carbocyclic analogs of nucleosides, Yao *et al.* [55] considered electronic descriptors such as charges, molecular orbital characteristics and polarity measures to account for the drug–receptor interaction effects and the computed octanol–water partition coefficient, solvent accessible surface areas and molecular volumes to model the drug delivery and size effects. The benefit of calculated parameters is that in contrast to experimental properties they are reproducible. They can be calculated for a set of compounds in question by defined software applying the same theoretical or methodological approximations. The errors associated with the assumptions needed for facilitating quantum calculations are considered to be constant within a series of related compounds and for most cases the direction of possible errors are known [5]. The additional advantage of quantum-chemical descriptors is that they allow characterization of an entire molecule as well as its fragments and substituents.

Table 25.1 gives a summary and classification of quantum-chemical descriptors available for contemporary QSAR/QSPR studies. The classification scheme in Table 25.1 closely reflects, combines and extends the classifications of quantum-chemical descriptors proposed by Karelson *et al.* [5] and later by Todeschini *et al.* [4]. Readers are also referred to these reviews for details about the descriptors of major groups, in terms of usage and physicochemical significance: atomic charges, molecular orbital characteristics, energy and polarity measures [5] as well as DFT-based descriptors (e.g., softness and hardness indices) [4]. Quantum-chemical measures that have recently been introduced in the literature in the context of potential application in QSAR studies are described below.

The special group of descriptors defined as electronic indices was derived by means of electronic indices methodology (EIM) to correlate structures of polycyclic aromatic hydrocarbons and their carcinogenic activity [56, 57]. This approach has been applied to investigate 5 α -reductase inhibitory properties of benzquinolizin-3-ones [13]. The EIM approach uses concepts of density of states (DOS, i.e., the number of electronic states per energy unit) and local density of states (LDOS, i.e., density of states calculated over a specific region or atom) to estimate the contributions of the specific regions of the molecules to physicochemical or biologic response (s) [56]. The EIM indices are represented by the values of relative HOMO and HOMO-1 contributions (η_H) to the LDOS over the ring that contains the highest bond order and by the critical value of their energy separation (ΔH). The authors showed that QSAR models for carcinogenic activity of PAHs based on EIM indices exhibit about 80% of predictive power [57] and that the EIM approach performs remarkably well in constructing rules and patterns of classification for benzo[c]quinolizin-3-ones according to their biologic activity [13].

Clare and co-authors [58–61] have introduced new descriptors, the frontier orbital phase angles (FOPAs), which together with flip regression and orbital nodal orientation calculation methods were specifically designed for QSAR modeling of drug-like compounds containing five- or six-membered aromatic rings. The authors suggested that the FOPA parameters (e.g., $S2\Theta_H$, $C2\Theta_H$, $S4\Theta_H$, $C4\Theta_H$ variables in Table 25.1) affect activity because they approximate the actual orientation of the nodes of the orbitals in the compounds and that the π -like orbitals of aromatic substances presumably interact with π -like orbitals of the receptor [58]. This approach was successfully used to establish drug–receptor correlations for phenylalkylamine hallucinogens [58] and to model inhibitory activities of some carbonic anhydrase inhibitors [59], flavonoid analogs [60] and phenylisopropylamines [61].

Quantum similarity descriptors are represented by Hodgkin–Richards indices [62], the reactivity based similarity indices [63] and by the most often used Carbó indices. Interestingly, both Carbó and Hodgkin–Richards indices were used in a molecular quantum similarity (MQS) study [64] to assess the difference in information that can be obtained from conceptual DFT descriptors, specifically the electron density, the shape function, the Fukui functions and the local softness. In contrast to the Hodgkin–Richards index, the Carbó index clearly revealed that within the set of congeneric steroids the density function and local softness contain different chemical information while the shape function and the Fukui function are

Table 25.1 Classification of quantum-chemical descriptors.

Symbol ^{a)}	Name/definition
Energy measures	
E_T (TE)	Total energy [4, 5, 20–22, 84, 107, 109, 112, 124, 125, 128, 136, 139, 146]
E_e (EE)	Electronic energy [68, 112, 125, 128]
CCR	Core–core repulsion energy [125, 128]
E_b	Binding energy [5]
ΔH_f°	Heat of formation [4, 5, 112, 125]
$\Delta\Delta H_f^\circ$	Relative heat of formation [5, 152]
ΔH_{prot} (ΔE)	Energy of protonation given as the difference between the total energy of the protonated and neutral species [5, 143]
IP	Ionization potential [4, 5, 104, 139]
EA	Electron affinity [4, 5, 104]
$\chi_{\text{MU,PU, etc.}}$	Atom or molecular electronegativity given in different scales (Mulliken, Pauling, etc.) [4, 82, 86, 98, 101, 114]
ESG _i	Sanderson group electronegativity calculated as the geometric mean for atoms comprising considered group [4]
χ_μ	Orbital electronegativity of the μ -th atomic orbital [4]
μ	Electronic chemical potential for a molecule of N_{el} electrons [4, 104]
Local quantum-chemical properties	
$Q(\mathbf{r})$	Electron density [4]
$\bar{I}(\mathbf{r})$	Average local ionization energy [4]
$P, P_{\mu\nu}, P_{\mu\mu}$	Charge density matrix and its elements [4]
B_{ij}	Bond index, a measure of the multiplicity of bonds between two atoms [4]
V_i	Valency index, the valency of the i -th atom as the sum of the valencies of its atomic orbitals [4]
F_i	Free valence index, a measure of the residual valency of the i -th atom in π -electron molecular orbitals [4]
$F'_i, F'_i \%$	General free valence index, a representation of the residual covalent binding capacity of the i -th atom [4]
$\nu(\mathbf{r})$	Composite nuclear potential for a given configuration of the nuclei of a molecule [4]
$S(\mathbf{r},s)$	Somoyai function, representing the difference between the electronic density $Q(\mathbf{r})$ and the composite nuclear potential $\nu(\mathbf{r})$ at a point $\mathbf{r}$ and providing the information about chemical bonding [4]
MEP	Molecular electrostatic potential – defines the interaction energy of a molecule with a unit positive charge at position $\mathbf{r}$ [4, 74]
$EP_i (P_i)$	Electrostatic potential on the specified i -th atom [82, 101, 116]
MNEP, LNEP ($P_{\text{min}}, P_{\text{max}}$)	The most negative and the least negative electrostatic potentials [82, 101]

(Continued)

Table 25.1 (Continued)

Symbol ^{a)}	Name/definition
F_k^α (FF)	Fukui function – defines electrophilicity associated with a site k in a molecule, where α represents local electrophilic quantities describing nucleophilic (+), electrophilic (–) and radical (0) attacks [4, 48, 66]
ω_g (ω)	Group or generalized electrophilicity [4, 48, 82, 98, 101, 104, 114]
Molecular orbital characteristics	
$E_{\text{HOMO}}, E_{\text{LUMO}}$ (HOMO, LUMO)	Energies of the highest occupied and the lowest unoccupied molecular orbitals [4, 5, 15, 21, 55, 60, 75, 82, 84, 86, 98, 108, 109, 113–115, 124, 125, 128, 136, 137, 142, 146, 147]
$E_{\text{LUMO}} - E_{\text{HOMO}}$ (GAP, Δ_{LH})	Difference between the HOMO and LUMO energies [4, 5, 55, 84, 98, 114, 124, 128, 147]
$E_{\text{LUMO}} + E_{\text{HOMO}}$	Sum of HOMO and LUMO energies [124, 128]
E_{SOMO} (SOMO)	Energy of the singly occupied molecular orbital [110]
$E_{\text{LUMO}} - E_{\text{SOMO}}$	Difference between the SOMO and LUMO energies [110]
$f_{\text{H/L}}$ (r_{LH})	HOMO/LUMO energy fraction, a stability index given by the ratio between HOMO and LUMO energies [4, 55]
HOP, LUP (HOPO, LUPO)	Energies of the highest occupied and the lowest unoccupied π orbitals [60, 61]
$\Phi_{\text{H}}, \Phi_{\text{L}}$	Angle between node in highest occupied π orbital and specific functional group [59]
$S2\Theta_{\text{H}}, C2\Theta_{\text{H}}, S4\Theta_{\text{H}}, C4\Theta_{\text{H}}$ $S2\Theta_{\text{L}}, C2\Theta_{\text{L}}, S4\Theta_{\text{L}}, C4\Theta_{\text{L}}$	Variables that account for orientation of nodes in π -like orbitals defined by sin and cos of the nodal angle in the highest occupied and lowest unoccupied molecular orbitals, respectively [60, 61]
η	Absolute hardness index [66, 82, 86, 98, 101, 104, 110, 114]
$\Delta\eta$	Activation hardness index represents the difference between absolute hardness of reactant and transition states [4]
S	Total softness index [66, 82, 86, 98, 101, 104]
Orbital electron densities	
$f_i^{\text{E}}, f_i^{\text{N}}$ (f_i^-, f_i^+)	Electrophilic and nucleophilic frontier electron density of atom i at HOMO and LUMO, respectively [4, 5, 115]
$F_i^{\text{E}}, F_i^{\text{N}}$ (F_i^-, F_i^+)	Indices of electrophilic and nucleophilic frontier electron density [4, 5, 113]
Superdelocalizability measures	
ES_i, NS_i (S_i^-, S_i^+)	Electrophilic and nucleophilic superdelocalizabilities measure, respectively, the availability of electrons in the i -th atom and the availability of space on the i -th atom for additional electron density [4, 5, 80, 142]
$\text{ES}_{\text{T}}, \text{NS}_{\text{T}}$ ($S_{\text{T}}^-, S_{\text{T}}^+$)	Total electrophilic and nucleophilic superdelocalizability indices [4, 5, 80, 142]
Polarity measures	
π_{ii}, π_{ij}	Self-atom and atom–atom polarizabilities [5]

(Continued)

Table 25.1 (Continued)

Symbol ^{a)}	Name/definition
$\alpha, \bar{\alpha}, \alpha$	Molecular polarizability [5, 55, 74, 105, 110, 114, 116, 125, 128, 136], average polarizability [5, 20–22, 108, 113, 139, 147], polarizability tensor [5]
β	First-order hyperpolarizability [122, 128]
γ	Second-order hyperpolarizability [68, 128]
μ (DM _T)	Total molecular dipole moment [5, 20–22, 55, 74, 75, 78, 82, 84, 86, 101, 108, 110, 114, 124, 125, 128, 136, 139, 146, 147]
D_x, D_y, D_z (DM _x , DM _y , DM _z)	Components of dipole moment along inertia axes [5, 82]
Θ (Q_{ii} , MQM)	Molecular quadrupole moment [20, 22, 98, 106, 107]
Φ	Average hexadecapole moment of a molecule [106]
Δ (SPP)	Submolecular polarity parameter, defined as the largest difference in electron charges between two atoms [5, 78]
APT	Atomic polar tensor [147]
Charges	
Q_A	Net atomic Mulliken charge at specific atom [5, 59, 86, 98, 105, 113, 115, 128, 136, 137, 140, 146, 147]
Q_A^+, Q_B^- ($Q_{\min}, Q_{\max}$ MPC, MNC)	The most positive and the most negative Mulliken atomic charges [5, 20–22, 55, 82, 84, 86, 101, 108, 109, 116, 124, 125, 128, 147]
Q_T, Q_A ($\sum Q_A$, SAC)	Sum of absolute values of charges of all atoms in a molecule or functional group [5, 82, 86, 101]
Q_T^2, Q_A^2 ($\sum Q_A^2$, SSC)	Sum of squares of charges of all atoms in a molecule or functional group [5, 82, 86, 101]
Q_m	Average of the absolute values of the charges on all atoms [5, 86, 101, 114]
Q_i	Electronic charge on the i -th atom [80, 142]
QTMS indices	
ρ	Electron density [46, 47, 49, 51, 52]
$\nabla^2 \rho$	Laplacian of electron density [46, 47, 49, 51, 52]
$\lambda_1, \lambda_2, \lambda_3$	Three Hessian eigenvalues [46, 49, 51, 52]
ε	Ellipticity of a bond at the bond critical point (BCP) – provides a measure of an extent to which charge is accumulated in a given plane [46, 47, 49, 51, 52]
$G(r)$	Lagrangian kinetic energy density [46, 49, 51, 52]
$K(r)$	Hamiltonian kinetic energy density [46, 47, 49, 51, 52]
Thermal properties ^{b)}	
E_{thermal}	Thermal energy [20–22, 105, 107, 109]
E_{int}	Internal energy [20]
C_v	Heat capacity at constant volume [20–22]
S	Entropy [20–22]

a) Possible variations of symbols adopted in the contemporary literature are given in parenthesis.

b) Thermal properties calculated at $T = 298.15$ K, $P = 1.00$ atm.

redundant; this was in agreement with other studies [64]. Carbó-Dorca and co-authors have also proposed several new molecular descriptors based on quantum similarity measures. These are (i) molecular quantum self-similarity measures (MQS-SMs) that were used to generate statistically significant QSAR models for steroids binding to corticosteroid-binding human globulin [29] as well as QSPR models for series of organic compounds [31] and (ii) an electron–electron repulsion energy descriptor, V_{ee} , tested on a widespread set of molecules as a complement to steric and electronic parameters in the description of molecular properties and bioresponses [65].

The QTMS approach takes advantage of the BCP space concept, providing topological similarity descriptors that are discrete distance-like measures defined in three or higher dimensional BCP space [7, 43, 44]. Some of the QTMS descriptors listed in Table 25.1 represent components of a so-called chemical vector that describes a bond in 3D BCP space (e.g., the electron density, ρ_b , the Laplacian of the electron density, $\nabla^2\rho$, and the ellipticity, ϵ); others characterize a bond by evaluating Lagrangian and Hamiltonian kinetic energy densities, $G(\mathbf{r})$ and $K(\mathbf{r})$, respectively [46–50]. In terms of chemical interpretation, it was shown that sometimes QTMS descriptors provide measures of σ (ρ_b and λ_3) and π ($\lambda_1 + \lambda_2$, ϵ) character of a bond or a simple measure of a covalent character of a bond ($\nabla^2\rho$). The ellipticity and Laplacian of electron density can also provide information about structural stability and local concentration of electronic charge, respectively [46, 51, 52]. QTMS descriptors have been used to build a wide variety of QSAR/QSPR models in medicinal and ecological chemistry [46], details of which are given in the publications of Popelier and co-authors (see also Chapter 24).

It is necessary to mention that QSAR studies in which quantum-chemical descriptors are obtained from calculations that account for the solvent effect are fairly rare [61, 66], due to the obvious reason of reducing computational time and costs, especially if it is assumed that the presence of solvent does not change significantly the geometrical and electronic characteristics of the molecule [67]. It has been noticed, however, that in certain cases this assumption is not valid [5]. In computational modeling of biologic macromolecules associated with rational drug-design, accounting for solvent can be crucial. Khandogin and York [66] have presented a set of descriptors for the characterization of macromolecules in solution that were obtained with modest computational cost using linear-scaling semiempirical methods combined with a conductor-like screening model (COSMO). The authors demonstrated the stability and convergence of derived descriptors and their applications to study several nucleocapsid proteins [66].

From Table 25.1 one can see that the total energy of the molecule, HOMO and LUMO energies, the HOMO–LUMO energy gap, the total molecular dipole moment, the molecular polarizability and Mulliken atomic charges can be ranked as the most frequently used electronic properties in the life science-related QSAR. The references cited in Table 25.1 refer readers to publications in which certain quantum-chemical descriptors were considered and then their “predictive power” evaluated by statistical methods. It is not always obvious at the beginning of a particular QSAR study what descriptors (quantum or classical) and in what amount

or combination must be selected. Two major approaches to the problem are worth mentioning. Some researchers choose the statistical or chemometric approach: they prefer starting from the entire pool of available descriptors and then perform computer-aided selection of significant descriptors before including them as variables into QSAR models [68–79]. Interestingly, when classical and quantum-chemical descriptors are combined in such an automated procedure, the chances of classical descriptors being selected for the final model are much higher than those for quantum-chemical descriptors. Every so often this fact makes it difficult or almost impossible to achieve a meaningful physicochemical interpretation of a predictive model. An alternative approach includes knowledge or experience-based initial selection, preferably of quantum-chemical (i.e., more interpretable) descriptors and then the possible addition of certain classical parameters to increase the accuracy and predictive ability of the final model, if needed [15, 55, 80–86]. However, the latter approach, which is better known from the historical perspective of QSAR, is sometimes considered as biased and it also may lead to an excessive amount of trial and error in the process of building predictive models when complex biologic phenomena are involved. The next section introduces the most commonly used and promising algorithms for the selection of descriptors, together with contemporary statistical methods.

25.3

Computational Approaches for Establishing Quantitative Structure–Activity Relationships

25.3.1

Selection of Descriptors

Contemporary QSAR has thousands of parameters available from experiment and *in silico* calculations that could potentially serve as independent variables (descriptors) in statistical analysis. It is already known, however, that utilization of an excessive number of descriptors leads to over-fitting of QSAR models and/or increases the risk of chance correlations. Despite the existence of “five golden rules” for building successful and meaningful QSAR models, formulated in 1973 by Hansch and Unger [87], the increasing complexity of biologic mechanisms on the one hand creates the need for considering a large variety of variables and, on the other, makes a knowledge-based approach to the identification of the most significant descriptors for a particular case extremely difficult. In his more recent review Kubinyi mentions the disadvantages of elimination procedures using forward, backward and stepwise regressions [1]. The conceptual description of mathematical approaches to the selection of descriptors that have shown to be efficient in generating optimal or near-optimal predictive models is presented below. A more general discussion of this topic, as well as of the process known as “feature selection” that allows identification and elimination of redundant or ineffective descriptors, can be found in a review of Nikolova and Jaworska [8] and the references therein.

Principal component analysis (PCA) performs reduction of data by generating linear combinations of original variables, that is, descriptors [57, 88–91]. The PCA method identifies correlated variables, groups them into linear combinations and generates an entire new set of uncorrelated (orthogonal) variables called principal components (PC). The direction of the first principal component (PC_1) is chosen to maximize the variance in the data; the next, PC_2 , is orthogonal to PC_1 and directed to maximize as much variance left as possible and so on. The process of data transformation is given by:

$$\mathbf{X} = \mathbf{TP}^T \quad (25.3)$$

where $\mathbf{X}$ represents the initial data matrix, $\mathbf{T}$ is a score matrix that defines the position of data points in a new coordinate system and $\mathbf{P}$ is a loadings matrix. The loadings indicate how much each original descriptor contributes to the corresponding PC. Scores and loadings allow the data points to be mapped into the new vector space defined by PCs [89, 91].

The decision tree method was employed by Smith *et al.* [92, 93] and later by Gubskaya *et al.* [79] to select the most significant descriptors for predicting the adsorption of fibrinogen onto polymer surfaces. Decision trees are usually used for description, classification and generalization of data [94]. The decision trees involved in the selection and analysis of descriptors classify data points by starting at the top of the tree (root node) and moving down the tree by creating a hierarchy of descriptor values on an “if-then-else” basis at each branch point until the terminal leaf (node) is reached. In these top-down constructions the data are recursively divided into subsets based upon the best classifying descriptors at each level (Figure 25.1). The C5 decision tree algorithm [95] employed by the authors evaluates the significance of each descriptor with respect to the set of experimental fibrinogen adsorption data using the concept of information gain introduced by Shannon [96, 97]. The conventional C5 algorithm was augmented by a Monte Carlo procedure to account for the experimental uncertainty [79]. All descriptors with the highest information gain were

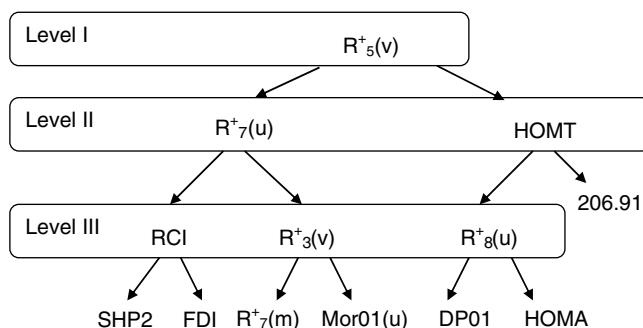


Figure 25.1 Schematic representation of C5 decision tree. The three most significant descriptors (level I and level II) were used to build an ANN model for predicting fibrinogen adsorption to polymeric surfaces [79].

extracted from thousands of Monte Carlo pseudo-experiments and summarized into a histogram. The three descriptors with the highest counts in this histogram were used as input variables to build QSAR models. It was shown that decision tree algorithms appear to be valuable tools for identifying the most relevant descriptors; however, so far they have only been used in the cases when classical descriptors are concerned. It would be beneficial from a methodological viewpoint to test the application of these algorithms on the QSAR models based exclusively on quantum-chemical properties whose biologic relevance to the particular bio-response was defined in advance by the knowledge-based approach.

A genetic algorithm (GA), a meta-heuristic method for the optimization of a function, was recommended by Kubinyi [1] as a method of choice for variable selection in QSAR [85, 98, 99]. The concept of the genetic algorithm is similar to that proposed in Darwin's theory of biologic evolution due to natural selection. In genetic algorithm terminology, an initial group ("population") of random organisms (sometimes called "chromosomes") evolves according to a fitness function that determines their survival. The algorithm searches for the "fittest" organisms through a selection, mutation and crossover genetic operation. "Genes" represent the properties of organisms; in the case of a feature selection these are descriptors. In other words, the method generates a set of potential solutions to a problem and then this set is iteratively modified and tested until an optimal solution is found. Genetic algorithms were successfully employed to select the most relevant descriptors from the large pools of variables containing various classes of classical and quantum-chemical descriptors [78, 100] as well as for the identification of the most significant characteristics from the relatively small sets obtained mainly by quantum simulations [76, 82, 98, 99, 101].

25.3.2

Linear Regression Techniques

Since 1964, the year the contemporary QSAR approach was born due to the contributions of C. Hansch, T. Fujita, S.M. Free and J.W. Wilson [102, 103], most QSAR models have been built using a multivariate regression technique. Regression analysis establishes a correlation between a dependent variable representing the biologic activity and multiple independent variables, that is, the descriptors (predictors). This correlation is most often expressed in the form of a multiple linear regression (MLR) equation as:

$$\text{activity} = \sum_i x_i \alpha_i + b \quad (25.4)$$

or, as in the case of a forward stepwise multiple regression (SMRL) technique [81, 98], as:

$$\text{activity} = \sum_i x_i (\alpha_i + \Delta \alpha_i) \quad (25.5)$$

where x_i denotes molecular descriptors, α_i and b are coefficients to be optimized, and $\Delta \alpha_i$ are coefficient errors. The best models are selected based on the correlation

coefficient (R or/and R^2), standard deviation (S) and the value F that represents the level of statistical significance of the model.

Clare and co-authors [59–61] have proposed a procedure called flip regression whereby MLR was applied to build QSAR models for benzene derivatives with each molecule independently in both possible orientations. For phenethylamine, for example, flipping includes rotating or reflecting the benzene ring in such a way that the charge and substituent values for two *o*- and two *m*-positions are interchanged. Then for N molecules 2^N regressions are generated and the one giving the best fit is selected [58].

The statement that “multiple regression is the most widely used mathematical technique in QSAR analysis” [2] is valid not only for all varieties of QSAR studies but also for those focused on the utilization of quantum-chemical descriptors [14, 20–22, 82, 104–115] and their combinations with other types of molecular characteristics [68, 70, 71, 76–78, 98–100, 116–123].

The partial least squares (PLS) regression method is the next multivariate linear regression technique commonly applied in QSAR modeling [51, 52, 67, 88, 124, 125]. It is often combined with principal component analysis or genetic algorithm to select the most appropriate input variables [52, 67, 85]. PLS establishes relationships between highly correlated input and output variables represented by arrays of data. It performs a reduction of the dimensionality of the raw data using both input and output data (i.e., $\mathbf{X}$ and $\mathbf{Y}$ matrices, respectively). Decomposition of $\mathbf{X}$ and $\mathbf{Y}$ is carried out simultaneously according to:

$$\mathbf{X} = \mathbf{TP} + \mathbf{D} \quad (25.6)$$

$$\mathbf{Y} = \mathbf{UQ} + \mathbf{F} \quad (25.7)$$

where $\mathbf{T}$ and $\mathbf{U}$ are the $\mathbf{X}$ - and $\mathbf{Y}$ -block score matrices, $\mathbf{P}$ and $\mathbf{Q}$ are the $\mathbf{X}$ and $\mathbf{Y}$ loadings, and $\mathbf{D}$ and $\mathbf{F}$ are residuals. A distinctive feature of this method is that it builds a regression model according to equation:

$$\mathbf{U} = \mathbf{BT} \quad (25.8)$$

with $\mathbf{U}$ and $\mathbf{T}$ representing scores (or projections) of dependent and independent variables, respectively. The N -way or multilinear partial least-squares method (N -PLS) is an extension of bilinear PLS designed for use in 3D-QSAR. N -PLS simultaneously decomposes and processes three-dimensional arrays of data such as GRID descriptors calculated by the comparative molecular field analysis (CoMFA) method or data generated by the quantum topological molecular similarity approach. Esteki *et al.* [51] have used both bilinear and N -PLS for QTMS indices-based QSAR modeling and prediction of acidity constant ($\text{p}K_a$) for some phenol derivatives.

25.3.3

Machine-Learning Algorithms

Since the late 1980s artificial intelligence methods have become an essential tool in QSAR analysis due to an increasing demand for accuracy and to the rapidly growing

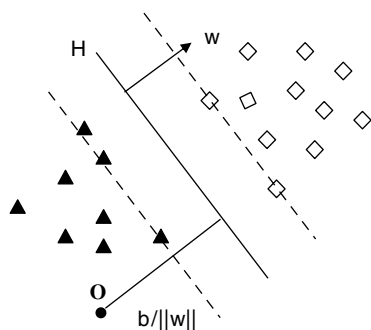


Figure 25.2 Schematic of SVM. The hyperplane H is shown by solid line and situated to maximize the margin, $d = 2/\|\mathbf{w}\|$ depicted by two dashed lines. Support vectors are located on the margin. Here $\mathbf{w}$ is the normal vector of hyperplane and $b/\|\mathbf{w}\|$ is its perpendicular distance from the origin.

number of SAR cases that exhibit highly nonlinear relationships. This section introduces the most promising SAR analysis nonlinear predictive methodologies and, when possible, their performance will be compared. In all cases mentioned here quantum-chemical descriptors were employed throughout or/and were found among the most significant variables responsible for accurate and meaningful correlations.

Not long ago a machine-learning technique called support vector machines (SVMs) became a part of the data analysis toolbox used for solving classification and regression problems in computational chemistry [126], specifically in drug-design [74, 83, 127], QSAR [73, 128], chemometrics and chemical engineering. The concept of SVM, originally introduced by Vapnik [129] and Lerner [130], is one of the major developments in statistical learning theory. Its principles can be briefly summarized as follows. The algorithm is designed to find an optimal hyperplane, H , between data points (i.e., all descriptors), $\mathbf{x}$, separating two distinct classes labeled as $y = 1$ and $y = -1$. The hyperplane, $\mathbf{w} \cdot \mathbf{x} + b$, is defined by its normal vector, $\mathbf{w}$, and the perpendicular distance from the origin $b/\|\mathbf{w}\|$ (Figure 25.2). The classification problem involves the optimization of Lagrangian multipliers α_i to generate a decision function $f(\mathbf{x})$ given by:

$$f(\mathbf{x}) = \text{sign} \left(\sum_{i=1}^l y_i \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b \right) \quad (25.9)$$

where $0 \leq \alpha_i \leq C$ and $\sum \alpha_i y_i = 0$ are constraints to be satisfied, C is a regularity parameter, $\mathbf{x}_i$ are support vectors (i.e., subset l of descriptors) and $K(\mathbf{x}, \mathbf{x}_i)$ is a Kernel function. A kernel function is defined in descriptor spaces of high dimensionality $\Phi(\mathbf{x})$ and $\Phi(\mathbf{x}_i)$ as $K(\mathbf{x}, \mathbf{x}_i) = \Phi(\mathbf{x}) \cdot \Phi(\mathbf{x}_i)$ and it can be computed without explicit use of $\Phi(\mathbf{x})$. To solve the regression problem, Equation 25.9 has to be rewritten as:

$$f(\mathbf{x}) = \sum_{i=1}^l y_i \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b, \quad (25.10)$$

where constraints are similar to those given for Equation (25.9). Several types of kernel functions are known [126]: linear, polynomial, Gaussian, exponential radial basis function (RBF) kernel and so on. The kernel function and the regularity parameter C are the only features that specify the SVM algorithm for a given data set. The Gaussian kernel is the most commonly used for support vector classification and regression. In addition to the conventional SVM algorithm, a few of its modifications (such as least squares (LS) SVM [131] and v -SVM [132]) were recently employed in medicinal chemistry to build successful QSAR models [83, 127].

Zheng *et al.* [128] reported predictions of receptor relative binding affinities for polybrominated diphenyl ethers using SVM and radial basis function neural network (RBFNN). After correlating relative binding affinities with seven of the most significant quantum-chemical descriptors, the authors concluded that SVM models generalize better than RBFNN models [128]. A similar conclusion was drawn by Liu *et al.* [73]; their analysis of classification models built with SVM and the k -nearest neighbor (k_{NN}) method for a novel series of cyclooxygenase selective inhibitors revealed that the performance of the k_{NN} algorithm is much less satisfactory than that of SVM. Comparison of nonlinear and linear QSAR models for a set of pyrazine-pyridine biheteroaryls, inhibitors of vascular endothelial growth factor receptor-2, showed that nonlinear models obtained using LS-SVMs perform better in terms of generalization and predictive ability than the multiple linear regression model [127]. Interestingly, the SVM and cluster-genetic algorithm-partial least squares discriminant analysis [133] classification models generated to predict chemical metabolism by human UDP-glucuronosyltransferase isoforms [83] demonstrated very similar and somewhat unsatisfactory performance with two-dimensional (2D) chemical descriptors employed. The authors applied an electronegativity equalization method (EEM) [134, 135] developed for the fast calculations of DFT-based molecular and atomic properties to compute quantum-chemical descriptors. Combination of 2D and EEM-derived descriptors using the so-called consensus approach made it possible to achieve significant improvement (up to 84%) of overall substrate and non-substrate predictability [83].

Artificial neural network (ANNs) are machine-learning algorithms that like SVM can be used to handle classification and regression problems. Classifying ANN estimates the so-called membership functions and doing so translates continuously changing output value into discrete nominal categories [84]. Such a procedure increases the robustness of the algorithm, simplifies and speeds up the training process and provides some measures of similarity between investigated objects (i.e., molecules). Regression ANN handles a prediction problem by estimating the value of a continuous output variable that is calculated for the known values of input parameters. It requires higher accuracy and therefore increases the time of the training process and the number of learning examples [84]. Additionally, regression ANN models must demonstrate extrapolation capabilities that, in contrast to radial basis function networks, are associated with the multilayer perceptron-type of NN architectures. Multilayer perceptrons are feed-forward, multilayer networks designed to provide adaptable frameworks for nonlinear function estimation. They consist of neurons or nodes arranged in layers: an input layer, one or more hidden

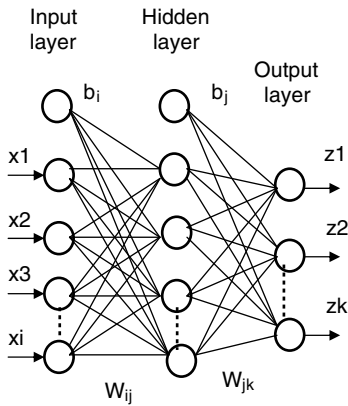


Figure 25.3 Structure of ANN represented by a two-layer perceptron with multi-output node, where the input variables and output values are denoted as x and z , respectively; b_i and b_j stay for biases; w_{ij} and w_{jk} are the connection weights between input/output, respectively, and a hidden nodes.

layers and an output layer. Multiple connections between neurons of adjacent layers are unidirectional (i.e., from input to output) and reflected in the connection weights (Figure 25.3). First, input neurons distribute the initial input variables (x_i) among hidden layer neurons without additional computations. Then the hidden layer variables (y_j) are calculated as:

$$y_j = f \left(\sum_i x_i w_{ij} + b_i \right) \quad (25.11)$$

where w_{ij} are the connection weights between input and hidden nodes, b_i is the bias of input layer, and $f(\varsigma) = (1 + e^{-k\varsigma})^{-1}$ is a sigmoid transfer function that varies between 0 and 1 with coefficient k to be user-specified. Finally, the output values (z_k) are computed as:

$$z_k = \sum_j y_j w_{jk} + b_j \quad (25.12)$$

where w_{jk} and b_j are the weights and the bias for the output layer, respectively. The value E represents the target error defined as:

$$E = \frac{1}{2} \sum_{k=0}^{s-1} (z_k - \hat{z}_k)^2 \quad (25.13)$$

where z_k and $\hat{z}_k$ are the predicted and experimentally determined output values, respectively.

Subsequent minimization of E modifies the connection weights to achieve the best fit.

In addition to conventional feed-forward (or as they are sometimes called, back-propagation) ANN algorithms [69, 82, 84, 91, 105, 136–139] and their variations, such as projection pursuit regression models [140], several alternative designs, namely, counter-propagation ANN architecture [141], Bayesian-regularized [101] and fuzzy ARTMAP [139] neural networks, were explored to handle complex quantitative relationships in biology-related systems. Counter-propagation ANNs are based on an unsupervised learning strategy that does not require a priori knowledge about the process to be modeled. Although this type of algorithm was not extensively used in QSAR, Novic *et al.* [141] used it to study the correlation of inhibitory activity of 105 flavonoid inhibitors of the enzyme p56^{lck} protein tyrosine kinase and proved that its results compare favorably with those obtained by classical back-propagation methods. The Bayesian regularization (BR) approach, instead of looking for global minima, finds locally the most probable network parameters, estimates their effectiveness and produces predictors that are reliable and well-matched to the data. Multiple linear regression model and BRANN combined with feature selection by genetic algorithm were recently utilized to model the biologic activity of luteinizing hormone-releasing hormone antagonists [101]. The better results achieved by the GA-BRANN suggest that nonlinear analysis is more suitable to model complex SAR of these non-peptide antagonists [101]. The fuzzy ARTMAP-based model was developed to construct log K_{ow} QSPR for 442 compounds and it demonstrated clear superiority with respect to back propagation ANN and multiple linear regression QSPR models, due to its ability to handle noisy data with fuzzy logic and to avoid the plasticity–stability dilemma of standard back propagation architectures [139].

There is a common belief that because SVM is based on structural risk minimization its predictions are better than those of other algorithms [126]. It was also suggested that SVMs perform better than ANNs because they provide a unique solution, whereas ANNs can become trapped in local minima and the optimum number of hidden neurons of ANN sometimes requires time-consuming computations. It is known, however, that ANNs can provide constantly good suboptimal solutions and that this is not always the case for SVMs. Thus, only an empirical comparison of results obtained by various machine-learning algorithms can demonstrate, for a particular SAR problem, the superiority of one method over the other [126].

25.4

Quantum-Chemical Descriptors in QSAR/QSPR Models

25.4.1

Biochemistry and Molecular Biology

Just over a decade ago, Karelson *et al.* [5] emphasized that quantum-chemical descriptors have a long history of use in QSAR studies in biochemistry and also indicated that a trend of such descriptors as HOMO–LUMO energies, frontier orbital

electron densities and superdelocalizabilities correlates well with various biologic activities. In the present section several examples of the most successful QSAR models will be shown and discussed without providing specific details of correlation equations, for which readers are referred to the original publications.

The way in which the contemporary QSAR approach pursues its twofold goal of understanding the chemical mechanisms associated with biochemical effects and providing practical recommendations for optimal molecular design can be demonstrated in investigations of inhibitory activity. The potency of the intermediate conductance Ca^{2+} -activated K^{+} channel (IKCa1) blockade by triarylmethane (TRAM) derivatives has been studied by Fernández and coworker [86]. The results showed that *ab initio* derived electronic properties in combination with topological (WHIM) descriptors are important parameters influencing the binding of TRAMs with IKCa1: global quantum-chemical properties (i.e., molecular dipole moments, HOMO and LUMO energies) describe the electronic environment of the molecules and local parameters (i.e., local Mulliken charges) help to locate molecular regions responsible for given bioactivity [86]. Analogous behavior in terms of selected descriptors tendency was observed by Safarpour *et al.* [82], who conducted QSAR analysis on the Ca^{2+} channel antagonist activity of some newly synthesized 1,4-dihydropyridine derivatives. From the six descriptors selected for the multiple linear regression model, five were related to the electronic (i.e., electrophilicity, electronegativity and dipole moment) or physicochemical (surface area and molar refractivity) properties of the whole molecules and one (electrostatic potential) described electronic properties of individual atoms [82]. Inhibition activity of flavonoids was investigated based on their ability to inhibit replication of the human immunodeficiency virus (HIV) [81] and to maintain the balance between neuronal excitation and inhibition in the central nervous system (CNS) by binding to the γ -aminobutyric acid type A [GABA(A)] receptor [77]. In the former case it was concluded that HIV-inhibitory activity is mainly the outcome of electronic interactions between atomic charges within flavonoids and possible receptor-like structures in the HIV or the lymphocyte itself. In the latter case it was shown that the binding affinities of selected flavonoids to GABA(A) receptor are highly dependent on conformational changes involved in the interactions [77]; in other words, no electronic properties were found to be significant after exploring the entire pool of 1176 classical and electronic variables. This example, however, appears to be an exception: the results of many other researchers clearly indicate a trend of preferable utilization of electronic properties in QSAR studies of inhibition activity [59–61, 85, 112–114].

Special types of inhibitory activity, such as the inhibition of parasite or bacterial growth, have been studied by several research groups [28, 71, 110, 142]. Katritzky and co-authors [71] reported QSAR models obtained using the B(est)MLR method and descriptors calculated with CODESSA PRO software [143] for two diverse sets of potentially active compounds against the D6 and NF54 strains of malaria. The mechanism of antimalarial activity showed the significance of charge-related interactions as well as the shape and branching of the molecules [71]. Utilization of kinetic energy DF similarity measures as descriptors by Gironés *et al.* [28] also led

to satisfactory correlations for all antimalarial activities in all studied molecular sets. Molina and co-authors [142] have developed a new QSAR strategy that includes a linear piecewise regression and discriminant analysis and examines the possibility of combining different types of molecular descriptors. Their best models for inhibition of respiration in *Escherichia coli* by 2-furylethylenes included only quantum-chemical descriptors such as local charges and electrophilic and nucleophilic superdelocalizabilities [142].

QSAR/QSPR approaches have been used over the years to develop models capable of estimating or predicting the toxic potency of various compounds ranging from small molecules to biosystems. The most recent achievements in this area include QSAR modeling of toxicity of various aromatic compounds [48, 49, 68, 100, 121, 124, 125] and halocarbons [15] to develop efficient and inexpensive methods for the estimation of their effects on human and environmental health. The quantum-chemical descriptors that quantify the electrophilic nature of the molecules were identified as premier quantities in the prediction of toxicity [15, 48, 100, 121, 124]. Again, in some cases quantum-chemical and classical descriptors were combined to fortify models obtained solely with classical descriptors [15, 68, 121].

The chemicals, which are known as potential carcinogens, have been in the focus of the scientific community for at least 70 years. Among these, polycyclic aromatic hydrocarbons take a special place: some of them are ranked as the strongest carcinogens, but some are identified as inactive. Several research groups have investigated the relationship between carcinogenic activity and the chemical structure of PAHs using quantum-mechanical calculations [56, 57, 124, 144, 145]. The latest studies include the work of Barone *et al.* [56], who proposed a new methodology, EIM, that utilizes electronic descriptors (Section 25.2.2), as well as the work of Vendrame *et al.* [57], who performed a comparative study of EIM and PCA-ANN approaches and confirmed that the key descriptors in EIM are indeed the relevant descriptors for classifying the carcinogenic activity of PAHs. In addition, Lu *et al.* [124] have reported a QSPR model for water solubility of PAHs obtained using electronic descriptors calculated at the B3LYP/6-31G(d) level and PLS statistics. Their results demonstrate the superiority of electronic parameters in comparison with known physicochemical properties and/or topological indices.

25.4.2

Medicinal Chemistry and Drug Design

In modern QSAR the distinction between biochemistry and molecular biology at one end and medicinal chemistry and drug-design at the other is somewhat artificial: in all these fields, biologic processes and phenomena are the matters of primary concern. Thus the present section will focus on examples of practical interest to medicine and pharmacology.

The blood–brain barrier (BBB) is represented by a complex cellular system that maintains the homeostasis of the CNS by separating the brain from systemic blood circulation. The ability of a drug to penetrate the BBB is of utmost importance

in the design of neurological drugs: CNS-active drugs require high penetration, while it is more desirable for drugs intended for peripheral activity to exhibit minimal penetration to prevent CNS-related side effects [118]. To build a QSAR model for the brain/plasma partition coefficient, $\log (C_{\text{brain}}/C_{\text{blood}})$ or $\log BB$, Hutter employed variables from the AM1 optimized geometry of 90 compounds [119]. The electrostatic potential and, related to it, the set of variables that accounts for the polarity of molecular surfaces were identified as the most significant descriptors of his 12-term MLR-based model [119]. Van Damme *et al.* [118] have developed and presented a new *in silico* model to predict $\log BB$, for a set of 82 structurally diverse compounds using a combination of classical and quantum-chemical descriptors. The final eight-parameter model among the others included several Mulliken charge-related descriptors and the dipole moment, confirming the known fact that non-polar molecules cross the barrier more easily than polar molecules [118]. Among a multitude of CNS-related drugs, valproic acid (VPA) is an established antiepileptic drug, which is generally well tolerated but has two serious potential side effects: hepatotoxicity and teratogenicity [98]. In an attempt to find a superior compound that would retain the anticonvulsant activity of the basic structure of VPA but would not cause the adverse side effects, Hashemianzadeh *et al.* [98] performed a QSAR study of 25 potent VPA derivatives utilizing DFT calculations and QTAIM to obtain quantum-chemical descriptors. Their statistically significant MLR model (with a correlation coefficient of 0.937) suggested that polarizability and electrostatic potential at certain carbon atoms of the drugs are strongly correlated with antiepileptic activity of these types of VPA derivatives [98].

Electronic descriptors were the major descriptors employed in QSAR studies of the anesthetic action of some polyhalogenated ethers [116] and the antioxidant properties of phenolic compounds [104]. Mehdipour *et al.* [116] reported an MLR equation (with correlation coefficient of 0.97) that clearly indicated the significant effects of coulombic (i.e., electrostatic potential and most positive charge descriptors), steric and polar interactions (molecular polarizability) as well as lipophilicity ($\log P$) on the anesthetic activity of the polyhalogenated ethers. Reis *et al.* [104] employed and compared the predictive quality of quantum-chemical descriptors potentially relevant to antioxidant activity that were calculated at AM1, PM3, HF and B3LYP levels of theory for 41 phenolic compounds. The best regression equations included E_{HOMO} , vertical ionization potential, IP_v , and charge on oxygen atom, Q_O . These descriptors, obtained at both HF and DFT/B3LYP levels, revealed that low values of $IP_{\text{VHOMO(DFT)}}$ (from Koopman's theorem) combined with negative charges on O7 lead to an increase in the antioxidant activity [104]. The antioxidative nature of hydroxyphenylureas has been investigated by Deeb *et al.* [120]; their eight parametric model consisted of five topological and three quantum-chemical descriptors: Q_m , Q_{max} and μ .

As mentioned in Section 25.2.2, QSAR studies that combine both classical and quantum-chemical descriptors are becoming more common. Alvarez-Ginatre and co-authors [76, 99] reported several predictive QSAR models of anabolic and androgenic activities for selected steroid analogs. In this study the combination of electronic and physicochemical descriptors helped to identify molecular shape,

hydrophobicity and electronic properties as three major factors responsible for these types of steroid activity. A similar approach to the initial selection of QSAR descriptors was employed for the discovery of less toxic, more selective and more effective agents to treat [55, 69, 123, 127] and prevent [73] cancer.

25.4.3

Material and Biomaterial Science

The use of QSAR/QSPR in the development of biomaterials and, in particular, materials for tissue-engineering applications is relatively new. The recent success of DFT in the accurate determination of electronic properties of biologically significant molecules has initiated QSPR studies in material science that focus specifically on the design of polymeric (bio)materials. DFT provides an invaluable tool for calculating quantum-chemical descriptors that demonstrate high potential in generating predictive QSPR models without the addition of classical descriptors for various classes of polymers represented by their repeat units. Yu, Gao, Liu and co-authors [20–22, 105–109, 136, 146, 147] have published a series of articles devoted to the prediction of the most important physicochemical properties of polymers, namely, refractive index [21], dielectric constant [109], glass transition temperature [105, 106, 109], melting point [136], cohesive energy [108], molar stiffness function (i.e., conformational property) [22], thermal decomposition [107] and reactivity parameters in free-radical polymerization [147]. It is somewhat surprising that no attempts to build QSAR models for prediction of biologic response (e.g., polymer adsorption [79, 92, 93] onto or cellular proliferation [89, 148] to the polymer surfaces) using quantum-chemical descriptors have been reported by the time of writing this chapter.

The conformational properties of polymer chains play a key role in synthesis (i.e., polymerization in solution), processing (i.e., solvent casting on thin films) and bioresponse onto the surfaces of biocompatible and biodegradable polymers. The molar stiffness function, K , is directly related to the intrinsic viscosity of polymer solutions and can be estimated using QSPR models. Yu *et al.* [22] used three quantum-chemical descriptors (quadrupole moment, thermal and total energies) to predict the molar stiffness function K for 47 vinyl polymers. A physically meaningful QSPR model with correlation coefficient 0.958 and mean error 7.1% was generated using stepwise MLR analysis. Yu and co-authors [147] also reported an accurate QSPR model (root mean square errors of 0.37 and 0.19, respectively) using back propagation ANN for predicting the reactivity parameters Q and e in radical copolymerization of vinyl monomers. The authors showed that the Mulliken charges and frontier molecular orbital energies are the descriptors most correlated with reactivity parameters [147].

An interesting comparison can be made between two independently generated QSAR models to predict the refractive index for the same representative set of polymers previously investigated by Bicerano [149]. Holder *et al.* [117] used CODESSA program [143] to calculate a total of 600 (classical and quantum-chemical) descriptors for each of the 60 polymers represented by dimer models. Quantum

Table 25.2 Comparison of QSAR models for T_g of polymeric materials.

Significant descriptors	Methods	Correlation coefficient	Class of polymers	References
PMA, ^{a)} M, ^{a)} U, E_{LUMO} , Q_{O} Θ , Φ	DFT	0.889, 0.898	Polyamides	Gao <i>et al.</i> [146]
	MLR, ANN		Polyvinyls, polyethylenes and polymethacrylates	Yu <i>et al.</i> [106]
	DFT			
L-1.356 ^{a)} E_{thermal} , α , Q_{C} E_{thermal} , E_{HF}	MLR	0.952	Polymethacrylates	Liu <i>et al.</i> [105]
	DFT	0.960, 0.991		
	MLR, ANN		Polyalkanes and polyacrylamides	Liu <i>et al.</i> [109]
	DFT			
	MLR	0.921		

a) Classical descriptors.

descriptors were obtained by the AM1 method. The final QSPR model with a correlation coefficient of 0.953 was obtained by MLR and featured two electronic descriptors: HOMO–LUMO gap and a polarizability index. Yu *et al.* [21] used refractive index data for 95 polymers from the same dataset and calculated only ten quantum-chemical descriptors for given monomers using DFT at the B3LYP/6-31G (d) level. In this case four descriptors (LUMO energy, molecular polarizability, heat capacity at constant volume and the most positive charge on hydrogen atom) were selected to build optimal QSPR models by means of stepwise MLR (correlation coefficient 0.926). These independently obtained and comparable results clearly identify the main electronic parameters that affect the values of refractive index for vinyl polymers.

The glass transition temperature, T_g , is one of the most important characteristics of amorphous polymers. Table 25.2 summarizes the results of several QSPR studies devoted to the prediction of T_g for various classes of polymers. There have been attempts in the past decade to evaluate the usefulness of different classes of descriptors in establishing correlations for certain types of activities/properties by comparing their statistical performance separately or in combination using the same or different data sets [150]. Clearly, from Table 25.2 (as well as from the previous example), it is impossible to provide any reliable generalizations in this regard and this fact can be used as the main argument against broad utilization of knowledge-based initial selection of descriptors.

25.5

Summary and Conclusions

Since QSAR/QSPR has been established as a methodology that allows one to estimate the properties of chemicals at a much lower cost than that of actual laboratory screening, it has been widely applied in all chemistry-related fields of the life

sciences. The necessity to produce robust and reliable QSAR models, the results of which would be of immense practical value, has led to the development of the "Guidance document on the validation of (Q)SAR models" that was recently introduced by a group of experts from the Organization for Economic Co-operation and Development [151]. The principles formulated in this document clearly reflect the original goal of QSAR to be a reliable predictive tool and to provide, when possible, mechanistic interpretation(s) of the model. Both the predictive ability of the model and the scientific insights into the mechanisms involved in a particular kind of biologic activity depend on the descriptors selected in the modeling process.

The use of quantum-chemical descriptors has an obvious advantage over other calculated or experimentally measured properties because they are reproducible (in the framework of the chosen approximation), they allow meaningful interpretation of QSAR models in terms of the mechanism of actions, metabolic or toxicological routes and, thus, can offer clear guidance for molecule optimization or design. In some cases it becomes necessary to consider both classical and quantum-chemical descriptors as potential candidates for successful QSAR modeling that sometimes dramatically increases the number of input variables. Among all available tools for automatic selection of descriptors, genetic algorithms have shown to be the most promising.

Rapid developments in combinatorial synthesis have facilitated the production of large amounts of chemical compounds whose activity cannot be easily estimated even by modern high-throughput screening techniques. The utilization of machine-learning methodologies such as artificial neural nets and support vector machines provides contemporary QSAR reliable ways of handling nonlinear statistics and noticeably increases accuracy and predictive power not only of large industrial-scale models but also of models generated for local scientific or testing purposes.

One of the major reasons for the acceleration of the use of quantum-chemical descriptors in QSAR modeling has been the development of DFT. Both relatively low costs and reasonable accuracy have led to the successful utilization of DFT for the calculation of a broad range of properties for larger molecules. Many researchers have confirmed the superior performance of DFT in comparison with semiempirical calculations. The concept of quantum similarity has stimulated the development of various theories and algorithms that have given rise to a new generation of descriptors known as quantum similarity measures. These descriptors are rapidly becoming competitive with conventional electronic parameters in solving QSAR problems in life science-related fields. Similar to the previous decade, the recent applications of quantum-chemical descriptors include prediction of inhibitory activity, chronic and acute toxicity, ligand–receptor binding affinity, antimicrobial activity, carcinogenesis and mutagenesis. Successful utilization of DFT-derived electronic parameters in the prediction of various properties of polymers has created the basis for *in silico* design of new biomaterials. Clearly, quantum-chemical descriptors have significant applicability potential in traditional areas of QSAR and their applications in novel and rapidly growing fields of biomedical science has yet to be explored.

Abbreviations

AIM	atoms in molecules (theory)
AM1	Austin model 1
ANN	artificial neural networks
BCP	bond critical point
BRANN	Bayesian regularization ANN
CNS	central nervous system
DF	density function
DFT	density functional theory
EIM	electronic indices methodology
GA	genetic algorithm
HF	Hartree–Fock
HOMO	highest occupied molecular orbital
LUMO	lowest unoccupied molecular orbital
MLR	multiple linear regression
MNDO	modified neglect of differential overlap
MP	Møller–Plesset
MQSM	molecular quantum similarity measure
PAH	polycyclic aromatic hydrocarbons
PCA	principal component analysis
PLS	partial least-squares
PM3	parametric model 3
QSAR/QSPR	quantitative structure activity/property relationship
QTMS	quantum topological molecular similarity
RBFNN	radial basis function neural network
SVM	support vector machines
WHIM	weighted holistic invariant molecular (descriptors)

Acknowledgments

The author is thankful to Professor C. F. Matta for the invitation to contribute a chapter to this book and his useful comments on its content. The constructive criticism and professional remarks of Drs Y. V. Lisnyak and V. Kholodovych are also deeply appreciated.

References

- 1 Kubinyi, H. (2002) *Quantum Struct.-Act. Relat.*, **21**, 348–356.
- 2 Selassie, C.D. (2003) *Burger's Medicinal Chemistry and Drug Discovery*, 6th edn, vol. 1 (ed. D.J. Abraham), John Wiley & Sons, Inc., New York.
- 3 Gramatica, P., Sumathy, K.V.C., and Suraj, S. (eds) (2008) A Strand Life Sciences Web Resource. http://www.qsarworld.com/qsar_archives.php, (September 2008).
- 4 Todeschini, R., Mannhold, R., Kubinyi, H., Consonni, V., and Timmerman, H.

- (2000) *Handbook of Molecular Descriptors*, John Wiley & Sons, Inc., New York.
- 5 Karelson, M., Lobanov, V.S., and Katritzky, A.R. (1996) *Chem. Rev.*, **96**, 1027–1043.
- 6 Parr, R.G. and Yang, W. (1989) *Density Functional Theory of Atoms and Molecules*, Oxford University Press, Oxford.
- 7 Koch, W. and Holthausen, M.C. (2001) *A Chemist's Guide to Density Functional Theory*, 2nd edn, Wiley-VCH Verlag GmbH, Weinheim.
- 8 Nikolova, N. and Jaworska, J. (2003) *QSAR Comb. Sci.*, **22**, 1006–1026.
- 9 Dewar, M.J.S., Zoebisch, E.G., Healy, E.F., and Stewart, J.J.P. (1985) *J. Am. Chem. Soc.*, **107**, 3902–3909.
- 10 Dewar, M.J.S. and Thiel, W. (1977) *J. Am. Chem. Soc.*, **99**, 4899–4907.
- 11 Stewart, J.J.P. (1989) *J. Comput. Chem.*, **10**, 209–220.
- 12 Levine, I.N. (1991) *Quantum Chemistry*, Prentice Hall, Englewood Cliffs, New Jersey.
- 13 Braga, S.F. and Galvão, D.S. (2004) *J. Chem. Inf. Comput. Sci.*, **44**, 1987–1997.
- 14 Trohalaki, S., Gifford, E., and Pachter, R. (2000) *J. Comput. Chem.*, **24**, 421–427.
- 15 Basak, S.C., Balasubramanian, K., Gute, B.D., Mills, D., Gorczynska, A., and Roszak, S. (2003) *J. Chem. Inf. Comput. Sci.*, **43**, 1103–1109.
- 16 Lee, C., Yang, W., and Parr, R.G. (1988) *Phys. Rev. B*, **37**, 785–789.
- 17 Becke, A.D. (1993) *J. Chem. Phys.*, **98**, 1372–1377.
- 18 Becke, A.D. (1993) *J. Chem. Phys.*, **98**, 5648–5652.
- 19 Wang, Z.-Y., Zhai, Z.-C., and Wang, L.-S. (2005) *QSAR Comb. Sci.*, **24**, 211–217.
- 20 Yu, X., Wang, X., Gao, J., Li, X., and Wang, H. (2005) *Polymer*, **46**, 9443–9451.
- 21 Yu, X., Yi, B., and Wang, X. (2007) *J. Comput. Chem.*, **28**, 2336–2341.
- 22 Yu, X., Yi, B., Xie, Z., Wang, X., and Liu, F. (2007) *Chemom. Intell. Lab. Syst.*, **87**, 247–251.
- 23 Hobza, P. and Syponer, J. (1999) *Chem. Rev.*, **99**, 3247–3276.
- 24 Gironés, X. and Carbó-Dorca, R. (2006) *QSAR Comb. Sci.*, **25**, 579–589.
- 25 Carbó-Dorca, R. (2007) *SAR QSAR Environ. Res.*, **18**, 265–284.
- 26 Gallegos, A., Robert, D., Gironés, X., and Carbó-Dorca, R. (2001) *J. Comput. Aided Mol. Des.*, **15**, 67–80.
- 27 Carbó-Dorca, R., Robert, D., Amat, L., Gironés, X., and Besalú, E. (2000) *Molecular Quantum Similarity in QSAR and Drug Design*, Springer Verlag, Berlin.
- 28 Gironés, X., Gallegos, A., and Carbó-Dorca, R. (2000) *J. Chem. Inf. Comput. Sci.*, **40**, 1400–1407.
- 29 Amat, L., Besalú, E., and Carbó-Dorca, R. (2001) *J. Chem. Inf. Comput. Sci.*, **41**, 978–991.
- 30 Lobato, M., Amat, L., Besalú, E., and Carbó-Dorca, R. (1997) *Quant. Struct.-Act. Relat.*, **16**, 465–472.
- 31 Poncet, R., Amat, L., and Carbó-Dorca, R. (1999) *J. Comput. Aided Mol. Des.*, **13**, 259–270.
- 32 Mezey, P.G. (1993) *Shape in Chemistry: An Introduction to Molecular Shape and Topology*, VCH Publishers, New York.
- 33 Mezey, P.G. (1996) *J. Chem. Inf. Comput. Sci.*, **36**, 1076–1081.
- 34 Mezey, P.G., Zimpel, Z., Warburton, P., Walker, P.D., Irvine, D.G., Dixon, D.G., and Greenberg, B. (1996) *J. Chem. Inf. Comput. Sci.*, **36**, 602–611.
- 35 Mezey, P.G. (1999) *Mol. Phys.*, **96**, 169–178.
- 36 Riess, I. and Münch, W. (1981) *Theor. Chim. Acta*, **58**, 295–300.
- 37 Matta, C.F. (2001) *J. Phys. Chem.*, **105**, 11088–11101.
- 38 Bader, R.F.W. (1990) *Atoms in Molecules: A Quantum Theory*, Oxford University Press, Oxford, UK.
- 39 Matta, C.F. and Bader, R.F.W. (2000) *Proteins Struct., Funct., Genet.*, **40**, 310–329.
- 40 Matta, C.F. and Bader, R.F.W. (2002) *Proteins Struct., Funct., Genet.*, **48**, 519–538.
- 41 Matta, C.F. and Bader, R.F.W. (2003) *Proteins Struct., Funct., Genet.*, **52**, 360–399.
- 42 Matta, C.F. and Boyd, R.J. (eds) (2007) *The Quantum Theory of Atoms in Molecules: From Solid state to DNA and Drug Design*, Wiley-VCH Verlag GmbH, Weinheim.
- 43 Popelier, P.L.A. (1999) *J. Phys. Chem.*, **103**, 2883–2890.
- 44 O'Brien, S.E. and Popelier, P.L.A. (1999) *Can. J. Chem.*, **77**, 28–36.
- 45 O'Brien, S.E. and Popelier, P.L.A. (2001) *J. Chem. Inf. Comput. Sci.*, **41**, 764–775.
- 46 Popelier, P.L.A., Smith, P.J., and Chaudry, U.A. (2004) *J. Comput. Aided Mol. Des.*, **18**, 709–718.

- 47 Loader, R.J., Singh, N., O'Malley, P.J., and Popelier, P.L.A. (2006) *Bioorg. Med. Chem. Lett.*, **16**, 1249–1254.
- 48 Roy, D.R., Parthasarathi, R., Subramanian, V., and Chattaraj, P.K. (2006) *QSAR Comb. Sci.*, **25**, 114–122.
- 49 Roy, K. and Popelier, P.L.A. (2008) *Bioorg. Med. Chem. Lett.*, **18**, 2604–2609.
- 50 Roy, K. and Popelier, P.L.A. (2008) *QSAR Comb. Sci.*, **27**, 1006–1012.
- 51 Esteki, M., Hemmateenejad, B., Khayamian, T., and Mohajeri, A. (2007) *Chem. Biol. Drug Des.*, **70**, 413–423.
- 52 Mohajeri, A., Hemmateenejad, B., Mehdipour, A., and Miri, R. (2008) *J. Mol. Graphics Modell.*, **26**, 1057–1065.
- 53 Jezierska, A., Panek, J., Ryng, S., Glowiak, T., and Koll, A. (2003) *J. Mol. Model.*, **9**, 159–163.
- 54 Mazza, C.B., Sukumar, N., Breneman, C.M., and Cramer, S.M. (2001) *Anal. Chem.*, **73**, 5457–5461.
- 55 Yao, S.-W., Lopes, V.H.C., Fernández, F., and García-Mera, X. (2003) *Bioorg. Med. Chem.*, **11**, 4999–5006.
- 56 Barone, P.M.V.B., Camilo, A.J., and Galvão, D.S. (1996) *Phys. Rev. Lett.*, **77**, 1186–1189.
- 57 Vendrame, R., Braga, R.S., Takahata, Y., and Galvão, D.S. (1999) *J. Chem. Inf. Comput. Sci.*, **39**, 1094–1104.
- 58 Clare, B.W. (2002) *J. Comput. Aided Mol. Des.*, **16**, 611–633.
- 59 Clare, B.W. and Supuran, C.T. (2005) *Bioorg. Med. Chem.*, **13**, 2197–2211.
- 60 Deeb, O. and Clare, B.W. (2007) *Chem. Biol. Drug Des.*, **70**, 437–449.
- 61 Deeb, O. and Clare, B.W. (2008) *Chem. Biol. Drug Des.*, **71**, 352–362.
- 62 Bowen-Jenkins, P.E. and Richards, W.G. (1986) *Int. J. Quantum Chem.*, **30**, 763–768.
- 63 Boon, G., Langenaeker, W., De Proft, F., De Winter, H., Tollenaere, J.P., and Geerlings, P. (2001) *J. Phys. Chem. A*, **105**, 8805–8814.
- 64 Bultinck, P. and Carbó-Dorca, R. (2005) *J. Chem. Sci.*, **117**, 425–435.
- 65 Gironés, X., Amat, L., Robert, D., and Carbó-Dorca, R. (2000) *J. Comput. Aided Mol. Des.*, **14**, 477–485.
- 66 Khandogin, J. and York, D.M. (2004) *Proteins Struct., Funct., Bioinf.*, **56**, 724–732.
- 67 Stenber, P., Norinder, U., Luthman, K., and Artursson, P. (2001) *J. Med. Chem.*, **44**, 1927–1937.
- 68 Maran, U., Karelson, M., and Katritzky, A.R. (1999) *QSAR Comb. Sci.*, **19**, 3–10.
- 69 Katritzky, A.R., Dobchev, D.A., Fara, D.C., and Karelson, M. (2005) *Bioorg. Med. Chem.*, **13**, 6598–6608.
- 70 Katritzky, A.R., Dobchev, D.A., Hür, E., Fara, D.C., and Karelson, M. (2005) *Bioorg. Med. Chem.*, **13**, 1623–1632.
- 71 Katritzky, A.R., Kulshyn, O.V., Stoyanova-Slavova, I., Dobchev, D.A., Kuanar, M., Fara, D.C., and Karelson, M. (2006) *Bioorg. Med. Chem.*, **14**, 2333–2357.
- 72 Katritzky, A.R., Pacureanu, L.M., Dobchev, D.A., Fara, D.C., Duchowicz, P.R., and Karelson, M. (2006) *Bioorg. Med. Chem.*, **14**, 4987–5002.
- 73 Liu, H.X., Zhang, R.S., Yao, X.J., Liu, M.C., Hu, Z.D., and Fan, B.T. (2004) *J. Comput. Aided Mol. Des.*, **18**, 389–399.
- 74 Kriegl, J.M., Arnhold, T., Beck, B., and Fox, T. (2005) *QSAR Comb. Sci.*, **24**, 491–502.
- 75 Gafourian, T., Safari, A., Adibkia, K., Parviz, F., and Nokhodchi, A. (2007) *J. Pharm. Sci.*, **96**, 3334–3351.
- 76 Alvarez-Ginatre, Y.A., Crespo-Otero, R., Marrero-Ponce, Y., Noheda-Marin, P., de la Vega, J.M.G., Montero-Cabrera, L.A., García, J.A.R., Caldera-Luzardo, J.A., and Alvarado, Y.J. (2008) *Bioorg. Med. Chem.*, **16**, 6448–6459.
- 77 Duchowicz, P.R., Vitale, M.G., Castro, E.A., Autino, J.C., Romanelli, G.P., and Bennardi, D.O. (2008) *Eur. J. Med. Chem.*, **43**, 1593–1602.
- 78 Mercader, A.G., Duchowicz, P.R., Fernández, F.M., Castro, E.A., Bennardi, D.O., Autino, J.C., and Romanelli, G.P. (2008) *Bioorg. Med. Chem.*, **16**, 7446–7470.
- 79 Gubskaya, A.V., Kholodovich, V., Knight, D., Kohn, J., and Welsh, W.J. (2007) *Polymer*, **48**, 5788–5801.
- 80 Estrada, E., Perdomo-López, I., and Torres-Labandeira, J.J. (2001) *J. Chem. Inf. Comput. Sci.*, **41**, 1561–1568.
- 81 Olivero-Verbel, J. and Pacheco-Londoño, L. (2002) *J. Chem. Inf. Comput. Sci.*, **42**, 1241–1246.
- 82 Safarpour, M.A., Hemmateenejad, B., and Jamali, M. (2003) *QSAR Comb. Sci.*, **22**, 997–1005.
- 83 Sorich, M.J., McKinnon, R.A., Miners, J.O., Winkler, D.A., and Smith,

- P.A. (2004) *J. Med. Chem.*, **47**, 5311–5317.
- 84 Szaleniec, M., Witko, M., Tadeusiewicz, R., and Goclon, J. (2006) *J. Comput. Aided Mol. Des.*, **20**, 145–157.
- 85 Dai, Y., Zhang, X., Wang, H., and Lu, Z. (2008) *J. Mol. Model.*, **14**, 807–812.
- 86 Fernández, M. and Caballero, J. (2008) *QSAR Comb. Sci.*, **27**, 866–875.
- 87 Unger, S.H. and Hansch, C. (1973) *J. Med. Chem.*, **16**, 745–749.
- 88 Mager, P.P. (1984) *Multidimensional Pharmacology: Design of Safer Drugs*, Academic Press, Inc., London.
- 89 Kholodovych, V., Smith, J.R., Knight, D., Abramson, S., Kohn, J., and Welsh, W.J. (2004) *Polymer*, **45**, 7367–7379.
- 90 Gini, G. and Lorenzini, M. (1999) *J. Chem. Inf. Comput. Sci.*, **39**, 1076–1080.
- 91 Molfetta, F.A.d., Angelotti, W.F.D., Romero, R.A.F., Montanari, C.A., and Silva, A.B.F.d. (2008) *J. Mol. Model.*, **14**, 975–985.
- 92 Smith, J.R., Knight, D., Kohn, J., Rasheed, K., Weber, N., Kholodovych, V., and Welsh, W.J. (2004) *J. Chem. Inf. Comput. Sci.*, **44**, 1088–1097.
- 93 Smith, J.R., Kholodovych, V., Knight, D., Kohn, J., and Welsh, W.J. (2005) *Polymer*, **46**, 4296–4306.
- 94 Murthy, S.K. (1998) *Data Min. Knowl. Discovery*, **2**, 345–389.
- 95 P. L. R. Research C5.0, v.5.0 (2002) 5.0 ed., St Ives NSW 2075, Australia, 2002.
- 96 Shannon, C.E. (1948) *Bell System Tech. J.*, **27**, 379–423.
- 97 Shannon, C.E. (1948) *Bell System Tech. J.*, **27**, 623–656.
- 98 Hashemianzadeh, M., Safarpour, M.A., Gholamjani-Moghaddam, K., and Mehdipour, A.R. (2008) *QSAR Comb. Sci.*, **27**, 469–474.
- 99 Alvarez-Ginatre, Y.M., Crespo, R., Montero-Cabrera, L.A., Ruiz-Garcia, J.A., Ponce, Y.M., Santana, R., Pardillo-Fontdevila, E., and Alonso-Becerra, E. (2005) *QSAR Comb. Sci.*, **24**, 218–225.
- 100 Isayev, O., Rasulev, B., Gorb, L., and Leszczynski, J. (2006) *Mol. Diversity*, **10**, 233–245.
- 101 Fernández, M. and Caballero, J. (2007) *J. Mol. Model.*, **13**, 465–476.
- 102 Hansch, C. and Fujita, T. (1964) *J. Am. Chem. Soc.*, **86**, 1616–1626.
- 103 Free, S.M. Jr and Wilson, J.W. (1964) *J. Med. Chem.*, **7**, 395–399.
- 104 Reis, M., Lobato, B., Lameira, J., Santos, A.S., and Alves, C.N. (2007) *Eur. J. Med. Chem.*, **42**, 440–446.
- 105 Liu, W., Yi, P., and Tang, Z. (2006) *QSAR Comb. Sci.*, **25**, 936–943.
- 106 Yu, X., Yi, B., Wang, X., and Xie, Z. (2007) *Chem. Phys.*, **332**, 115–118.
- 107 Yu, X., Xie, Z., Yi, B., Wang, X., and Liu, F. (2007) *Eur. Polym. J.*, **43**, 818–823.
- 108 Yu, X., Wang, X., Li, X., Gao, J., and Wang, H. (2006) *J. Polym. Sci., Part B: Polym. Phys.*, **44**, 409–415.
- 109 Liu, A., Wang, X., Wang, L., Wang, H., and Wang, H. (2007) *Eur. Polym. J.*, **43**, 989–995.
- 110 Chaviara, A.T., Kioseoglou, E.E., Pantazaki, A.A., Tsipis, A.C., Karipidis, P.A., Kyriakidis, D.A., and Bolos, C.A. (2008) *J. Inorg. Biochem.*, **102**, 1749–1764.
- 111 Song, Y., Zhou, J., Zi, S., Xie, J., and Ye, Y. (2005) *Bioorg. Med. Chem.*, **13**, 3169–3173.
- 112 Pasha, F.A., Neaz, M.M., Cho, S.J., and Kang, S.B. (2007) *Chem. Biol. Drug Des.*, **70**, 520–529.
- 113 Wan, J., Zhang, L., Yang, G., and Zhan, C.-G. (2004) *J. Chem. Inf. Comput. Sci.*, **44**, 2099–2105.
- 114 Eroglu, E. and Türkmen, H. (2007) *J. Mol. Graphics Modell.*, **26**, 701–708.
- 115 Zhang, L., Wan, J., and Yang, G. (2004) *Bioorg. Med. Chem.*, **12**, 6183–6191.
- 116 Mehdipour, A.R., Hemmateenejad, B., and Miri, R. (2007) *Chem. Biol. Drug Des.*, **69**, 362–368.
- 117 Holder, A.J., Ye, L., Eick, J.D., and Chappelow, C.C. (2006) *QSAR Comb. Sci.*, **25**, 905–911.
- 118 Van Damme, S., Langenaeker, W., and Bultinck, P. (2008) *J. Mol. Graphics Modell.*, **26**, 1223–1236.
- 119 Hutter, M.C. (2003) *J. Comput. Aided Mol. Des.*, **17**, 415–433.
- 120 Deeb, O., Youssef, K.M., and Hemmateenejad, B. (2008) *QSAR Comb. Sci.*, **26**, 417–424.
- 121 Toropov, A.A., Rasulev, B.F., and Leszczynski, J. (2008) *Bioorg. Med. Chem.*, **16**, 5999–6008.
- 122 Katritzky, A.R., Pacureanu, L., Dobchev, D., and Karelson, M. (2007) *J. Mol. Model.*, **13**, 951–963.

- 123 Matysiak, J. (2008) *QSAR Comb. Sci.*, **27**, 607–617.
- 124 Lu, G.-N., Dang, Z., Tao, X.-Q., Yang, C., and Yi, X.-Y. (2008) *QSAR Comb. Sci.*, **27**, 618–626.
- 125 Niu, J., Long, X., and Shi, S. (2007) *J. Mol. Model.*, **13**, 163–169.
- 126 Ivanciuc, O. (2007) *Reviews in Computational Chemistry*, vol. 23 (eds K.B. Lipkowitz and T.R. Cundari), Wiley-VCH Verlag GmbH, Weinheim.
- 127 Li, J., Qin, J., Liu, H., Yao, X., Liu, M., and Hu, Z. (2008) *QSAR Comb. Sci.*, **27**, 157–164.
- 128 Zheng, G., Xiao, M., and Lu, X. (2007) *QSAR Comb. Sci.*, **26**, 536–541.
- 129 Vapnik, V. (1995) *The Nature of Statistical Learning Theory*, Springer, New York.
- 130 Vapnik, V. and Lerner, A. (1963) *Automat. Remote Contr.*, **24**, 774–780.
- 131 Suykens, J.A.K. and Vandewalle, J. (1999) *Neural Process. Lett.*, **9**, 293–300.
- 132 Scholkopf, B., Smola, A.J., Williamson, R.C., and Bartlett, P.L. (2000) *Neural Comput.*, **12**, 1207–1245.
- 133 Sorich, M.J., Miners, J.O., McKinnon, R.A., and Smith, P.A. (2004) *Mol. Pharmacol.*, **65**, 301–308.
- 134 Bultinck, P., Langenaeker, W., Lahorte, P., De Proft, F., Geerlings, P., Waroquier, M., and Tollenaere, J.P. (2002) *J. Phys. Chem. B*, **106**, 7887–7894.
- 135 Bultinck, P. and Carbó-Dorca, R. (2002) *Chem. Phys. Lett.*, **364**, 357–362.
- 136 Gao, J., Wang, X., Yu, X., Li, X., and Wang, H. (2006) *J. Mol. Model.*, **12**, 521–527.
- 137 Tang, Y., Chen, K.-X., Jiang, H.-L., and Ji, R.-Y. (1998) *Eur. J. Med. Chem.*, **33**, 647–658.
- 138 Hu, L.-H., Chen, G.-H., and Chau, R.M.-W. (2006) *J. Mol. Graphics Modell.*, **24**, 244–253.
- 139 Yaffe, D., Cohen, Y., Espinosa, G., Arenas, A., and Giralt, F. (2002) *J. Chem. Inf. Comput. Sci.*, **42**, 162–183.
- 140 Nguyen-Cong, V. and Rode, B.M. (1996) *Eur. J. Med. Chem.*, **31**, 479–484.
- 141 Novic, M., Nikolovska-Coleska, Z., and Solmajer, T. (1997) *J. Chem. Inf. Comput. Sci.*, **37**, 990–998.
- 142 Molina, E., Estrada, E., Nodarse, D., Torres, L.A., González, H., and Uriarte, E. (2008) *Int. J. Quantum Chem.*, **108**, 1856–1871.
- 143 Katritzky, A.R., Karelson, M., and Petrukhin, R. (2001–2005) CODESSA PRO, University of Florida, Florida.
- 144 Coulson, C.A. (1953) *Adv. Cancer Res.*, **1**, 1–56.
- 145 Pullman, A. and Pullman, B. (1955) *Adv. Cancer Res.*, **3**, 117–169.
- 146 Gao, J., Wang, X., Li, X., Yu, X., and Wang, H. (2006) *J. Mol. Model.*, **12**, 513–520.
- 147 Yu, X., Liu, W., Liu, F., and Wang, X. (2008) *J. Mol. Model.*, **14**, 1065–1070.
- 148 Kholodovich, V., Gubskaya, A.V., Bohrer, M., Harris, N., Knight, D., Kohn, J., and Welsh, W.J. (2008) *Polymer*, **49**, 2435–2439.
- 149 Bicerano, J. (2002) *Prediction of Polymer Properties*, Marcel Dekker, New York.
- 150 Katritzky, A.R. and Gordeeva, E.V. (1993) *J. Chem. Inf. Comput. Sci.*, **33**, 835–857.
- 151 Environment Directorate OECD (2007) *OECD Environment Health and Safety Publications*, Environment Directorate OECD, Paris.
- 152 Ferrari, A.M., Sgobba, M., Gamberini, M.C., and Rastelli, G. (2007) *Eur. J. Med. Chem.*, **42**, 1028–1031.

26

Platinum Complexes as Anti-Cancer Drugs: Modeling of Structure, Activation and Function

*Konstantinos Gkionis, Mark Hicks, Arturo Robertazzi, J. Grant Hill,
and James A. Platts*

26.1

Introduction to Cisplatin Chemistry and Biochemistry

Cisplatin (**1**, *cis*-[Pt Cl₂ (NH₃)₂]) is one of the best-selling anti-cancer drugs of the last four decades (Figure 26.1). First synthesized in the nineteenth century [1], interest was sparked in the 1960s following Rosenberg's serendipitous discovery of cytotoxicity [2, 3], ultimately leading to approval by the US Food and Drug Administration in 1978. Marketed as Platinol, this deceptively simple complex is now widely used as an effective first line treatment for many cancers.

Despite the success of cisplatin as an anti-cancer drug (it is active against testicular cancer, ovarian cancer, cervical cancer, colorectal cancer and relapsed lymphoma), its toxicity in tumor cells is coupled with several drawbacks [4] that have stimulated interest in the development of improved platinum drugs, and consequently in understanding the molecular mechanism that explains the biological activity of platinum compounds [5]. This led to the discovery of carboplatin (**2**, azanide cyclobutane-1,1-dicarboxylic acid platinum) and oxaliplatin (**3**, cyclohexane-1,2-diamine oxalate platinum), which are both widely in use for the treatment of cancers, as well as other complexes (e.g., nedaplatin, lobaplatin, heptaplatin) that have been approved for use in some parts of the world but not globally [6].

Carboplatin is similar to cisplatin in its proposed mechanism of action, and is better tolerated by the body but less efficient than cisplatin due to a chelate effect. Carboplatin is used in the treatment of ovarian cancer, cervical cancer, head and neck cancer, non-small cell lung cancer and relapsed lymphoma. Oxaliplatin is a third-generation platinum anti-cancer drug with a diaminocyclohexane (DACH) entity, predominantly used in the treatment of colorectal cancer. All of these complexes show antitumor activity due to the formation of cytotoxic lesions on DNA with platinum adducts, preventing replication and eventually causing cell death.

The method by which platinum complexes enter cells is a matter of debate, with both passive diffusion across membranes and active transport by copper transporter

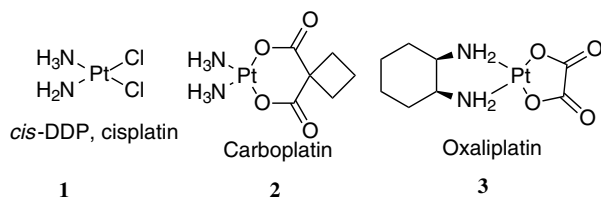


Figure 26.1 Structures of approved platinum drugs.

proteins suggested. However, once into the cytoplasm, cisplatin becomes hydrated with two chloride ligands being replaced by two water ligands to form a positively charged species. Products of its hydrolysis have the ability to interact with nucleophilic molecules within the cell; such molecules include DNA, RNA and proteins. The current accepted thesis is that cisplatin induces its cytotoxic properties through binding to nuclear DNA. Platinum drugs favor binding to N7 atoms of the imidazole rings of guanosine (G) and adenosine (A) bases of DNA, potentially resulting in monoadducts, interstrand and intrastrand crosslinks, the latter being the most likely binding mode. Among all the possible intrastrand crosslinks, there are more 1,2-d(GpG) crosslinks than any other, resulting in a significant DNA distortion. Figure 26.2 shows a view of intrastrand 1,2-d(GpG) cisplatin-DNA complex.

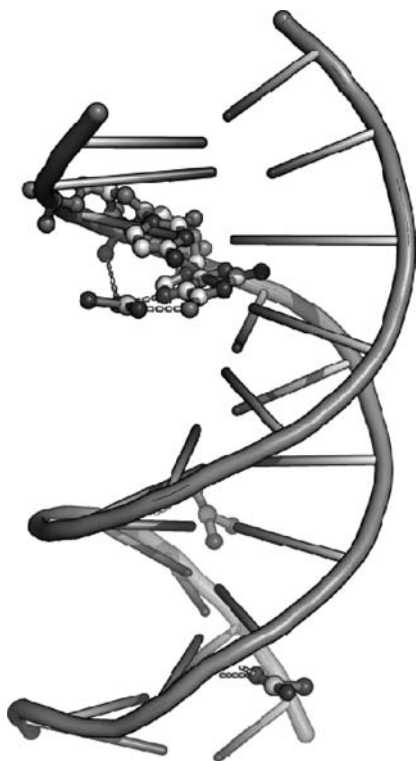


Figure 26.2 Intrastrand 1,2-d(GpG)-cisplatin complex.

Oxaliplatin forms fewer crosslinks than cisplatin at equimolar concentrations as its adducts are bulkier and more hydrophobic than those formed from cisplatin or carboplatin, leading to different effects in the cell. The general consensus is that 1,2-intrastrand crosslinks are responsible for the observed cytotoxicity, as comparative studies with inactive *trans*-diamine-dichloroplatinum (transplatin) show that this is unable to form 1,2-intrastrand crosslinks, but does form 1,3-intra- and interstrand crosslinks.

The major limitations associated with use of these complexes are the toxicity of and resistance to the drugs. Toxicities range from mild to severe nephrotoxicity to peripheral neurotoxicity, with the latter being the most serious. A notable distinction between cisplatin and carboplatin is the difference in the spectrum of toxicities observed between the two. The most common toxicity as a consequence of treatment with oxaliplatin is peripheral neuropathy. In addition, several mechanisms of resistance to these drugs by tumor cells have been observed. Some tumors have a natural resistance to platinum drugs while others develop resistance after the initial treatment. Resistance to cisplatin has been more extensively studied than resistance to carboplatin. However, the suggested resistance mechanisms for both cisplatin and carboplatin are similar, if not identical. The formation of Pt-DNA adducts by cisplatin can be limited (resisted) by reduced accumulation of the drug, enhanced drug efflux and inactivation by coordination to sulfur-containing proteins. Resistance also occurs through enhanced repair of Pt-DNA adducts and increased tolerance of the resulting DNA damage.

These resistance and toxicity issues have led to intensive research efforts into finding new drug candidates with better potency and/or selectivity and reduced side effects. Promising avenues of research include multinuclear platinum complexes, in which two or three Pt centers are coupled together via linker groups, increasing the extent of DNA damage. Biologically active ligands can be incorporated into the platinum complexes, either to increase the concentration of platinum into the cell or to act in combination with DNA lesions. Increased bulk and rigidity of the nitrogen ligands has been proposed to reduce deactivation of the drug by slowing reaction with sulfur-containing molecules such as glutathione. To further improve the affinity of cisplatin for nucleic acids, an alternative approach was also proposed, which involves the tethering of different DNA-binding ligands, such as oligonucleotides, intercalators and DNA-groove binders, to a cisplatin-derivative [7]. In addition, trans-platinum complexes have come back into vogue recently, with NH_3 ligands replaced by heterocyclic ligands such as pyridine, imidazole, thiazole and so on.

Metals other than platinum have also been a focus for the discovery of novel drugs: promising candidates include titanium, ruthenium, rhodium and gold. Most operate in a similar fashion to cisplatin, with cationic and nucleophilic metal centers forming adducts with DNA, although specific details may differ.

Quantum chemical methods can play a significant role in gaining a deeper insight into the mechanism of action of these drugs, and hence suggest novel avenues for research into new and improved treatments. Specific areas of interest include the geometrical and electronic structure of both the drugs themselves and their DNA adducts, the mechanisms and potential energy surfaces of activation and subsequent

reaction with DNA, and means by which cells distribute through the body and enter cells.

This chapter discusses some selected work in these areas, including some by our own group as well as the many groups working worldwide on these topics. We focus on studies based on *ab initio* and/or density functional theory (DFT) methods, leaving to one side the large body of literature using molecular mechanics methods.

26.2

Calculation of Cisplatin Structure, Activation and DNA Interactions

One of the first efforts to apply quantum chemistry to gain a better understanding of cisplatin chemistry was reported in 1985 by Basch *et al.* [8], who used Hartree–Fock (HF) methods together with a double- ζ Gaussian basis set and pseudopotential on Pt to compare cis and trans isomers of $\text{PtCl}_2(\text{NH}_3)_2$ and hydrated complexes. In this way, they found that the trans isomer is about 19 kcal mol^{-1} ($1 \text{ kcal} = 4.184 \text{ kJ}$) more stable than the cis one, a difference attributed to lesser repulsion between chlorides. Geometry optimization gave rather longer Pt–N bonds in cisplatin compared to transplatin, interpreted as evidence for a substantial trans effect in the former, and also geometrical evidence for N–H $\cdots$ Cl hydrogen bonding in the cis isomer.

Recognizing the need for electron correlation to obtain accurate results, Carloni *et al.* used gradient corrected exchange correlation DFT functionals and periodic plane-wave basis sets to examine the isomers [9]. This reproduced Pt–N and Pt–Cl bond lengths to within the estimated error in experimental data (obtained from X-ray diffraction) and confirmed the stability of the trans over the cis isomer, albeit by 8 kcal mol^{-1} at this level. These authors also found evidence for N–H $\cdots$ Cl hydrogen bonds, not least in non-zero values of Mayer bond orders between donor and acceptor atoms. Vibrational spectra were also calculated for cisplatin and transplatin in the range $100\text{--}3500 \text{ cm}^{-1}$, resulting in very good agreement with the available experimental data, especially in the Pt–N and Pt–Cl stretching frequencies, as well as some fine details in the N–H region that are highly sensitive to the symmetry and environment. The bonding in these complexes was examined via difference densities, obtained by subtraction of fragment electron density from that of the entire complex.

Following the work of Carloni *et al.*, Pavankumar *et al.* carried out a systematic study of the structure, bonding, charge density and vibrational frequencies of cisplatin, with particular focus on the dependence of these properties with respect to theoretical methods and basis sets [10]. Electron correlation was included using Møller–Plesset (MP) perturbation theory. Fourteen different basis sets on Cl, N and H were employed, and combined with two pseudopotential schemes, namely those of Stevens, Basch and Krauss (SBK) and of Hay and Wadt (LanL2DZ). Finally, three different possible conformations were explored, that is, two with C_{2v} and one with C_s symmetry; all three were confirmed as minima by harmonic frequency calculation. Agreement with experimental geometry improved with use of larger basis sets, and

with higher orders of the MP series, and was also noticeably better with SBK rather than LanL2DZ pseudopotentials. Vibrational frequency analysis showed considerable variation in calculated frequencies with method and basis set, and the authors proposed MP2/6-311 + + G(2d,2p) as the best choice, giving low overall errors across the range of values observed. Bonding within cisplatin was analyzed using the calculated molecular orbitals, electrostatic potential and electron density analysis.

While the structure of cisplatin itself is of undoubted interest, the fact that it must be activated by hydration has led to significant interest in the interactions and reactions of cisplatin with water. In this context, “hydration” can have two quite different meanings: one refers to the solvation of an intact cisplatin molecule, with water in the second solvation shell; the second refers to the reaction of cisplatin with one or more water molecules, leading to substitution of one or both chloride ligands. Both are important in the chemistry and biochemistry of cisplatin, and have been examined using theoretical methods.

Kozelka and coworkers have investigated potential energy surfaces (PESs) for interaction of a single water molecule with cisplatin in various orientations [11]. From this, favored cisplatin–water conformations and distance dependence were observed. Significant differences between these parameters from HF, MP2 and DFT levels of theory indicated the importance of dispersion energy in determining the details of interaction with water. Intriguingly, these studies demonstrated the existence of “inverse” hydration [12], in which $\text{Pt} \cdots \text{H}-\text{O}$ interactions dominate, as well as the more expected $\text{Pt} \cdots \text{O}$ orientations. Similar PES curves were recently established for a larger number of possible orientations of cisplatin–water complexes by Lopes *et al.* [13], who also extracted Lennard-Jones parameters for faster/larger-scale simulation of cisplatin hydration from these results. Robertazzi *et al.* [14] used DFT methods to examine the optimum geometry of 1:1 cisplatin–water complexes, reporting just three stable minima and characterizing the interactions present in each on the basis of electron density. Coordination to platinum induces substantial polarization of chloride and ammine ligands, causing the N–H groups to act as strong hydrogen bond donors and Cl as strong hydrogen bond acceptors. Extending this treatment gave an estimate of the first solvation shell of cisplatin in a 10:1 complex that contained $\text{N}-\text{H} \cdots \text{O}$, $\text{O}-\text{H} \cdots \text{Cl}$ and $\text{O}-\text{H} \cdots \text{O}$ hydrogen bonds. In addition, analysis of the electron density revealed that the hydrogen bonds between cisplatin and water become slightly weaker when explicit solvation is included, whereas water $\cdots$ water interactions are strongly enhanced by their proximity to cisplatin.

Characterization of the hydrolysis of cisplatin has been of much interest, including the barriers to successive replacement of chlorides, the mechanisms of this process and the overall energy changes accompanying such reactions. Zhang *et al.* [15] reported DFT studies of the hydrolysis reaction using a range of popular exchange-correlation functionals, pseudopotentials on Pt and a self-consistent reaction field (SCRF) estimation of aqueous solvation. They found the reaction to proceed via a five-coordinate, trigonal bipyramidal transition state, described as belonging to the S_N2 class of substitution reaction. This mechanism gives rise to a barrier of about

23 kcal mol⁻¹, slightly in excess of an experimental value of 20 kcal mol⁻¹. Comparison of gas-phase with solvated data indicated that solvation has a major effect on calculated barriers, while barely changing reaction thermodynamics. Since this work, similar studies have tested variations of the level of theory, basis set and treatment of solvation, but have in general found little difference in calculated barriers to reaction and predicted rate constants for hydrolysis. For instance, Robertazzi followed the progress of hydrolysis instigated by one of the water molecules in their 10:1 water-cisplatin complex, that is, including explicit solvation rather than implicit SCRF effects; while evident differences were found in the mechanism of chloride substitution, the energetics of the reaction were hardly affected by the inclusion of explicit solvation.

Deubel and coworker have investigated all the hydrolysis steps of cisplatin employing state-of-the-art DFT calculations [16]. The computational scheme was carefully chosen to tackle problems arising from the estimation of solvation free energies and entropic effects. Such an approach was corroborated by the close agreement of experimental and calculated pK_a values of cisplatin and its hydrolysis products. Interestingly, calculations predicted very similar activation barriers for the three hydrolysis steps, ranging between 25 and 27 kcal mol⁻¹, with reaction free energies of 0–2 kcal mol⁻¹. In addition, a critical comparison of these results with previous studies [17, 18] indicated that the evident disagreement of the second and third hydrolysis barriers, previously predicted to be strongly endergonic, may be due to factors such as the choice of the reference state to calculate the reaction energy and barrier as well as the estimation of free solvation energy and entropic effects. Notably, these calculations supported, for the first time, the experimental hypothesis that the diaqua form of cisplatin is the active species.

Once activated via hydrolysis, the interaction of aquated cisplatin with DNA is the generally accepted mechanism for the observed cytotoxicity of the drug. Details of this interaction have therefore been of great interest to theoretical researchers. In one of the first studies Basch *et al.* used HF methods to probe binding of Pt(NH₃)₂²⁺ to guanine, adenine, cytosine and thymine [19]. This showed preferential binding to the N7 position of guanine (see Figure 26.3 for numbering), with reasonably strong

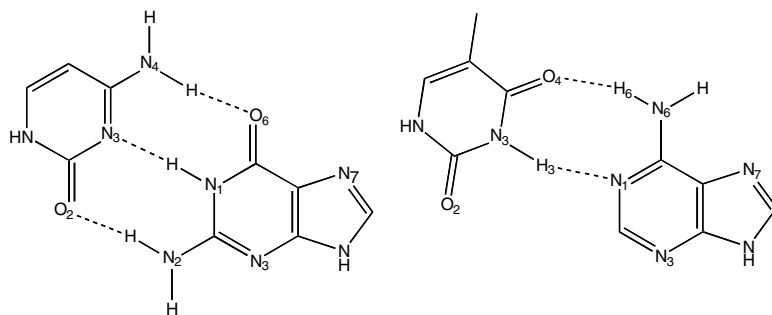


Figure 26.3 Numbering of Watson–Crick paired CG and TA pairs.

binding to N and O sites across most other bases. The order of strength of binding was proposed to be $G(N7) > C(N3) > C(O2) > G(O6) > A(N3)$, $A(N1) > A(N7) > G(N3) > T(O4) > T(O2)$, with approximately 19 kcal mol^{-1} difference between $G(N7)$ and the next most stable site. Bifunctional, or chelating, binding to both N7 and O6 of a single guanine was not found to be a stable orientation, supporting the suggestion that 1,2- and 1,3-intrastrand binding might be the major source of DNA damage by cisplatin. The effects of binding to N7 and O6 sites were monitored via Mulliken population analysis, indicating substantial polarization of the electron density of guanine by the Pt fragment.

Baik *et al.* used DFT methods, pseudopotentials and SCRF solvation to address the key question of why cisplatin binds preferentially to guanine over adenine [20]. The N7 sites of these bases are electronically similar, and are also exposed in the major groove of standard B-DNA. This detailed study confirmed previous findings that cisplatin shows a thermodynamic preference for binding at guanine over adenine, with gas-phase SCF values differing by around 15 kcal mol^{-1} . Hydrogen bonds from coordinated N–H groups of cisplatin to O6 of guanine and N6 of adenine were clearly observed, the latter involving significant pyramidalization of the NH_2 group. The strengths of these hydrogen bonds were estimated at about 7 and 5 kcal mol^{-1} , respectively, and are hence insufficient to account for the difference in stability. Solvation reduces binding energies substantially and also closes the difference between guanine and adenine binding, due to the greater solvation energy of guanine. Zero-point energy (ZPE) and entropy effects, calculated from harmonic vibrational frequency data, reduce the difference in binding further still, giving a $\Delta\Delta G(\text{Sol})$ of about 7 kcal mol^{-1} . These authors also examined the kinetics of guanine and adenine adduct formation from platinum chloroaqua complexes, consistently finding lower barriers for reaction with guanine, quoting values of $24.6 \text{ kcal mol}^{-1}$ for guanine and $30.2 \text{ kcal mol}^{-1}$ for adenine. These differences in barriers were further probed by an energy decomposition scheme and in terms of the frontier molecular orbitals of the nucleic acid bases.

Carloni *et al.* used Car–Parrinello molecular dynamics (CPMD) to monitor the reaction of activated cisplatin with DNA fragments [21], particularly guanine-phosphate-guanine (dGpG), for which X-ray crystallographic data were available. Simulation of the $Pt(NH_3)_2^{2+}$ adduct of this in water was stable, and gave average Pt–N bond lengths of 2.05(7) and 2.03(5) Å (estimated standard deviation in parentheses) for ammonia and guanine ligands, respectively, and angles fluctuating by about 3° from the X-ray values. The angle formed between coordinated guanines was notably smaller than observed in the solid state, which was ascribed to the effect of solvation. Hydrogen bonding patterns to solvent water and to the 5'-phosphate group were monitored across 2 ps of dynamical simulation.

Burda and Leszczynski used the popular B3LYP DFT method, supplemented by MP2 data where necessary, to monitor Pt–DNA interactions, and in particular the structure and energetics of bridged structures with two purine bases [22]. Hydrogen bonds and trans effects were observed to play a role in determining the relative energies of PtA_2 , PtG_2 and $PtAG$ adducts, surprisingly showing that the mixed

complex PtAG is relatively stable, but that PtG₂ is more so. Molecular orbitals and natural bond orbitals (NBOs) were used to rationalize these observations in terms of donation from and back-donation to ligands. Robertazzi *et al.* tackled a similar problem, but this time using topological analysis of the electron density as the principal tool for analysis [23]. Stationary (or critical) points in the electron density unambiguously identify bonds, and properties evaluated at these points can be related to the strength and other properties of the bonds. In this way, the ubiquity of hydrogen bonds in cisplatin adducts of DNA becomes clear: these consist not only of hydrogen bonds from the coordinated N–H ammonia ligands but also between nucleic acid bases. Estimating the strength of these hydrogen bonds on the basis of electron density properties allowed the remaining covalent binding energy to be computed, quantifying the increased Pt affinity of G(N7) over other potential binding sites. The effect of platination on the standard Watson–Crick base pairing of guanine with cytosine was examined in similar fashion: in general, the overall pairing energy of G with C is little affected by the presence of Pt, but the strengths of individual hydrogen bonds (and hence the relative orientation of the duplex DNA) are altered markedly. A clear increase in the Pt–N bond strength in Pt-G ≡ C over Pt-G was also noted.

Mantri *et al.* used DFT to address why cisplatin forms AG links but not GA ones, where the difference lies in the directionality of the nucleic acid chains in DNA, taking d(pApGpA) and smaller segments of this as models of DNA [24]. This work showed little thermodynamic difference between AG and GA adducts, whereas kinetic differences for formation of the two motifs are large. The transition state leading to AG formation is stabilized by hydrogen bonding to a backbone phosphate group, whereas the right-handed nature of the DNA helix prevents this interaction in the GA orientation. The geometrical constraints of binding a single Pt center to two adjacent nucleic acid bases also induces large changes in the backbone dihedrals of the DNA chain, especially in the puckering of the sugar rings, allowing better relaxation and greater tilt angle between purine bases.

The studies discussed above use small segments of DNA as models of larger scale behavior. While there are clearly many lessons that can be drawn from this approach, most experimental data on cisplatin-DNA binding are obtained for larger DNA oligomers, typically duplex DNA structures of between 6 and 16 bases in length. Full quantum mechanical simulation of such large structures is likely to be computationally prohibitive for many years to come so an alternative approach is required. Hybrid QM/MM methods, in which a small region of a larger structure is treated with an accurate quantum mechanical method while the remainder is described by faster molecular mechanics (MM) methods, show much promise in this regard. However, only a few examples of this approach have been published to date.

Spiegel *et al.* coupled CPMD methods to both GROMOS and AMBER MM force-fields to model cisplatin binding to larger DNA oligomers such as d(CCTCTG*G*TCTCC)-d(GGAGACCAGAGG), where * indicates the location of cisplatin binding [25]. This was prompted by observation of significant differences between X-ray and NMR data. Their QM/MM partitioning scheme included

$[\text{Pt}(\text{NH}_3)_2]^{2+}$, two guanines and, in one case, a single phosphate group in the QM region, linked to the remaining MM region by hydrogen link atoms. Single point energies were used to check this partition by extending the QM region to include more bases, including the paired cytosines and flanking thymines in the same strand as the Pt-bound guanine. Simulation of this system over 5–7 ps allowed the structure to relax and geometrical features used to monitor DNA structures such as rise, roll, axis bend, buckle and propeller twist were checked. In this way, the simulated structure was found to be more similar to the NMR than to the X-ray structure, perhaps unsurprisingly since the former is obtained from solution phase. Hydrogen bonds between ammonia ligands and phosphate groups were lost over the course of the simulation, but those to O6 of guanine persisted throughout. “Docking” of $[\text{Pt}(\text{NH}_3)_2]^{2+}$ to the regular structure of B-DNA resulted in a highly distorted structure, but this relaxed to an essentially identical structure to that obtained from the experimental starting point after just a few picoseconds of simulation.

Robertazzi and Platts used the ONIOM procedure for QM/MM to couple DFT methods with AMBER, starting with relatively small models of a single strand of DNA [26]. (See Chapters 2 and 3 for reviews on the ONIOM-QM/MM method.) To account for possible dispersion interactions between nucleobases, the BHandH functional was employed in combination with medium/large-sized basis sets. The partitioning employed here included $[\text{Pt}(\text{NH}_3)_2]^{2+}$ and two–four nucleic acid bases in the QM region, with their associated sugar-phosphate backbones in the MM region. Compared with isolated bases, the effect of the backbone was not found to alter the main trends in binding energies, such as the preference for G over A. However, subtle differences in energies and especially optimized geometries were evident, particularly when the Watson–Crick paired cytosine (for G) or thymine (for A) and their respective sugar-phosphate backbones were included. A similar partitioning was then employed to study cisplatin binding to a larger DNA oligomer, d (CCTG*G*TCC)·d (GGACCAGG), solvated by around 400 water molecules, whose structure is known from NMR experiments. This DFT/AMBER approach satisfactorily reproduced the experimental structure, and again revealed subtle differences in structure and binding than would be obtained from a purely QM approach.

Matsui *et al.* examined whether coordination of cisplatin and transplatin to DNA can induce proton transfer reactions between the nucleobases, using two- and four-base pair models [27]. In this study the ONIOM method was employed for geometry optimization of the four base pair model, using the mPW1PW91 DFT functional and the universal force field (UFF), while the proton transfer reactions were modeled fully at the DFT level using the SDD ECP basis set on the platinum atom and the 6-31G (d,p) basis set on the rest of the atoms. The inclusion of large-scale DNA structure through the ONIOM method did little to change the coordination geometry around Pt. Comparison with available experimental structural data indicated generally good agreement for the platinum coordination site, but slightly larger differences in the lengths of the many hydrogen bonds present within these complexes. Single proton transfer reactions were found to be feasible within the platinated GC pair, but multiple proton transfers were not.

26.3

Platinum-Based Alternatives

The long history in drug development, small size and well-defined coordination geometry of cisplatin make it an ideal candidate for study using quantum chemical methods. Accordingly, fewer publications on the structure and properties of non-cisplatin drugs such as carboplatin have appeared in the literature. Tornaghi performed one of the first comparisons of cisplatin with carboplatin using DFT [28], reporting good general agreement with the experimental X-ray structure. Discrepancies of 3% in bond lengths and 4% in angles were found and explained as errors in DFT/basis set as well as environmental effects in the crystal. Calculation of harmonic frequencies allowed bands in the infrared spectrum to be assigned, with distinct Pt–N bands but Pt–O modes that are strongly coupled to motions of the six-membered ring. Molecular orbital energies showed significant differences to cisplatin, including a rather larger HOMO–LUMO gap due to coupling with ligand-based orbitals. In both cases the LUMO is almost solely the $d_{x^2-y^2}$ orbital, but the HOMO differs, with strong mixing of Pt with O orbitals in cisplatin.

Wysokiński and Michalska compared the performance of several DFT methods in calculating the structure and vibrational frequencies of cisplatin and carboplatin [29]. Findings for cisplatin were not very different from those discussed above. Carboplatin was found to have essentially planar coordination around Pt with the six-membered ring in a boat conformation and a slightly puckered cyclobutane fragment. The exchange-correlation functional mPW1PW91 gave the best agreement with the X-ray crystallographic structure, although differences due to theoretical methods were small. Vibrational spectra were also well calculated by mPW1PW91, in terms of both the normal mode frequencies and the intensities of infrared and/or Raman bands. Recently, Wysokiński *et al.* revisited their predictions of the Raman spectrum of carboplatin, including new measurements of this spectrum, confirming the accuracy of their chosen mPW1PW91 functional [30]. Highly detailed analysis of the vibrational modes within carboplatin was presented, allowing confident assignment of the entire experimental spectrum. NBO analysis was used to study the charge distribution and bonding in carboplatin, indicating a charge of +0.81 on Pt and –1.07 on N, and evidence for a trans-effect from N to O.

Pavelka *et al.* have monitored the hydrolysis reaction of carboplatin, which is of interest since this drug is designed to undergo slower *in vivo* activation than cisplatin due to the chelating nature of the leaving group [31]. Using B3LYP with the Stuttgart–Dresden pseudopotentials, along with SCRF aqueous solvation, they found a barrier of about 31 kcal mol^{–1} for initial disruption of the chelated carboplatin structure, with the reaction proceeding via a five-coordinate transition state similar to that now well established for cisplatin. This first reaction is endothermic by around 15 kcal mol^{–1}. The second hydrolysis step, which leads to loss of the malonato ligand, has a lower barrier of about 22 kcal mol^{–1}, such that the first step appears to be rate limiting. The effects of pH were monitored by comparing hydroxide (HO[–]) with aqua (H₂O) ligands – acidification being known experimentally to speed up this reaction. Indeed, with protonated ligands, the rate-limiting barrier was reduced to around

23 kcal mol⁻¹. It was proposed that changes to the ligand structure on protonation lead to formation of stronger hydrogen bonds, which then act to stabilize the relevant transition state.

Sarmah and Deka have used DFT and SCRF solvation to compare solvation and reactivity indices for cisplatin, carboplatin and oxaliplatin [32]. These indices include global and local hardness and softness, electrophilicity and frontier molecular orbital (Fukui) functions. Optimized geometries for cisplatin and carboplatin were similar to those discussed above, while that for oxaliplatin was similar to the X-ray crystallographic structure, with a chair-like geometry of the cyclohexane ring and NH₂ groups in the equatorial position. The calculated reactivity indices reproduced the experimental trend in reactivity of cisplatin, carboplatin and oxaliplatin. These values were also used to construct a QSAR equation for the experimental cytotoxicity (IC₅₀) of seven platinum complexes, in which a single electrophilicity parameter explained 90% of the variance in the experimental data.

JM118 is a candidate platinum drug that includes a hydrophobic cyclohexane ring, designed to slow hydrolysis and improve cell uptake. Zhu *et al.* used standard B3LYP, LanL2DZ and SCRF methods to examine the hydrolysis of this complex [33]. Similar mechanisms to other Pt(II) complexes were found, with trigonal bipyramidal transition states. Details of the energy barrier were found to be sensitive to the model used, and inclusion of one or more explicit water molecules was necessary as they act to solvate the anionic leaving group. This work confirmed that JM118 should undergo slower hydrolysis than cisplatin, hence allowing time for the complex to reach its cellular target before being deactivated, and also hopefully reducing some of the toxic side effects associated with cisplatin.

Other ligand architectures studied using theoretical methods include a complex with orotic acid (vitamin B₁₃), which is known to act as a biological carrier for metal ions such as Mg²⁺ and Ca²⁺ [34]. As with previous work on carboplatin, the Raman spectrum of this complex was used as the principal source of experimental data, and again mPW1PW91 performed well in reproducing this. NBO analysis was used to examine the bonding and potential hydrogen bonding within this complex. Dos Santos *et al.* studied complexes of Pt(Cl)₂ with tetracycline, a potent broad-spectrum anti-microbial compound [35]. The range of possible coordination sites of the platinum moiety were explored with HF, MP2 and B3LYP methods, leading to 14 separate coordinated forms, in all of which the tetracycline acts as a chelating ligand, with the most stable being coordinated through one N and one O center, both in the gas phase and SCRF water. Changes in predicted ¹³C NMR chemical shift data from the values for free tetracycline were used to compare with experiment. Hydrolysis of the tetracycline-Pt(Cl)₂ complex was also examined, and found to be broadly similar to previous studies of other Pt(II) complexes.

Synthesis of complexes containing more than one Pt center is an increasingly popular strategy in discovery of new drugs; it is proposed that they give 1,2-intra-strand adducts more selectively than single Pt complexes. Deubel has studied dinuclear complexes bridged by hydroxo groups, including their formation from single Pt complexes, activation by hydrolysis and reactions with guanine [36]. Barriers to reaction were found to be higher than for mononuclear compounds. No evidence for

direct Pt–Pt bonding interactions was found on the basis of calculated electron density – instead a ring critical point was located between Pt nuclei. Magistrato *et al.* studied dinuclear Pt complexes bridged by azoles, including those linked by adjacent (1,2) N and non-adjacent (1,3) atoms in the azole, using QM/MM methods to follow complexation to a decamer of duplex DNA [37]. Geometries of the complexes themselves were close (3% relative error) to the X-ray crystallographic values, with the typical square-planar coordination geometry expected of Pt(II). DNA complexes were close to NMR structures, with a RMS deviation of about 0.6 Å; local structural parameters such as “rise”, “roll” and “tilt” were similar to NMR, but major differences to the local structure around Pt from equivalent cisplatin structures were evident. It was suggested that the lack of major changes in the DNA duplex structure may help the Pt adduct escape recognition by repair enzymes, and hence lead to amelioration of the problem of resistance to treatment observed with cisplatin drug therapy.

The above results for the diplatinum azole-bridged complexes were well reproduced by Spiegel *et al.* using force field parameters derived from *ab initio* forces generated by QM/MM calculations following the “force matching” methodology [38]. The parameters were used to perform 10 ns MD simulations, in which the (1,2)-azole derivative–DNA adduct structural characteristics were well reproduced, while larger discrepancies were observed for the (1,3) derivative. This was attributed to the fact that the distortions induced by the (1,3) complex are larger than those of the (1,2). The MD simulations based on *ab initio* forces, apart from reproducing the QM/MM results, have the advantage of implicitly incorporating solvent and temperature effects into the simulation.

QM/MM MD simulations for modeling drug–DNA interactions were reviewed in 2006 by Spiegel and Magistrato [39]. In this work, emphasis was given to cisplatin- and dinuclear azole-bridged-DNA studies as well as to heterocyclic antitumor antibiotics that bind covalently to DNA. Despite the drawback of the limited time scale of the methodology that is pointed out by the authors, the usefulness of QM/MM MD simulations in providing information on the induced structural distortions on the DNA helix upon drug complexation and on the underlying chemistry of the relevant processes is highlighted. As a result, it is concluded that this approach can contribute to the understanding of recognition processes and is a promising method for future studies of non-DNA drug targets.

To improve upon cisplatin activity, Reedijk and coworkers recently synthesized a multifunctional drug (Figure 26.4) combining a cisplatin-derivative and a copper-based artificial DNA-cleaving agent, Cu(3-Clip-Phen) [40]. This was found to bind to DNA from the major groove via the cisplatin subunit and the minor groove via the copper subunit. This combination resulted in the typical effects of cisplatin, for example, DNA bending, together with double-strand cleavage promoted by the copper center, leading to promising biological results [41]. Two theoretical studies focus on the copper phenanthroline subunit: Robertazzi *et al.* [42] employed B3LYP and BLYP DFT functionals to characterize the structural and electronic properties of the parent compound copper-1,10-phenanthroline complex, as well as the more effective derivatives Cu(2-Clip-Phen) and Cu(3-Clip-phen) (phen = phenanthroline), with the two aromatic rings linked by a serinol bridge (Clip). The interaction with a

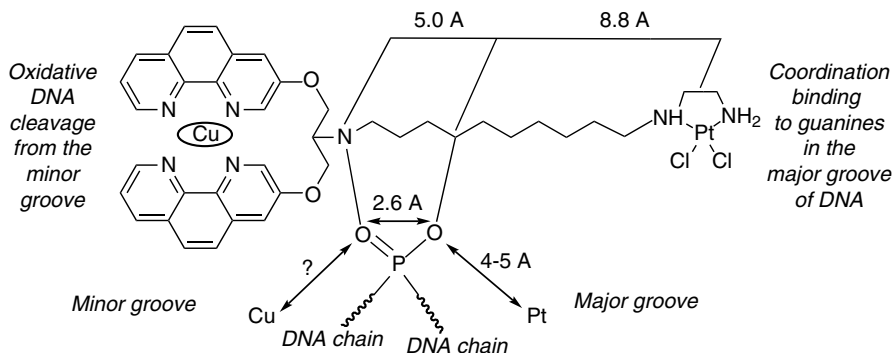


Figure 26.4 Schematic view of Cu(3-Clip-Phen) and its interaction with DNA. (Reproduced with permission of the American Chemical Society from Reference [40].)

DNA fragment was then explored by docking calculations, indicating pseudo-planarity of Cu(3-Clip-Phen) to be one of the key factors of activity. In addition, preliminary calculations [42b] on the entire complex showed that the geometries of cisplatin and copper subunits hardly change when these are combined together. This suggested, in line with experiments, that DNA binding of ditopic cisplatin-copper complex is similar to that of the single components. Further theoretical studies are required to shed light on the structural and electronic properties of such an intriguing example of a potential multifunctional drug.

26.4

Non-platinum Alternatives

Platinum is by no means the only metal that can form DNA adducts and hence show potential anti-cancer properties. The long history and precedent for Pt complexes, along with the highly regular stereochemistry associated with the common oxidation states, mean that structure–activity rules for Pt complexes are well-established, but in recent years complexes of other transition metals, most notably titanium, ruthenium and rhodium, have become prominent in the literature. Robertazzi and Platts examined the entire d-block of transition metals for their interactions with guanine, and their effect on guanine-cytosine pairing, using DFT and QTAIM methods, keeping the oxidation state, overall charge and ligands as similar as possible to those found in cisplatin [43]. Most metals show a clear preference for the N7 coordination site, as is found in Pt complexes, but the “early” transition metals such as titanium and vanadium are thermodynamically more stable when complexed to the O6 site, with a crossover region in group 6 (chromium). The manganese group showed the weakest binding to guanine, whereas the nickel group exhibited the strongest binding.

The effect of metal complexation on base pairing to cytosine was strikingly different depending on whether the N7 or O6 site is preferred, since O6 is also

involved as a hydrogen-bond acceptor in this pairing. As with cisplatin, the overall energy of pairing is barely affected by N7 coordination, although individual hydrogen bonds are affected. In contrast, almost all O6 complexes exhibit much weaker pairing and distorted geometry of base pairs, due to the almost complete loss of the $N_4-H_4 \cdots O_6$ hydrogen bond to cytosine. Molecular electrostatic potentials of O6 and N7 complexes and the differences from free guanine were used to rationalize the observed changes.

Ruthenium complexes show much promise as new, non-platinum drugs. One of the most popular is $ImH[trans-Ru(III)Cl_4(DMSO)(Im)]$ (Im = imidazole), which has been termed NAMI-A for short. This complex, a strong antimetastatic agent, recently completed phase I clinical trials. The mechanism of action of this potential drug remains unknown, but similarities in the activation of the complex by substitution of chlorides with water have been noted, this being required for biological activity. Recently, three studies have been reported on the hydrolysis of this complex. Chen *et al.* [44] used DFT with SCRF solvation to establish the pseudo-octahedral nature of transition states and a barrier of 23 kcal mol^{-1} , in reasonable agreement with experiment. Bešker presented similar work [45] using larger basis sets and explicit water molecules, obtaining better agreement with experiment. Significantly, they showed that the first hydrolysis step is faster than the second.

Vargiu *et al.* [46] studied the same process along with the analogous one in $ImH[trans-RuCl_4(Im)_2]$ (ICR), with a view to understanding the difference in biological activity between these complexes, for example, ICR is active against primary tumors and NAMI-A is an antimetastatic agent. Unlike previous studies, this work considered both Ru(II) and Ru(III) states, all the possible hydrolysis routes (dimethyl sulfoxide and imidazole hydrolysis were also studied) and the reduction potentials for the most relevant metabolites. Reduction is indeed believed to play a key role in the biological activity of these complexes, that is, the kinetically inert Ru(III) may be converted into the more labile Ru(II) complex. Similar reaction profiles to previous work for Ru(III) were found, but their data suggested that Ru(II) di-aqua complexes should be more abundant and possibly play a more important role in the biological activity of NAMI-A. Reduction of ICR was more difficult than for NAMI-A, and had less effect on the overall hydrolysis path.

Chiorescu *et al.* [47] used DFT and SCRF methods to study the Ru(III)/Ru(II) process in more detail, reporting “unprecedented accuracy” in their predictions of experimental data for 61 ruthenium complexes in four solvents, including for NAMI-A. The effects of basis set, and especially details of the SCRF models, were systematically tested, with radii for atoms that make up the solute cavity a particular focus.

Ruthenium features in another class of promising compounds first developed and tested by Sadler [48], in which arene and ethylenediamine ligands are coordinated to Ru, with a single chloride ion. As with other drugs, this chloride is believed to be lost by hydrolysis, leaving a single site for DNA adduct formation. Peacock *et al.* studied the hydrolysis and nucleobase binding of some ruthenium complexes, including O, O- and N,N-chelating ligands, and compared these to the behavior of equivalent

osmium compounds, using DFT and SCRF methods [49]. This work showed that hydrolysis of the O,O-complexes was significantly faster than that of N,N-complexes.

Bešker *et al.* used DFT to study the binding of some simple ruthenium complexes such as $[\text{Ru}(\text{NH}_3)_5]^{n+}$ to nucleobases, and also to some common fragments of amino acids, as well as hydrolysis of chloride complexes [50]. The complexes conform to the expected octahedral geometry and contain multiple hydrogen bonds (typically to O6 of guanine or N6 of adenine). Bond dissociation energies of Ru(II) and Ru(III) complexes were calculated, using SCRF solvation models, leading to similar trends in binding to Pt complexes and stronger binding for Ru(III) than Ru(II). Some trends between bond strengths and proton affinity and/or $\text{p}K_a$ of the corresponding bases were observed, suggesting that the former might be a useful guide to biologic activity.

Gossens *et al.* [51] used standard DFT and MP2 methods as well as CPMD to study the binding of these complexes to different DNA bases, finding a similar order to the preference for binding site observed for cisplatin and related species, $\text{G}(\text{N}7) \gg \text{C}(\text{O}2) \approx \text{C}(\text{N}3) > \text{A}(\text{N}7) > \text{G}(\text{O}6) > \text{OH}2$. Very low barriers for the rotation of the arene were found, indicating that the ligand can re-orient itself to maximize or minimize interactions with its environment. Strong hydrogen bonding was observed between coordinated N–H groups of the ethylenediamine ligand and O6 of guanine, as was flexibility of the backbone N–C–C–N dihedral of the ethylenediamine ligand. The complex to N7 of adenine was found to be significantly less stable than the guanine complex, in accord with experimental findings that greater preference for guanine is found with these Ru complexes than with cisplatin-type complexes. Evidence for a weak hydrogen bond to N6 of adenine was found, which requires a change of hybridization (and hence destabilization) of the amino group.

Gkionis *et al.* tackled a similar problem, but studied more combinations of arene and base [52]. Significantly, many of these combinations are designed to incorporate stabilizing base–arene stacking interactions, which are poorly modeled (if at all) by conventional DFT methods. Because of this, two DFT methods that give better performance for stacking, namely BHandH and M05, were employed. Comparison with X-ray crystallographic data indicated generally good agreement, even for those base–arene combinations where these contacts are present, lending credence to the methods employed. The interaction strengths of these contacts were also checked by use of local MP2 methods, which generally provide better estimates of such weak interactions. As in Reference [51], a clear preference for guanine over all other bases considered was evident, and the binding energy difference between guanine and adenine was rather larger than that found for cisplatin in the gas phase. SCRF solvation reduced this difference, but still leaves a clear preference for the N7 position of guanine. QTAIM analysis was used to identify the contributions of hydrogen bonding and stacking to the stability of each complex, leaving the remainder as the inherent strength of the Ru–N (or Ru–O) covalent bond. The strengths of these bonds were rationalized on the basis of MO energies as well as the weak interactions present, indicating that while the latter are only a small fraction of the total binding energy they play a significant role in modulating the trend in binding energies. Such interactions have been proposed to play a major role in the biologic activity of these

complexes, since larger arene ligands may intercalate into the duplex DNA structure, possibly acting synergistically with the formation of covalent Ru adducts.

Dorcier *et al.* [53] undertook a combined experimental and theoretical study of organometallic ruthenium(II) and osmium(II) anti-cancer complexes binding to an oligonucleotide. They firstly characterized such complexes by means of NMR, mass spectroscopy and crystal X-ray diffraction. Then they explored the biologic activity by investigating the interaction between these complexes with a DNA fragment, employing electrospray ionization mass spectroscopy. Unlike Sadler's compounds, loss of arene groups was observed in certain cases, which may represent a new binding mechanism for a metal–DNA adduct. DFT calculations were then used to rationalize these findings. These calculations suggested that several factors, including change of the metal center (ruthenium versus osmium), methylation/protonation of the ligand and nature of the arene group (*p*-cymene versus benzene) affected the metal–arene binding energies.

Rhodium is another transition metal that shows promise for treatment of cancers. Deubel recently reported DFT studies of the formation of dirhodium–DNA adducts [54]. Unlike most other metal complexes, dirhodium preferentially forms adducts with adenine rather than guanine, and in doing so appears to stabilize an unusual tautomeric form of the nucleobase that can act as a bridging ligand between the two rhodium centers. DFT methods identified the binding energy of various tautomers of adenine to rhodium in both monodentate and bidentate forms, supporting the experimental observation of preference for adenine over guanine. Transition states for formation of bidentate complexes, and for the necessary tautomerization, were identified and associated with barriers of between 20 and 30 kcal mol^{−1}. Modification of the DNA interactions of such complexes, for instance by increasing the lability of carboxylate leaving groups, was also examined, leading to concrete proposals for synthesis of new drug targets.

A similar problem was examined by Burda and Gu [55], who used B3LYP, NBO analysis and electrostatic potentials to examine the structures, stabilities and properties of dirhodium adducts to adenine and guanine. Head-to-head and head-to-tail orientations of nucleobases were considered and again the stability of adenine adducts was noted. However, since formation of these adducts requires proton transfer (tautomerization), the energetics of this also affects the final thermodynamic stability of complexes. This is easier in guanine than in adenine, counteracting the apparent stability of adenine adducts.

Šponer *et al.* [56] used DFT to examine titanocene, [Ti(Cp)₂Cl₂], Cp = cyclopentadiene. Despite superficial similarities to cisplatin, the mode of action of this complex is rather more complex, and may involve loss of Cp as well as chloride ligands. A range of alkylammonium-substituted titanocenes, designed to increase aqueous solubility without reducing biologic activity, were examined and the effects of the substituents on electronic properties and Cp binding energies determined. This work suggested that proton-induced loss of Cp might play an important role in the biochemistry of these complexes, and the energy of protonation shows some correlation with *in vivo* anti-tumor activity. This supports the hypothesis that both Cl and Cp are lost during the action of this drug.

26.5

Absorption, Distribution, Metabolism, Excretion (ADME) Aspects

It is increasingly acknowledged that favorable physicochemical properties of molecules should be taken into account alongside the search for high *in vivo* or *in vitro* activity in the search for new drugs. The factors that lead to a molecule having appropriate chemical properties are generally grouped under the heading of absorption, distribution, metabolism, and excretion (ADME) properties. Absorption and distribution factors include aqueous solubility, lipophilicity, bioavailability and transport across barriers found in cell walls, intestinal walls or between blood and brain. Metabolism is vital for many metal complexes, since the active species is not usually the one administered to patients. The hydrolysis of cisplatin is one example of this, but is so fundamental to the chemistry and biochemistry of this drug that it is not often described as an ADME problem.

An important factor in the design of new organic drugs is the lipophilicity of the molecule, usually taken as the logarithm of the partition coefficient of the species between *n*-octanol and water, $\log P$. Methods to estimate the $\log P$ of typical organic species are common, but applications to metal complexes are scarce, and have been limited by a lack of experimental data. Several years ago we set out a method to estimate $\log P$ for platinum complexes, based on statistical correlations between exposed surface areas of polar, halogen and metal atoms [57]. Geometries of 24 compounds, mostly Pt(II) but also some Pt(IV), were obtained by HF optimization, and surface areas calculated from these geometries. Acceptable correlations were found and the predicted $\log P$ was shown to have an exponential relation with the uptake of platinum compounds into cells. Subsequently, we collaborated with colleagues who measured $\log P$ data for a further 24 Pt(II) complexes by RP-HPLC [58]. Geometries of each compound were optimized using B3LYP and properties related to lipophilicity such as dipole moment, polarizability and electrostatic potentials were calculated at these structures. Rather better statistical relations were found than in our initial study, which seemed to stem from the fact that all measurements were carried out using the same protocol in the same laboratory, rather than being gathered from many diverse literature sources.

More recently, we collated a literature dataset consisting of 43 Pt(II) complexes and a further 21 Pt(IV) complexes, each of which has $\log P$ reported [59]. DFT optimization and property calculations were carried out as before, but the methods used previously did not work as well as might be hoped. The lack of reproducibility across different methods is certainly one source of error, but the limited variety of complexes used in previous studies also seems to be a limitation. We were able to construct acceptable models of $\log P$ by use of a genetic algorithm to select the calculated descriptors that best model the literature data [60], but the neatness of previous work is lost in this fashion. The model was then used to calculate $\log P$ for some metabolites of cisplatin, which cannot be easily measured, and used to rationalize observations on the distribution and uptake of these metabolites.

Metabolism is an important factor in metal drugs in two ways: *in vivo* activation of complexes to their active forms and deactivation by proteins and other species.

A recent alternative to variations on the theme of cisplatin has been the use of Pt(IV) complexes, which typically have octahedral coordination geometry. These are kinetically more inert than Pt(II) analogues and hence can be orally administered with fewer side effects. The Pt(IV) complex is seen as a “pro-drug,” which must be altered by reduction to the Pt(II) form before it can form DNA adducts and exert its biologic effect. Reduction potentials for the Pt(II)/Pt(IV) couple are therefore an important aspect of design: if reduction is too difficult the concentration of active species will be reduced, while if reduction is too easy the pro-drug will not reach its target intact. Reduction potentials for common ligands and metals can be predicted from tabulated constants, as exemplified in the work of Lever’s group [61]. Computational attempts to calculate reduction potentials directly are scarce though [62].

Deactivation of the active species is widely thought to occur through complexation by sulfur ligands, especially the side chains of cysteine and methionine amino acids, and by glutathione. Increased levels of glutathione have been associated with the onset of resistance to cisplatin therapy, such that this aspect has received a great deal of attention. Zimmermann *et al.* [63] used B3LYP to model the interactions of cisplatin and its metabolites with some model thiol ligands. Monodentate and chelate complexes were found to be stable and strong Pt–S bonds were found, in accordance with the hard–soft acid–base (HSAB) principle. Reaction of cisplatin with cysteine is exothermic and thermodynamically competitive with complexation with guanine, whereas methionine is a weaker ligand to Pt. Da Silva came to similar conclusions from their DFT and *ab initio* studies [64]. Predicted rate constants for substitution with sulfur ligands were in good agreement with experimental data.

Deubel and coworker investigated the potential implications of ammine loss from cisplatin, linking this phenomenon to inactivation, storage and resistance of cisplatin [65]. In particular, they employed the B3LYP functional and implicit solvation to study several cisplatin complexes with various biomolecules (including nitrogen heterocycles, neutral and anionic sulfur ligands) to evaluate trans influence and trans effect of the ligands. Upon binding of cisplatin hydrolysis products with these ligands, loss of ammine was predicted as the predominant reaction. It was found that the charge of Pt(II) center has little effect for ammine displacement, while anionic and neutral sulfur ligands exert the strongest trans influence/effect.

References

- 1 Peyrone, M. (1845) *Ann. Chem. Pharm.*, **51**, 129.
- 2 Rosenberg, B., Van Camp, L.V., and Krigas, T. (1965) *Nature*, **205**, 698.
- 3 Rosenberg, B., Van Camp, L.V., Trosko, J.E., and Mansour, V.H. (1969) *Nature*, **222**, 385.
- 4 Wong, E. and Giandomenico, C.M. (1999) *Chem. Rev.*, **99**, 2451.
- 5 Reedijk, J. (1996) *Chem. Commun.*, 801.
- 6 Galanski, M., Jakupc, M.A., and Keppler, B.K. (2005) *Curr. Med. Chem.*, **12**, 2075.
- 7 Reedijk, J. (2005) *Coord. Chem. Rev.*, **249**, 2845–2853.
- 8 Basch, H., Krauss, M., Stevens, W.J., and Cohen, D. (1985) *Inorg. Chem.*, **24**, 3313–3317.
- 9 Carloni, P., Andreoni, W., Hutter, J., Curioni, A., Giannozzi, P., and Parinello, M. (1995) *Chem. Phys. Lett.*, **234**, 50–56.

- 10 Pavankumar, P.N.V., Seetharamulu, P., Yao, S., Saxe, J.D., Reddy, D.G., and Hausheer, F.H. (1999) *J. Comput. Chem.*, **20**, 365–382.
- 11 Kozelka, J., Bergès, J., Attias, R., and Fraita, J. (2000) *Angew. Chem. Int. Ed.*, **39**, 198–201.
- 12 Bergès, J., Caillet, J., Langlet, J., and Kozelka, J. (2001) *Chem. Phys. Lett.*, **344**, 573–577.
- 13 Lopes, J.F., Rocha, W.R., Dos Santos, H.F., and De Almeida, W.B. (2008) *J. Chem. Phys.*, **128**, 165103.
- 14 Robertazzi, A. and Platts, J.A. (2004) *J. Comput. Chem.*, **25**, 1060–1067.
- 15 Zhang, Y., Guo, Z., and You, X.-Z. (2001) *J. Am. Chem. Soc.*, **123**, 9378–9387.
- 16 Lau, J.K.-C. and Deubel, D.V. (2006) *J. Chem. Theory Comput.*, **2**, 103–106.
- 17 Raber, J., Zhu, C., and Eriksson, L.A. (2004) *Mol. Phys.*, **102**, 2537–2544.
- 18 Burda, J.V., Zeizinger, M., and Leszczynski J. (2005) *J. Comput. Chem.*, **26**, 907–914.
- 19 Basch, H., Krauss, M., Stevens, W.J., and Cohen, D. (1986) *Inorg. Chem.*, **25**, 684–688.
- 20 Baik, M.-H., Friesner, R.A., and Lippard, S.J. (2003) *J. Am. Chem. Soc.*, **125**, 14082–14092.
- 21 Carloni, P., Sprik, M., and Andreoni, W. (2000) *J. Phys. Chem. B*, **104**, 823–835.
- 22 Burda, J.V. and Leszczynski, J. (2003) *Inorg. Chem.*, **42**, 7162–7172.
- 23 Robertazzi, A. and Platts, J.A. (2005) *Inorg. Chem.*, **44**, 267–274.
- 24 Mantri, Y., Lippard, S.J., and Baik, M.-H. (2007) *J. Am. Chem. Soc.*, **129**, 5023–5030.
- 25 Spiegel, K., Rothlisberger, U., and Carloni, P. (2004) *J. Phys. Chem. B*, **108**, 2699–2707.
- 26 Robertazzi, A. and Platts, J.A. (2006) *Chem.–Eur. J.*, **12**, 5747–5756.
- 27 Matsui, T., Shigeta, Y., and Hirao, K. (2007) *J. Phys. Chem. B*, **111**, 1176–1181.
- 28 Tornaghi, E., Andreoni, W., Carloni, P., Hutter, J., and Parinello, M. (1995) *Chem. Phys. Lett.*, **246**, 469–474.
- 29 Wysokiński, R. and Michalska, D. (2001) *J. Comput. Chem.*, **22**, 901–912.
- 30 Wysokiński, R., Kuduk-Jaworska, J., and Michalska, D. (2006) *J. Mol. Struct. (THEOCHEM)*, **758**, 169–179.
- 31 Pavelka, M., Lucas, M.F.A., and Russo, N. (2007) *Chem.–Eur. J.*, **13**, 10108–10116.
- 32 Sarmah, P. and Deka, R.C. (2008) *Int. J. Quantum Chem.*, **108**, 1400–1409.
- 33 Zhu, C., Raber, J., and Eriksson, L.A. (2005) *J. Phys. Chem. B*, **109**, 12195–12205.
- 34 Wysokiński, R., Hernik, K., Szostak, R., and Michalska, D. (2007) *Chem. Phys.*, **333**, 37–48.
- 35 Dos Santos, H.F., Marcial, B.L., De Miranda, C.F., Costa, L.A.S., and De Almeida, W.B. (2006) *J. Inorg. Biochem.*, **100**, 1594–1605.
- 36 Deubel, D.V. (2006) *J. Am. Chem. Soc.*, **128**, 1654–1663.
- 37 Magistrato, A., Ruggerone, P., Spiegel, K., Carloni, P., and Reedijk, J. (2006) *J. Phys. Chem. B*, **110**, 3604–3613.
- 38 Spiegel, K., Magistrato, A., Maurer, P., Ruggerone, P., Rothlisberger, U., Carloni, P., Reedijk, J., and Klein, M.L. (2008) *J. Comput. Chem.*, **29**, 38–49.
- 39 Spiegel, K. and Magistrato, A. (2006) *Org. Biomol. Chem.*, **4**, 2507–2517.
- 40 De Hoog, P., Boldron, C., Gamez, P., Slidregt-Bol, K., Roland, I., Pitié, M., Kiss, R., Meunier, B., and Reedijk, J. (2007) *J. Med. Chem.*, **50**, 3148–3152.
- 41 De Hoog, P., Louwerse, M.J., Gamez, P., Pitié, M., Baerends, E.J., Meunier, B., and Reedijk, J. (2008) *Eur. J. Inorg. Chem.*, **4**, 612–619.
- 42 (a) Robertazzi, A., Magistrato, A., De Hoog, P., Carloni, P., and Reedijk, J. (2007) *J. Inorg. Chem.*, **46**, 5873–5881.
(b) Robertazzi, A., Vargui, A.V., Magistrato, A., Ruggerone, P., Carloni, P., de Hoog, P., Reedijk, J., (2009) *J. Phys. Chem. B*, **113**, 10881–10890.
- 43 Robertazzi, A. and Platts, J.A. (2005) *J. Biol. Inorg. Chem.*, **10**, 854–866.
- 44 Chen, J., Chen, L., Liao, S., Zheng, K., and Ji, L. (2007) *J. Phys. Chem. B*, **111**, 7862–7869.
- 45 Bešker, N., Coletti, C., Marrone, A., and Re, N. (2008) *J. Phys. Chem. B*, **112**, 3871–3875.
- 46 Vargui, A.V., Robertazzi, A., Magistrato, A., Ruggerone, P., and Carloni, P. (2008) *J. Phys. Chem. B*, **112**, 4401–4409.
- 47 Chiorescu, I., Deubel, D.V., Arion, V.B., and Keppler, B.K. (2008) *J. Chem. Theory Comput.*, **4**, 499–506.
- 48 Morris, R.E., Aird, R.E., Murdoch, P., Del, S., Chen, H., Cummings, J., Hughes, N.D., Parsons, S., Parkin, A., Boyd, G., Jodrell, D.I., and Sadler, P.J. (2001) *J. Med. Chem.*, **44**, 3616–3621.
- 49 Peacock, A.F.A., Melchart, M., Deeth, R.J., Habtemariam, A., Parsons, S., and

- Sadler, P.J. (2007) *Chem.-Eur. J.*, **13**, 2601–2613.
- 50 Bešker, N., Coletti, C., Marrone, A., and Re, N. (2007) *J. Phys. Chem. B*, **111**, 9955–9964.
- 51 Gossens, C., Tavernelli, I., and Rothlisberger, U. (2007) *J. Chem. Theory Comput.*, **3**, 1212–1222.
- 52 Gkionis, K., Platts, J.A., and Hill, J.G. (2008) *Inorg. Chem.*, **47**, 3893–3902.
- 53 Dorcier, A., Dyson, P.J., Gossens, C., Rothlisberger, U., Scopelliti, R., and Tavernelli, I. (2005) *Organometallics*, **24**, 2114–2123.
- 54 Deubel, D.V. (2008) *J. Am. Chem. Soc.*, **130**, 665–675.
- 55 Burda, J.V. and Gu, J. (2008) *J. Inorg. Biochem.*, **102**, 53–62.
- 56 Šponer, J.E., Leszczynski, J., and Šponer, J. (2006) *J. Phys. Chem. B*, **110**, 19632–19636.
- 57 Platts, J.A., Hibbs, D.E., Hambley, T.W., and Hall, M.D. (2001) *J. Med. Chem.*, **44**, 472–474.
- 58 Platts, J.A., Oldfield, S.P., Reif, M.M., Palmucci, A., Gabano, E., and Osella, D. (2006) *J. Inorg. Biochem.*, **100**, 1199–1207.
- 59 Tetko, I.V., Jaroszewicz, I., Platts, J.A., and Kuduk-Jaworska, J. (2008) *J. Inorg. Biochem.*, **102**, 1424–1437.
- 60 Oldfield, S.P., Hall, M.D., and Platts, J.A. (2007) *J. Med. Chem.*, **50**, 5227–5237.
- 61 Lever, A.B.P. (1990) *Inorg. Chem.*, **29**, 1271–1285.
- 62 Dobrogorskaia-Méreau, I.A.I. and Nemukhin, A.V. (2005) *J. Comput. Chem.*, **26**, 865–870.
- 63 Zimmermann, T., Zeizinger, M., and Burda, J.V. (2005) *J. Inorg. Biochem.*, **99**, 2184–2196.
- 64 Da Silva, V.J., Costa, L.A.S., and Dos Santos, H.F. (2008) *Int. J. Quantum Chem.*, **108**, 401–414.
- 65 Lau, J.K.-L. and Deubel, D.V. (2005) *Chem. Eur. J.*, **11**, 2849–2855.

27

Protein Misfolding: The Quantum Biochemical Search for a Solution to Alzheimer's Disease

Donald F. Weaver

27.1

Introduction

“Men ought to know that from the brain, and from the brain alone, arises our pleasures, joys, laughter and jests, as well as our sorrows, pains, grief and tears. Through it, we think, we see, we hear, and we distinguish the ugly from the beautiful, the bad from the good, the pleasant from the unpleasant . . . It is this same brain which makes us mad or delirious, inspiring us with dread or fear . . . These things that we suffer all come from the brain – when it is not healthy.”

Hippocrates, 5th century BCE

It is astonishing how trivial amounts of certain molecules can dramatically affect the brain, the mind, the thoughts and the lives of people who are ill; likewise, it is truly awe inspiring how administering 2 mg of a drug can transform an unconscious person back to their normal conscious self. Since science is a very human activity, it is crucial that we develop tools to facilitate our overcoming the “molecule to mind” hurdles that impede drug discovery, thereby enabling us to devise truly effective therapeutics for the myriad of tragic brain disorders that afflict humankind. Quantum biochemistry is such a tool. Quantum biochemistry is a powerful and far-reaching tool, permitting a molecular level conceptualization of the mechanistic processes that underlie the clinical phenomenology of human disease.

Quantum biochemistry may offer particularly significant insights into the class of brain diseases that arise from the pathological process termed “protein misfolding.” Over the course of the past decade, a growing body of evidence has shown that protein misfolding is an important biochemical process underlying multiple devastating neurological diseases, including Alzheimer's disease, Parkinson's disease and mad cow disease [1]. This chapter will discuss and explore the utility of quantum biochemistry as a tool to be used in the battle against protein misfolding disorders.

27.2

Protein Folding and Misfolding

Like the other major biochemical macromolecules, including polysaccharides and nucleic acids, proteins are essential parts of organisms, participating in virtually every cellular process. Chemically, proteins are linear polymers (i.e., polypeptides) built from 20 different L- α -amino acids; all amino acids possess common structural features, including an α -carbon to which an amino group, a carboxyl group and a variable side chain are bonded. Biochemically, proteins are composed of amino acids arranged in a linear polypeptide chain and joined together along a backbone by peptide (amide) bonds between the carboxyl and amino groups of adjacent amino acid residues; the sequence of amino acids in a protein is defined by the sequence of a gene, encoded in the genetic code.

27.2.1

Protein Folding

Protein folding is the biophysical process by which a polypeptide folds into its characteristic and functional three-dimensional structure – since structure determines function, protein folding is a crucial way of regulating biological activity. The manner in which a newly synthesized chain of amino acids transforms itself into a perfectly folded protein depends both on the intrinsic properties of the amino acid sequence and on multiple contributing influences from within the cellular milieu. Biochemically, four distinct aspects of a protein's structure and shape are recognized:

- **Primary structure:** the sequence of amino acids.
- **Secondary structure:** regularly repeating local structures stabilized by hydrogen bonds; most common examples are the α -helix (a right-handed coiled conformation in which every backbone N–H group donates a hydrogen bond to the backbone C=O group of the amino acid four residues earlier within the primary structure) and β -sheet (consisting of peptide strands connected laterally by three or more hydrogen bonds, forming a twisted, pleated sheet; these β -strands are arranged adjacent to other strands and form an extensive hydrogen bond network with their neighbors, in which the N–H groups in the backbone of one strand establish hydrogen bonds with the C=O groups in the backbone of the adjacent strands).
- **Tertiary structure:** Tertiary structure determines the overall shape of a single protein molecule by controlling the spatial relationship of the various secondary structural regions to one another. The tertiary structure is stabilized by nonlocal interactions, most commonly the formation of a hydrophobic core, but also through hydrogen bonds, electrostatic interactions (salt bridges) and even disulfide bonds.
- **Quaternary structure:** the shape or structure that results from the interaction of more than one protein/peptide molecule, usually called protein subunits in this context, which function as part of the larger assembly or protein complex.

This ordered hierarchy of structure, extending from primary to quaternary, defines the shape of the protein and thus its biological function. Many proteins are enzymes, which catalyze biochemical reactions, being vital to cellular metabolism; other proteins have structural/mechanical functions, such as the proteins in the cytoskeleton, which form a system of scaffolding to maintain cell shape; still other proteins are important in cell signaling, immune responses, cell adhesion and the cell cycle. Because of the importance of correctly folded proteins to the overall cellular function and health (and thus the organism as a whole), it is not surprising that there exists an elaborate quality control system to prevent protein misfolding.

27.2.2

Protein Misfolding

When the protein folding process escapes from the cellular quality-control mechanisms, protein misfolding occurs. In turn, aggregation of these misfolded proteins heralds a wide range of highly debilitating and increasingly prevalent diseases. For example, aggregated misfolded proteins are associated with extracellular amyloid-aggregation illnesses such as Alzheimer's disease and familial amyloid polyneuropathy, as well as intracytoplasmic aggregation diseases such as Huntington's and Parkinson's disease [1]. The prion-related illnesses such as Creutzfeldt–Jakob disease and bovine spongiform encephalopathy (mad cow disease) are also protein misfolding disorders. These diverse degenerative brain diseases are associated with the deposition of misfolded proteins into insoluble, extracellular aggregates and/or intracellular inclusions; it is not clear whether the aggregates are the cause or merely a reflection of the loss of protein homeostasis – the balance between synthesis, folding, aggregation and protein turnover. Nevertheless, protein misfolding is an integral part of these severe neurodegenerative disorders and the discovery of therapeutics for these disorders is emerging as one of the most important neuropharmacological priorities of the twenty-first century.

27.3

Quantum Biochemistry in the Study of Protein Misfolding

Although there is a crucial need for therapeutics to deal with the problem of protein misfolding, conventional methods of drug design are not applicable to this unique biochemical problem. Traditionally, rational drug design involves having the three-dimensional structure (usually an X-ray structure) of a receptor protein in hand, and then insightfully engineering therapeutic molecules to interact with this protein. However, this tried and true approach simply will not work for protein misfolding disorders. Determining the geometries and experimental structures of aberrant misfolded proteins has thus far proven to be impossible; there are no available crystal structures, there are no satisfactory NMR structures. This lack of rigorous structural data for misfolded proteins is a major stumbling block to the development of therapeutics.

Since there are no experimental structures, the opportunity exists for studying the three-dimensional properties of misfolded proteins using quantum biochemistry methodologies. This is one area in which quantum mechanics and molecular mechanics calculations can offer important insights for understanding the three-dimensional properties of disease producing proteins (and for which no other techniques are applicable). Since the problem of protein misfolding is immense, the full spectrum of quantum biochemistry methods (including quantum mechanics, molecular mechanics and bioinformatics) must be employed. Elsewhere in this book, the various quantum mechanics techniques have been described in detail; accordingly, this chapter will briefly review molecular mechanics. Moreover, when studying macromolecular problems such as protein misfolding, molecular mechanics is a useful tool.

27.3.1

Molecular Mechanics

Molecular mechanics arises from the principles of classical mechanics, rather than those of quantum mechanics. Quantum mechanics is based upon an explicit consideration of electrons and electron properties. Molecular mechanics, on the other hand, does not consider electrons explicitly. The term molecular mechanics refers to a heavily parameterized computational method that leads to accurate geometries and accurate relative energies for different conformations of molecules. The molecular mechanics procedure employs the fundamental equations of vibrational spectroscopy, and represents a natural evolution of the notions that atoms are held together by bonds and that additional interactions exist between non-bonded atoms. The essential idea of molecular mechanics is that a molecule is a collection of particles held together by elastic or harmonic forces, which can be defined individually in terms of potential energy functions. The sum of these various potential energy equations comprises a multidimensional energy function termed the force field, which describes the restoring forces acting on a molecule when the minimal potential energy is perturbed. The force field approach supposes that bonds have natural lengths and angles, and that molecules relax their geometries to assume these values. The incorporation of van der Waals potential functions and electrostatic terms allows the inclusion of steric interactions and electrostatic effects. In strained systems, molecules will deform in predictable ways, with strain energies that can be readily calculated. Thus, molecular mechanics uses an empirically derived set of simple classical mechanical equations, and is in principle well suited to provide accurate *a priori* structures and energies for peptides or other biochemical macromolecules of pharmacological interest.

Molecular mechanics lies conceptually between quantum mechanics and classical mechanics, in that data obtained from quantum mechanical calculations are incorporated into a theoretical framework established by the classical equations of motion. The Born–Oppenheimer approximation, used in quantum mechanics, states that Schrödinger's equation can be separated into a part that describes the motion of

electrons and a part that describes the motion of nuclei, and that these can be treated independently. Quantum mechanics is concerned with the properties of electrons; molecular mechanics is concerned with the nuclei, while electrons are implicitly treated in a classical electrostatic manner.

The heart of molecular mechanics is the force field equation [2]. A typical molecular mechanics force field is shown below:

$$V = V_r + V_\theta + V_\omega + V_{\text{inv}} + V_{\text{nb}} + V_{\text{hb}} + V_{\text{cross}} \quad (27.1)$$

or:

$$V = \sum k_r (r - r_0)^2 + \sum k_\theta (\theta - \theta_0)^2 + \sum \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \\ + \sum_{i < j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{\epsilon r_{ij}} \right) + \sum \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + V_{\text{cross}} \quad (27.2)$$

where V_r represents bond length energies; V_θ represents bond angle energies; V_ω represents dihedral angle energies; V_{nb} represents non-bonded interaction energies (van der Waals and electrostatic); V_{hb} represents hydrogen bonding interactions. Typically, the bond stretching and bending functions are derived from Hooke's law of harmonic potentials; a truncated Fourier series approach to the torsional energy permits accurate reproduction of conformational preferences. The molecular mechanics method is extremely parameter dependent. A force field equation that has been empirically parameterized for calculating peptides must be used for peptides; it cannot be applied to nucleic acids without being re-parameterized for that particular class of molecules.

Molecular mechanics calculations provide reasonable structural information about both small and large molecules, containing 10–10 000 atoms. Therefore, molecular mechanics calculations are useful for studying larger molecules such as the misfolded proteins that have been implicated in the pathogenesis of neurodegenerative diseases such as Alzheimer's disease.

27.4

Alzheimer's Disease: A Disorder of Protein Misfolding

Alzheimer's disease (AD) is a progressive neurodegenerative disease that first manifests with mild cognitive, language and behavioral symptoms, which gradually worsen in severity and eventually lead to dementia. AD is the most common cause of dementia, accounting for approximately 75–80% of cases; it affects 11% of people aged 80–84, and 24% of those aged 85–93 years [1]. There is no remission in the progression of Alzheimer's disease, nor are there any disease-stabilizing drugs currently available [3]. As such, onset of the disease is inevitably followed by increasing mental and physical incapacitation, loss of independent living, institutionalization and death. There is usually an 8–10 year period from symptom onset until death.

Alzheimer's – A Protein Misfolding Disorder

AD is characterized by two neuropathological hallmarks: extracellular deposits of misfolded β -amyloid protein (producing amyloid or “senile” plaques) and intracellular deposits of misfolded tau protein (producing neurofibrillary tangles [NFTs]). Because of these misfolded protein deposits, the AD brain is also characterized by a dramatic loss of neurons and synapses, particularly in areas involving higher order cognitive functions such as the basal forebrain and hippocampus. Also, the levels of many neurotransmitters are greatly reduced, including serotonin, dopamine, glutamate and especially acetylcholine. These reduced neurotransmitter levels are responsible for the broad and profound clinical manifestations of AD; that is, memory impairment, cognitive deficits, restlessness and depression [4].

Protein Misfolding of Beta-Amyloid

The peptide known as β -amyloid ($A\beta$) is widely regarded as the chief molecular culprit in AD. $A\beta$ is a 40–43 amino acid protein that forms extraneuronal aggregates having a β -pleated structure. $A\beta$ is initially formed within the amyloid precursor protein (APP), a 695–770 amino acid transmembrane protein found in virtually all brain cells [4]. The sequence of $A\beta$ 1–43 is shown in Figure 27.1, using the single-letter amino acid abbreviations; the region contained within the transmembrane sequence of APP is identified, as are the sites of action of three APP-processing enzymes, α -, β - and γ -secretase. The roles of these enzymes are discussed below.

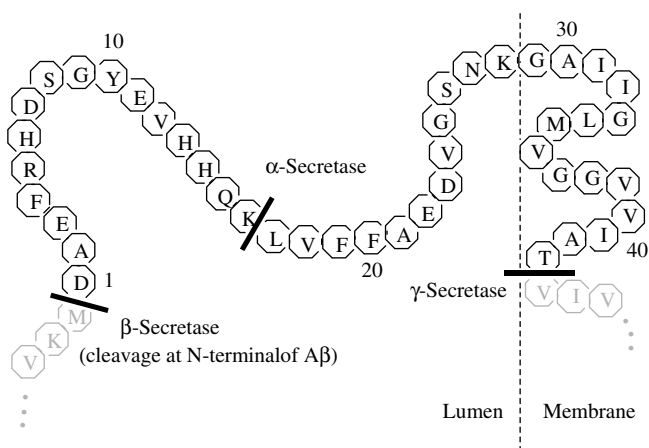


Figure 27.1 Sequence of A β 1–43 and sites of secretase cleavage on amyloid precursor protein (APP). γ -Secretase has low specificity, cleaving the amyloid precursor protein (APP) anywhere between residues 39 to 43 of A β (cleavage after residue 43 shown here).

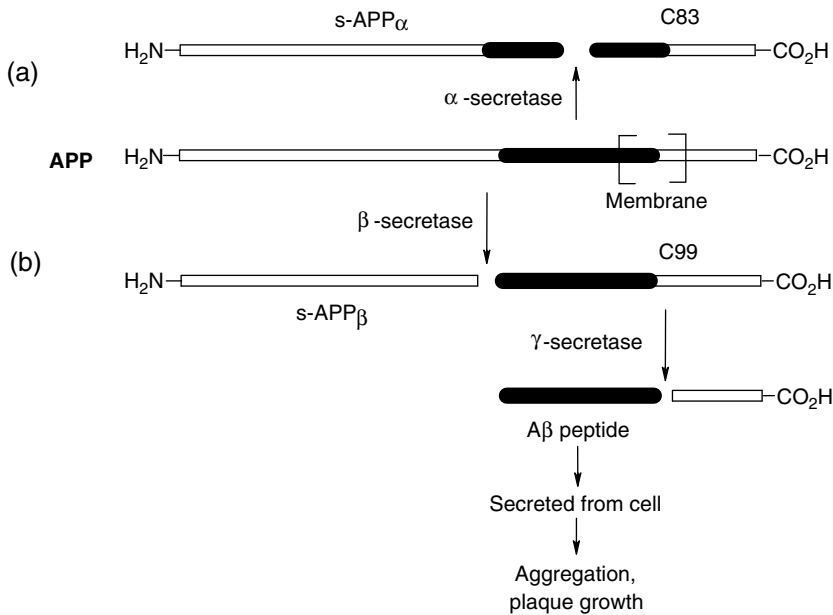


Figure 27.2 Processing of amyloid precursor protein (APP): (a) “normal” cleavage within A β region by α -secretase; (b) pathogenic cleavage of APP by β - and γ -secretase, liberating A β , which can self-assemble into neurotoxic aggregates and become incorporated into growing plaques.

APP is normally cleaved within the A β domain by α -secretase (Figure 27.2), liberating a soluble N-terminal fragment (s-APP α) and a membrane-bound C-terminal fragment (C83). Alternatively, APP can be cleaved by β -secretase at the N-terminus of the A β domain, giving s-APP β and C99. The latter membrane-bound fragment then undergoes intra-membrane cleavage by γ -secretase at the C-terminus of A β , resulting in the liberation of A β into the cell [5]. The secretion of A β follows, allowing the peptide to participate in extracellular aggregation and become incorporated into growing plaques [6]. This cascade of events is depicted in Figure 27.2.

Aggregates of misfolded A β , resulting either from an overabundance of A β production or from an impairment of the brain's ability to dispose of the peptide, are believed to be the molecular event causally related to AD. There is debate as to what level of A β aggregation is responsible for the observed neurotoxicity. Historically, large aggregates, present as senile plaques in the brains of AD patients, were suspected of being the brain-destructive species. Recent evidence however, has suggested that smaller, diffusible A β oligomers may be the chief mediators of neurotoxicity in AD [7–11]. Research into these assemblies has alternatively identified A β dimers, trimers, dodecamers and other small oligomers as the principal toxic species protein misfolding [12, 13]. Regardless of whether it is small, oligomeric A β species or large fibrils that are of greatest detriment to neurons, preventing or even reversing the self-assembly of A β would be expected to

mitigate protein misfolding mediated neuronal damage and offer a therapeutic approach to AD.

27.5

Quantum Biochemistry and Designing Drugs for Alzheimer's Disease

How does one sit down in front of a blank computer screen and think about designing drugs for the treatment of AD? In theory, approaches to the treatment of AD can be conceptually divided into five categories: (i) inhibition of A β production or secretion, (ii) inhibition of A β aggregation/deposition, (iii) vaccination against misfolded A β , (iv) treatment of the neurotoxic inflammatory and/or oxidative stresses that occur secondary to A β deposition or (v) neurotransmitter-based therapies to symptomatically deal with the neurotransmitter deficiency arising from the inability of A β -damaged neurons to produce neurotransmitters [14–16].

In this work, inhibition of A β aggregation/deposition will be targeted for multiple reasons. We are seeking to devise a disease-modifying agent for AD. AD arises from an immuno-inflammatory process in which protein misfolding of β -amyloid initiates a neurotoxic cascade that triggers sequential activation of the innate and adaptive immune systems, leading to complement activation, microglial stimulation, chemokine/cytokine release and ultimately neuronal death [5]. Given that protein misfolding is a key initiating factor in AD pathology, an A β anti-aggregant could in principle arrest the neurotoxic immunological cascade at its onset, exerting a disease-altering effect. Therefore, the focused goal of our research is to employ quantum biochemistry as a tool with which to discover an A β anti-aggregant compound.

To initiate our search for an A β anti-aggregant, we devised a computational model of non-toxic monomeric (unaggregated) A β in order to identify receptor targets that would enable either the prevention or interruption of the aggregation process; this *in silico* receptor model is referred to as the HHQK-BBXB model (and its existence is supported by various experimental data). In 1998, Giulian *et al.* suggested that the H₁₃H₁₄Q₁₅K₁₆ tetrapeptide domain within A β provided a structural basis for the immunopathology of AD, noting that A β _{13–16} was necessary for initial A β aggregation and subsequent microglial activation through a cell surface mechanism mediated by a glycosaminoglycan called heparan sulfate proteoglycan [26]. Recently, we extended this HHQK immune-mediating motif to a more general HHQK-BBXB model (where B denotes any basic cationic residue) [17]. Based on molecular mechanics calculations, we identified the HHQK-BBXB receptor within A β as a 39.6 Å² triangle with Lys₁₆ at one corner and the two histidine residues at the other corners; *in silico* simulations suggest that binding to this receptor should prevent the initiation and propagation of A β aggregation.

The HHQK-BBXB receptor consists of three positively charged basic residues in a 1-2-4 arrangement within a tetrapeptide motif. Since binding to a single amino acid would not impart sufficient intermolecular binding selectivity, it was determined that a molecule capable of binding to BBXB at either two or three of the “B” residues was

required. Assuming an α -helical conformation and using a molecular mechanics energy minimized geometry of HHQK, the 1–2, 1–4 and 2–4 inter-residue side-chain charge separations are 9.0, 13.5 and 10.8 Å, respectively. To establish energetically favorable intermolecular interactions with these cationic basic B-type residues, the two preferred methods are via an anionic group (forming a cationic–*anionic* Coulombic interaction) or an aromatic group (forming a cationic–*aromatic* [cation– π] interaction).

Accordingly, we pursued two approaches to devising therapeutics for AD capable of binding to the microenvironment containing the HHQK peptidic motif, thereby inhibiting A β aggregation. The first approach was to identify a molecule capable of forming an *anionic*–cationic Coulombic interaction with the histidine or lysine residues of the HHQK-BBxB motif; the second approach was to identify a molecule capable of forming an *aromatic*–*cationic* interaction with the histidine or lysine residues of the HHQK-BBxB motif.

27.5.1

Approach 1 – Homotaurine

Homotaurine (3-aminopropane-1-sulfonic acid; 3-APS) is a structural analog of taurine. 3-APS belongs to a family of therapeutic mono- and polysulfated/sulfonated molecules computationally designed by Weaver and coworkers as agents to prevent neurotoxic A β aggregation (see US Patents 5,643,562(1997); 5,728,375(1998); 5,972,328(1999)). Because of its molecular geometry and the presence of a sulfonate moiety, 3-APS functions as a glycosaminoglycan mimetic, binding to the E11-V12-H13-H14 tetrapeptide segment of A β , in a fashion similar to heparan sulfate proteoglycan, thereby preventing the overlapping HHQK region of A β from facilitating the aggregation process.

During the design process, quantum biochemistry calculations were used to evaluate the interaction between 3-APS and A β . The starting conformation of A β was obtained from the RCSB (Research Collaboratory for Structural Bioinformatics) Protein Data Bank (PDB ID: 1AML). The amyloid peptide conformational space was searched using molecular dynamics, Monte Carlo and genetic algorithm techniques to ensure a local energy minimum in the region of the starting conformation. Further energy minimizations were carried out using a molecular mechanics approach employing the CHARMM force field. Calculations were done both *in vacuo* and in a 30 Å box of TIP3P image water molecules; these calculations were performed using the CHARMM force field employing a steepest descent algorithm followed by a conjugate gradient energy minimization algorithm.

These calculations confirmed that *in silico* the preferred binding location for 3-APS on A β is the EVHH domain. EVHH has a negatively charged residue on one end (E) and positively charged residue(s) (H) at the other; correspondingly, 3-APS has a positively charged ammonium at one end and a negatively charged anionic sulfur moiety at the other. The intermolecular interaction between 3-APS and the EVHH domain within A β is geometrically and energetically favored, with a binding energy of 45 kcal mol^{−1} (1 kcal = 4.184 kJ) within the CHARMM force field. Also using the

CHARMm force field, the 3-APS molecule was positioned at 38 other points systematically along the A β primary structure; all of these other binding site interactions were $>10 \text{ kcal mol}^{-1}$ greater in energy. By binding to the EVHH motif, 3-APS has the capacity to inhibit the HHQK mediated aggregation of A β .

We next explored the 3-APS/amyloid interaction at a higher level of theory. Given the large nature of the β -amyloid peptide, the HHQK region of interest was isolated and capped with amide groups. The backbone of the peptide segment was constrained and the system was minimized using the CHARMm force field to achieve a low energy, stable state [2]. AM1 RHF singlet state calculations were then performed to calculate the heat of formation of the tetrapeptide [18]. 3-APS was constructed in a zwitterionic state and minimized using the CHARMm force field in MOE to identify a low energy conformation [19]. The zwitterionic 3-APS was then oriented with the negatively charged functional group oriented towards the His13, His 14 and Lys16 amino acid side chains with multiple possible orientations being examined; these systems were first optimized using the CHARMm force field, followed by subsequent energy minimizations at the AM1 semiempirical molecular orbital level of theory. These calculations confirmed an energetically and geometrically favorable interaction between 3-APS and β -amyloid.

27.5.2

Approach 2 – Melatonin

Quantum biochemical calculations also suggested that melatonin should be able to bind to B-type residues within the HHQK-BBxB motif, via an aromatic–cationic interaction involving the indole moiety within melatonin. To evaluate this aromatic–cationic interaction, we first evaluated multiple protein structures within the Protein Data Bank to determine if indole rings residue demonstrate a propensity for associating with B-type residues, such as histidine or lysine. We employed data mining methods that we had already developed to study noncovalent interactions between aromatic rings and peptidic functional groups [20–23]. Applying these same methodologies, we studied 1029 X-ray protein structures from the PDB, identifying 3241 histidines of which 16% were closely associated with an indole ring (indicating an energetically favorable interaction between indole rings and histidine side-chains).

To rigorously evaluate the capacity of indole rings to establish a cation– π interaction with a B-type amino acid, a detailed set of quantum mechanics calculations was next performed. The ability of methylammonium (MA; model of the lysine side-chain) and 4-methylimidazolium (4-MI, model of the histidine side-chain) to bind in an energetically favorable manner to nine different aromatic rings (benzene, benzofuran, benzothiophene, furan, indole, naphthalene, pyridine, pyrrole and thiophene) was calculated using 3-21G (RHF) and 6-31G(d) (RHF and MP2) level molecular orbital quantum mechanics calculations. These calculations on model systems showed that the pyridine and indole heteroaromatic rings bound to both MA and 4-MI with the greatest binding energies ($27\text{--}30 \text{ kcal mol}^{-1}$). Next, to further study the geometry of the interaction, additional quantum mechanics calculations were

performed using density functional theory computations at the B3LYP/6-31G* level on complexes between NH_4^+ (representing the side-chain terminus of lysine) and the indole aromatic ring. This series of calculations showed that in the optimal intermolecular structure the NH_4^+ was located directly over the hexa-atomic ring of the indole (rather than the penta-atomic ring) with the shortest distance between a hydrogen atom of NH_4^+ and the indole plane being 2.1 Å. Thus, rigorous quantum mechanics calculations confirmed that the indole ring of melatonin can bind in an energetically and geometrically favorable manner to isolated BBXB side-chains.

These calculations were then extended to full peptide structures. The ability of melatonin to bind individually to the HHQK motif within A β was modeled using molecular mechanics calculations as implemented in the CHARMM force field. The starting conformation of A β was obtained from the RCSB Protein Data Bank (PDB ID: 1AML). An initial A β peptide conformational energy minimization was carried out with the force field employing a conjugate gradient energy minimization algorithm. Within this force field level of theory, the ability of melatonin to interact with the HHQK-BBXB motif of A β was then explicitly simulated. Using multiple different starting orientations, melatonin was docked at a 3 Å separation from the BBXB domain at more than ten starting points; then, the complexes were allowed to relax through a minimization cycle. The system was solvated explicitly for an all-atom minimization, followed by computation of binding energies under distance-dependent solvation. Pentane was used as a negative control, with the cutoff for binding deemed to be 5 kcal mol⁻¹ better than pentane. These molecular mechanics simulations confirmed an energetically favorable interaction between the indole moiety of melatonin and the HHQK domain within A β .

Additional molecular modeling studies were also performed to preliminarily explore the detailed molecular basis of melatonin-mediated inhibition of amyloid aggregation. Starting with β -amyloid in a pre-aggregated α -helical conformation, melatonin was placed in 126 different starting positions (three different starting points along each of the 42 residues). Melatonin bound with the greatest binding affinity at residues 13–14 in A β 1–42. These two residues are part of the HHQK domain. These *in silico* calculations show that melatonin is able to bind to the BBXB receptor motif in A β [24].

27.6 Conclusions

Quantum biochemistry is a tool, a very powerful tool, especially when applied to disease processes such as AD and related protein misfolding disorders. This chapter demonstrates the utility of quantum biochemistry in identifying potential new chemical entities for the treatment of AD. Additional work in the Weaver laboratory has demonstrated the ability of these classes of compounds to prevent A β aggregation *in vitro* using various experimental biochemical assays. 3-APS, designated with the drug name tramiprosate, has completed Phase I, II and III human clinical trials [25].

"There is no more common error than to assume that, because prolonged and accurate mathematical calculations have been made, the application of the result to some fact of nature is absolutely certain."

A.N. Whitehead, *An Introduction to Mathematics*

Acknowledgments

D.F.W. acknowledges salary support from a Canada Research Chair, Tier 1, in Neuroscience. This work was funded, in part, by operating and infrastructure grants from the American Health Assistance Foundation, the Canada Foundation for Innovation, and the Atlantic Innovation Fund of the Atlantic Canada Opportunities Agency.

References

- 1 Nussbaum, R.L. and Ellis, C.E. (2003) Alzheimer's disease and Parkinson's disease. *New Engl. J. Med.*, **348**, 1356–1364.
- 2 MacKerell, A.D. Jr, Bashford, D., Bellott, M., Dunbrack, R.L. Jr, Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, W.E. III, Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics of proteins. *J. Phys. Chem. B*, **102**, 3586.
- 3 Selkoe, D.J. and Schenk, D. (2003) Alzheimer's disease: molecular understanding predicts amyloid-based therapeutics. *Annu. Rev. Pharmacol. Toxicol.*, **43**, 545–584.
- 4 Selkoe, D.J. (1993) Physiological production of the beta-amyloid protein and the mechanism of Alzheimer's disease. *Trends Neurosci.*, **16**, 403–409.
- 5 Mattson, M.P. (2004) Pathways towards and away from Alzheimer's disease. *Nature*, **430**, 631–639.
- 6 Selkoe, D.J. (2005) Defining molecular targets to prevent Alzheimer disease. *Arch. Neurol.*, **62**, 192–195.
- 7 Lambert, M.P., Barlow, A.K., Chromy, B.A., Edwards, C., Freed, R., Liosatos, M., Morgan, T.E., Rozovsky, I., Trommer, B., Viola, K.L., Wals, P., Zhang, C., Finch, C.E., Krafft, G.A., and Klein, W.L. (1998) Diffusible, nonfibrillar ligands derived from Abeta1-42 are potent central nervous system neurotoxins. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 6448–6453.
- 8 Walsh, D.M. and Selkoe, D.J. (2004) Deciphering the molecular basis of memory failure in Alzheimer's disease. *Neuron*, **44**, 181–193.
- 9 Walsh, D.M. and Selkoe, D.J. (2004) Oligomers on the brain: the emerging role of soluble protein aggregates in neurodegeneration. *Protein Pept. Lett.*, **11**, 213–228.
- 10 Haass, C. and Selkoe, D.J. (2007) Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid beta-peptide. *Nat. Rev. Mol. Cell Biol.*, **8**, 101–112.
- 11 Walsh, D.M. and Selkoe, D.J. (2007) A beta oligomers - a decade of discovery. *J. Neurochem.*, **101**, 1172–1184.
- 12 Townsend, M., Shankar, G.M., Mehta, T., Walsh, D.M., and Selkoe, D.J. (2006) Effects of secreted oligomers of amyloid beta-protein on hippocampal synaptic plasticity: a potent role for trimers. *J. Physiol.*, **572**, 477–492.

- 13 Walsh, D.M., Klyubin, I., Shankar, G.M., Townsend, M., Fadeeva, J.V., Betts, V., Podlisny, M.B., Cleary, J.P., Ashe, K.H., Rowan, M.J., and Selkoe, D.J. (2005) The role of cell-derived oligomers of Abeta in Alzheimer's disease and avenues for therapeutic intervention. *Biochem. Soc. Trans.*, **33**, 1087–1090.
- 14 Melnikova, I. (2007) Therapies for Alzheimer's disease. *Nat. Rev. Drug Discov.*, **6**, 341–342.
- 15 Wolfe, M.S. (2002) Secretase as a target for Alzheimer's disease. *Curr. Top. Med. Chem.*, **2**, 371–383.
- 16 Janus, C., Pearson, J., McLaurin, J., Mathews, P.M., Jiang, Y., Schmidt, S.D., Chishti, M.A., Horne, P., Heslin, D., French, J., Mount, H.T., Nixon, R.A., Mercken, M., Bergeron, C., Fraser, P.E., St-George-Hyslop, P., and Westaway, D. (2000) A beta peptide immunization reduces behavioural impairment and plaques in a model of Alzheimer's disease. *Nature*, **408**, 979–982.
- 17 Stephenson, V., Heyding, A., and Weaver, D.F. (2005) The promiscuous drug concept with applications to Alzheimer's disease. *FEBS Lett.*, **579**, 1338–1342.
- 18 Dewar, M.J.S., Zoebisch, E.G., Healy, E.F., and Stewart, J.J.P. (1985) Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.*, **107**, 3902.
- 19 Chemical Computing Group, Inc., Molecular Operating Environment, 2007.09, Montreal, Quebec, Canada.
- 20 Duan, G., Smith, V.H. Jr, and Weaver, D.F. (1999) An *ab initio* and data mining study on aromatic-amide interactions. *Chem. Phys. Lett.*, **310**, 323–332.
- 21 Duan, G., Smith, V.H. Jr, and Weaver, D.F. (2000) A data mining and *ab initio* quantum mechanics study of the interaction between the aromatic and backbone amide groups in proteins. *Int. J. Quantum Chem.*, **80**, 44–60.
- 22 Duan, G., Smith, V.H. Jr, and Weaver, D.F. (2000) Characterization of aromatic-amide (side-chain) interactions in proteins through systematic *ab initio* calculations and data mining analyses. *J. Phys. Chem. A*, **104**, 4521–4532.
- 23 Duan, G., Smith, V.H. Jr, and Weaver, D.F. (2002) A data mining, *ab initio* quantum mechanics and molecular mechanics study on the conformation of phenylalanine and on its interaction with neighbouring backbone amide groups in protein structure. *Int. J. Quantum Chem.*, **90**, 669–683.
- 24 Pappolla, M., Bozner, P., Soto, C., Shao, H., Robakis, N.K., Zagorski, M., Frangione, B., and Ghiso, J. (1998) Inhibition of Alzheimer beta-fibrillogenesis by melatonin. *J. Biol. Chem.*, **273**, 7185–7188.
- 25 Gervais, F., Paquette, J., Morissette, C., Krzywkowski, P., Yu, M., Azzi, M., Lacombe, D., Kong, X., Aman, A., Laurin, J., Szarek, W.A., and Tremblay, P. (2007) Targeting soluble Abeta peptide with tramiprosate for the treatment of brain amyloidosis. *Neurobiol. Aging*, **28**, 537–547.
- 26 Giulian, D., Haverkamp, L.J., Yu, J., Karshin, W., Tom, D., Li, J., Kazanskaia, A., Kirkpatrick, J., and Roher, A.E. (1998) The HHQK domain of beta-amyloid provides a structural basis for the immunopathology of Alzheimer's disease. *J. Biol. Chem.*, **273**, 29719–29726.

28

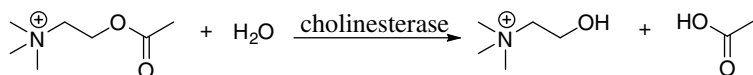
Targeting Butyrylcholinesterase for Alzheimer's Disease Therapy

Katherine V. Darvesh, Ian R. Pottie, Robert S. McDonald,
Earl Martin, and Sultan Darvesh

28.1

Butyrylcholinesterase and the Regulation of Cholinergic Neurotransmission

Acetylcholine acts as a neurotransmitter by being released from the nerve cell into the synaptic cleft where it stimulates another nerve cell by binding to its receptor. Once the signal is initiated, the neurotransmitter must be degraded to prevent prolonged stimulation. Cholinesterases deactivate acetylcholine by catalyzing its hydrolysis, thereby regulating neurotransmission between nerve cells. Cholinesterases belong to a family of hydrolase enzymes that exploit the nucleophilic potential of a serine hydroxyl group and the acid/base chemistry of histidine and glutamate to effect the hydrolysis of complex biomolecules such as esters and amides. Two cholinesterases are widely distributed in animals, acetylcholinesterase (AChE, EC 3.1.1.7) and butyrylcholinesterase (BuChE, EC 3.1.1.8). Both enzymes are able to efficiently catalyze the hydrolysis of the neurotransmitter acetylcholine (Scheme 28.1), thereby regulating cholinergic neurotransmission.



Scheme 28.1 Hydrolysis of acetylcholine by cholinesterases.

Although the two enzymes bear many similarities in terms of overall structure, amino acid sequence and mechanism of catalysis [1–3], there are numerous differences, not the least of which relates to substrate specificity. AChE is highly specific for acetylcholine as substrate whereas BuChE can hydrolyze not only acetylcholine but also a wide variety of larger esters and amides [4–6]. These observed differences in substrate specificity for AChE and BuChE can be largely attributed to distinct differences in the volume of the region where the hydrolysis reaction takes

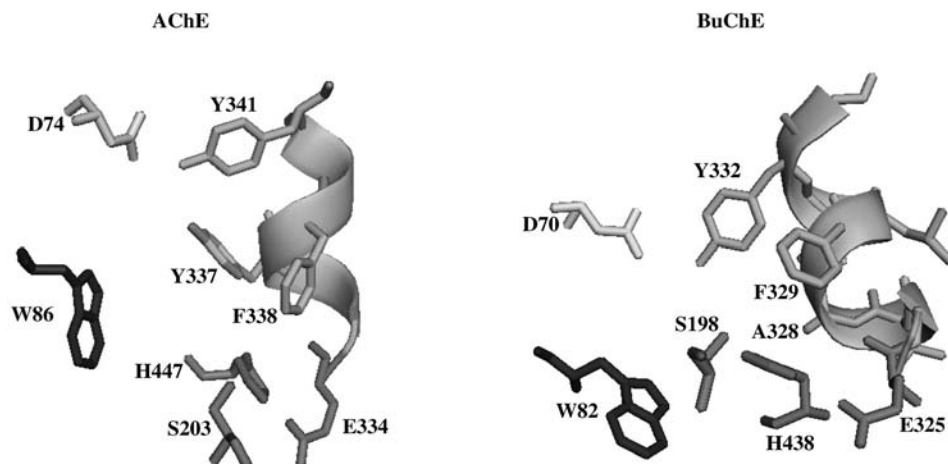


Figure 28.1 Active site gorge regions of (a) AChE and (b) BuChE. The figure was generated using crystal structure coordinates of cholinesterase [7, 8] from the Protein Data Bank [36], using The PyMOL Molecular Graphics System [37].

place and in the unique arrangement of amino acid residues in that region of the two enzymes. These details have been worked out over the past decade through X-ray diffraction analysis of purified crystalline AChE [7] and BuChE [8]. Unlike many enzymes that have catalytic sites at the protein surface, the catalytic site of a cholinesterase is near the bottom of a 20 Å deep depression in the protein tertiary structure that is referred to as the active site gorge. In AChE, this depression is partially lined by 14 amino acids with bulky aromatic side chains while the BuChE gorge has only eight such residues. A consequence of this is that the available estimated volume of the BuChE active site gorge ($\sim 500 \text{ \AA}^3$) is more than 1.5 times greater than that of AChE ($\sim 300 \text{ \AA}^3$) [9]. However, both enzymes have the same catalytic triad arrangement (Figure 28.1), consisting of serine, histidine and glutamate, which co-operate in converting the alcohol group of the serine into a strong nucleophile through proton transfer [10].

In addition to the catalytic triad, certain other regions have been identified as important in contributing to the similarities and differences exhibited by AChE and BuChE. The Ω -loop contains a negatively charged aspartate residue (D74 in AChE, D70 in BuChE) that is part of the peripheral anionic site for initial binding of cationic substances, such as acetylcholine, near the top of the active site gorge. Further down the gorge, on the same Ω -loop, is a tryptophan residue (W86 in AChE, W82 in BuChE), the aromatic ring system of which provides a π -cation interaction with the same substrate as it is positioned over the catalytic triad. Groups attached to the carbonyl at the other end of the choline ester are accommodated within a loop (the acyl loop), on the opposite side of the gorge, in a binding site termed the acyl pocket (Figure 28.2). In AChE, this pocket is restricted in size, being lined by two aromatic residues (F295 and F297), while in BuChE these side chains are replaced

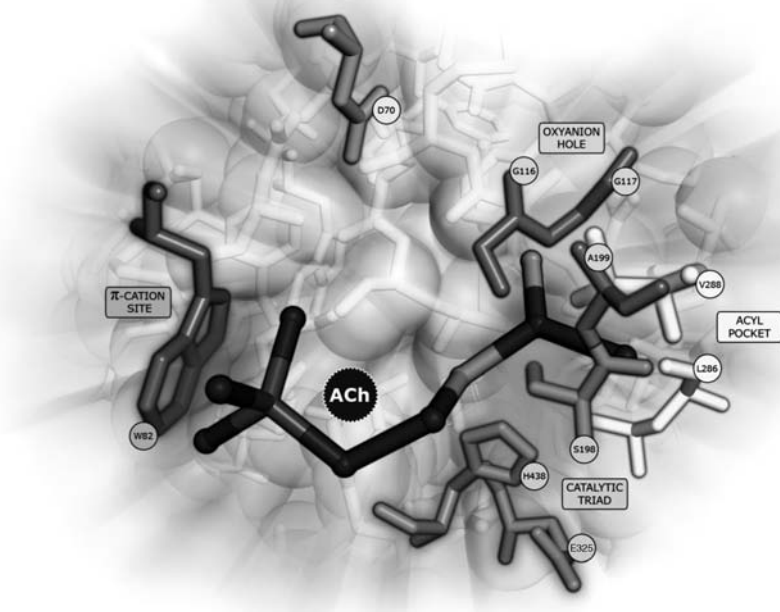


Figure 28.2 Schematic representation of the butyrylcholinesterase active site gorge with acetylcholine (ACh) positioned prior to catalysis. The figure was generated using crystal structure coordinates of butyrylcholinesterase [8] from the Protein Data Bank [36], using The PyMOL Molecular Graphics System [37].

by the smaller alkyl groups of leucine and valine (L286 and V288), which permit the binding of much larger groups attached to the substrate carbonyl. Thus, although acetylcholine is effectively hydrolyzed by both AChE and BuChE, only BuChE shows any efficient hydrolysis of structurally diverse substrates such as butyrylcholine, benzoylcholine, succinylcholine, acetylsalicylic acid, cocaine, heroin, esters of fatty acids and other larger molecules, such as ghrelin [3], all of which are esters but are very diverse in their overall structures. Another important region in the active site gorge is the helical region [11] (E-helix) (Figure 28.1) that contains a tyrosine residue that is part of the peripheral anionic site (Y341 in AChE; Y332 in BuChE) and the anionic glutamate that is a component of the catalytic triad (E334 in AChE, E325 in BuChE). Prior to the catalytic event, these electrostatic interactions of substrate with enzyme ensure the proper alignment of the electrophilic carbonyl carbon over the nucleophilic oxygen of the catalytic serine residue, S203 in AChE and S198 in BuChE (see Figure 28.2). A further arrangement of amino acid residues, G121, G122 and A204 in AChE and G116, G117 and A199 in BuChE

form an oxyanion hole that helps stabilize the alkoxide of the tetrahedral intermediate in the first step of the catalytic process by forming hydrogen bonds ($-N-H \cdots O-$) with the negatively charged carbonyl oxygen of the ester substrate transition state.

28.2

Butyrylcholinesterase: The Significant other Cholinesterase, in Sickness and in Health

For the better part of a century, studies of cholinesterase function have focused on the highly specific AChE as a regulator of cholinergic systems [2, 10]. The importance of cholinesterases in neurotransmission is confirmed by the observed improvement in memory and behavior in neurodegenerative disorders that exhibit low levels of brain acetylcholine, such as in Alzheimer's disease (AD), by treating patients with cholinesterase inhibitors such as donepezil, galantamine and rivastigmine [12].

In recent years several lines of evidence have suggested that the particular inhibition of BuChE may also be important in the treatment of neurodegenerative diseases. The specific inhibition of BuChE has been shown to increase brain levels of acetylcholine [2] and to produce improvement in cognition and behavior [13]. BuChE is expressed in those regions of the brain involved in cognition and behavior [3, 14]. Furthermore, AChE knockout mice are viable, indicating the importance of BuChE in compensating for the absence of AChE [15]. A further indication that BuChE inhibition could be important in dementia is the observed high levels of this enzyme found to be associated with the neuritic plaques and neurofibrillary tangles characteristic of brain tissue lesions in AD [16, 17]. Finally, all of the cholinesterase inhibitors found effective in treating symptoms in AD inhibit both AChE and BuChE [3]. Hence it is difficult to determine, at present, which cholinesterase inhibition is producing the bulk of the observed positive effects of treatment with cholinesterase inhibitors. These observations have guided a portion of the work in our laboratory towards the development of specific and potent inhibitors of BuChE for the treatment of neurodegenerative disorders. One series of inhibitors being considered for this purpose is a collection of derivatives based on phenothiazine (Figure 28.3).

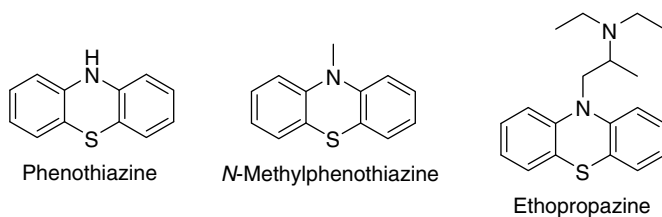
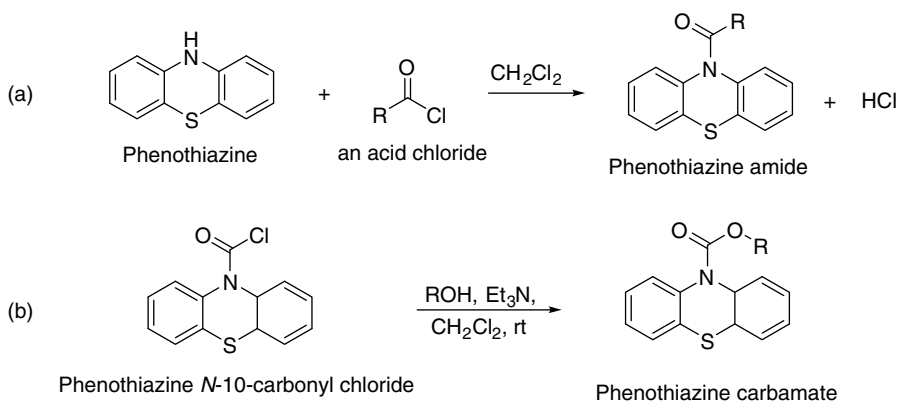


Figure 28.3 Examples of compounds based on the phenothiazine scaffold.

28.3

Optimizing Specific Inhibitors of Butyrylcholinesterase Based on the Phenothiazine Scaffold

Phenothiazine and many of its derivatives are known, among other things, to be inhibitors of human BuChE, but not AChE [10, 18], and several of these molecules, such as ethopropazine (Figure 28.3), are potent inhibitors of that enzyme. To study the structure–activity relationships of phenothiazines as cholinesterase inhibitors, a series of *N*-10-carbonyl derivatives of phenothiazine were synthesized and examined for the ability to inhibit butyrylcholinesterase. These compounds included two distinct classes, amides, prepared from phenothiazine itself [19, 20] (Scheme 28.2, reaction a) and carbamates prepared from phenothiazine-*N*-10-carbonyl chloride [11] (Scheme 28.2, reaction b). Substituents on the carbonyl side of these derivatives included alkyl, cycloalkyl, alkylaryl and aryl groups (Tables 28.1 and 28.2).



Scheme 28.2 Synthesis of phenothiazine derivatives: (a) amide and (b) carbamate.

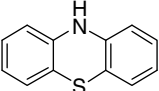
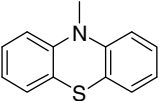
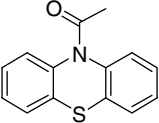
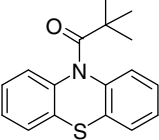
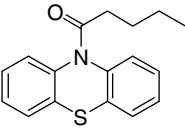
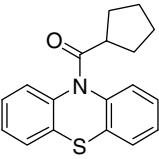
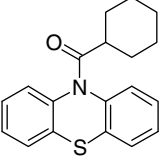
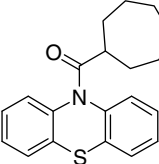
Each synthetic compound was tested for its ability to inhibit AChE and/or BuChE. Assays were performed using the methodology of Ellman *et al.* [21] that employs esters of thiocholine as substrate analogues for acetylcholine and the non-limiting Ellman reagent, 5,5'-dithio-*bis*(2-nitrobenzoic acid) (DTNB), to detect thioester hydrolysis as it occurs (Scheme 28.3).

28.4

Biological Evaluation of Phenothiazine Derivatives as Cholinesterase Inhibitors

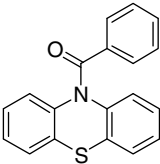
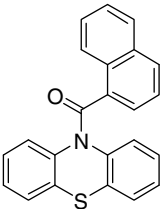
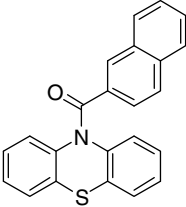
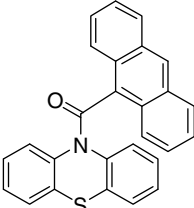
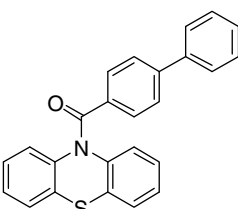
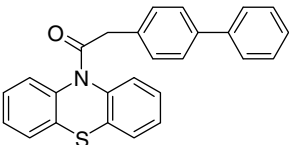
Kinetic analysis of thiocholine hydrolysis using the Ellman method, in the presence and absence of one of the phenothiazine derivatives, provided an indication of the type of inhibition, if any, whether the inhibitor was affecting substrate affinity (K_m value), maximum velocity ($V_{\max}$ value) or both, and the relative potency of inhibition (K_i or k_a values). The phenothiazine derivatives examined were found to be of two general inhibitor classes. One class, the reversible inhibitors, blocks substrate

Table 28.1 Inhibition constants (K_i) and molecular volumes for *N*-10-phenothiazine amides.

Entry	Structure	Compound #	AChE (K_i μM) ^{a)}	BuChE (K_i μM) ^{a)}	Molecular volume ($\text{\AA}^3$)
1		1	None	31.8 ± 10.4	212
2		2	None	36.9 ± 2.7	223
3		3	38.9 ± 3.5	35.8 ± 2.5	244
4		4	29.5 ± 4.5	6.33 ± 0.18	308
5		5	None	10.2 ± 1.5	315
6		6	None	4.21 ± 0.4	315
7		7	None	1.26 ± 0.22	343
8		8	None	0.86 ± 0.16	362

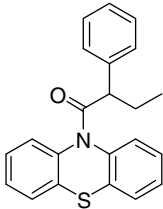
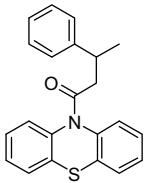
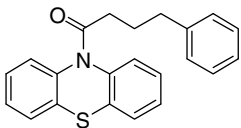
(Continued)

Table 28.1 (Continued)

Entry	Structure	Compound #	AChE (K_i μ M) ^{a)}	BuChE (K_i μ M) ^{a)}	Molecular volume ($\text{\AA}^3$)
9		9	None	5.8 ± 0.6	314
10		10	None	0.088 ± 0.0007	383
11		11	None	0.47 ± 0.03	379
12		12	None	0.0035 ± 0.0006	410
13		13	None	None	389
14		14	None	1.7 ± 0.4	430

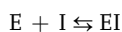
(Continued)

Table 28.1 (Continued)

Entry	Structure	Compound #	AChE (K_i μ M) ^{a)}	BuChE (K_i μ M) ^{a)}	Molecular volume ($\text{\AA}^3$)
15		15	None	0.40 ± 0.03	383
16		16	None	1.7 ± 0.2	387
17		17	None	0.22 ± 0.4	392

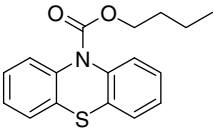
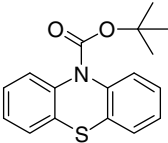
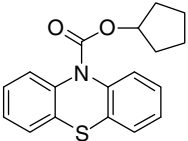
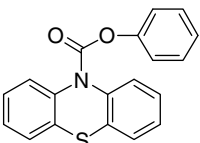
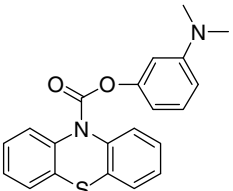
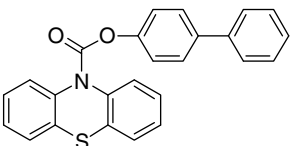
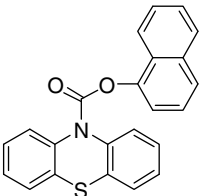
a) None refers to no detectable inhibitory activity up to the solubility limit of these compounds (3.33×10^{-5} – 1.67×10^{-4} M).

access to the enzyme. In general, a good reversible inhibitor is a molecule that can bind specifically, and with considerable affinity, to a target enzyme through electrostatic interactions that permit rapid association–disassociation of the enzyme–inhibitor complex (EI):



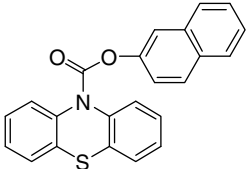
The relative potency of a reversible inhibitor can be quantified in several ways based on the effect produced by the inhibitor on the rate of substrate conversion to product in the enzyme-catalyzed reaction. One common method for evaluating inhibitor potency involves developing a dose–response curve that involves examining the effect of inhibitor concentration on the rate of product formation at a fixed concentration of substrate and enzyme [22]. The inhibitor dose that gives half-maximum inhibition of the enzyme-catalyzed reaction (IC_{50} value) can be estimated from such a plot and is a relative measure of inhibitor potency that is especially useful in assays involving crude biological systems. One disadvantage of the dose–response approach is that, although it provides a rapid means of comparing inhibitor potencies for a given enzyme, it gives little additional information, such as the type of reversible inhibition (competitive, noncompetitive, etc.) and what substrate/product kinetic parameters are being affected.

Table 28.2 Inhibition constants (K_i), deactivation constants (k_a) and molecular volumes for *N*-10-phenothiazine carbamates.^{a)}

Entry	Structure	Compound #	$10^{-3}\text{AChE } k_a$ ($\text{M}^{-1} \text{min}^{-1}$)	BuChE K_i (μM)	Molecular volume ($\text{\AA}^3$)
1		18	0.135 ± 0.064	13.4 ± 1.3	305
2		19	0.219 ± 0.08	4.3 ± 0.9	305
3		20	0.492 ± 0.185	3.19 ± 1.0	312
4		21	46.0 ± 12.0	0.57 ± 0.1	316
5		22	$19\ 200.0 \pm 1.201$	1.91 ± 0.14	366
6		23	910.0 ± 326.9	No inhibition ^{b)}	400
7		24	1650.0 ± 51.3	0.036 ± 0.01	367

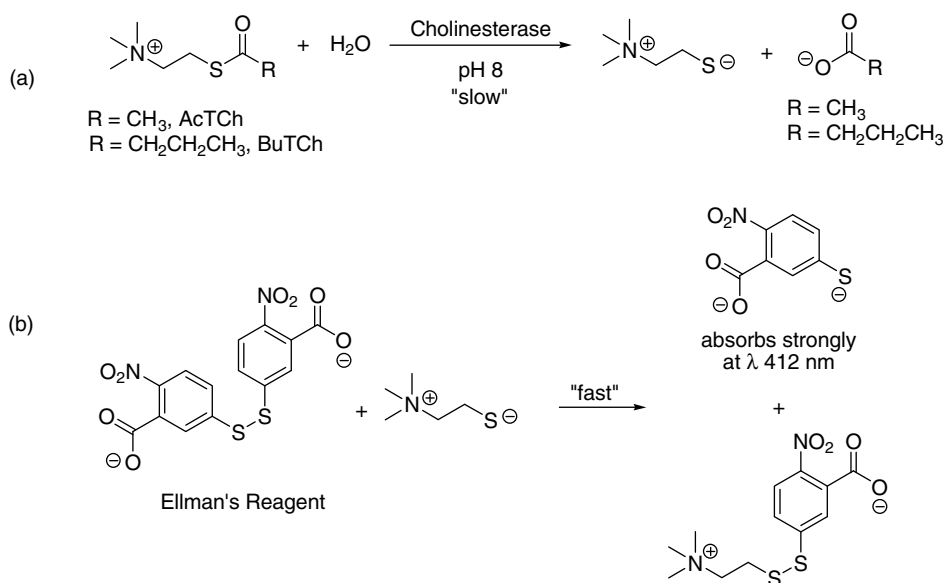
(Continued)

Table 28.2 (Continued)

Entry	Structure	Compound #	$10^{-3}\text{AChE } k_a$ ($\text{M}^{-1} \text{min}^{-1}$)	BuChE K_i (μM)	Molecular volume ($\text{\AA}^3$)
8		25	11.2 ± 0.96	0.12 ± 0.008	367

- a) k_a values for inhibition by rivastigmine: [3] AChE = $1.13 \times 10^3 \text{ M}^{-1} \text{min}^{-1}$; BuChE = $129 \times 10^3 \text{ M}^{-1} \text{min}^{-1}$. K_i values for inhibition by donepezil: [3] AChE = $0.024 \mu\text{M}$; BuChE = $2.21 \mu\text{M}$. K_i values for inhibition by galantamine: [3] AChE = $0.52 \mu\text{M}$; BuChE = $2.09 \mu\text{M}$.
- b) No inhibition detected up to $[I] = 2.36 \times 10^{-5} \text{ M}$, the solubility limit.

Another approach involves a more detailed examination of the effect of inhibitor concentration on the conversion of substrate into product under conditions that allow a fixed amount of the enzyme to obey Michaelis–Menten kinetics [23]. This involves varying substrate concentration and measuring the rate of product formation at several different inhibitor concentrations and employing a linear relationship of the observed data such as the Lineweaver–Burk double reciprocal plot (Figure 28.4) [24]. A re-plot of the slopes of these lines against inhibitor concentration gives the dissociation equilibrium constant for the enzyme–inhibitor complex ($K_i = [E][I]/[EI]$)



Scheme 28.3 Ellman's method for detecting cholinesterase catalyzed hydrolysis of choline thioesters.

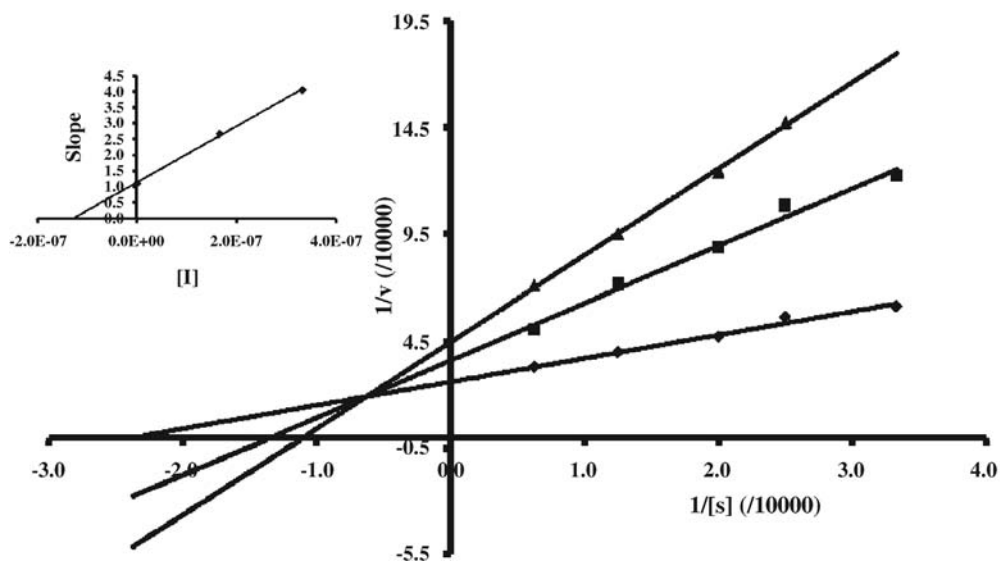


Figure 28.4 Lineweaver–Burk plot for naphthalen-1-yl(10*H*-phenothiazin-10-yl)methanone (**10**) of BuChE hydrolysis of butyrylthiocholine in the absence ($\blacklozenge$) and presence ($\blacksquare = 9.7 \times 10^{-8}$ M and $\blacktriangle = 1.9 \times 10^{-7}$ M) of compound **11**. Replot of the slopes versus inhibitor concentration (M) gave the inhibitor constant (K_i value M) as the x -intercept (inset).

as the x -intercept (Figure 28.4, inset). By definition the smaller the K_i value the greater the affinity for the enzyme. This treatment provides a more accurate determination of the inhibitor concentration producing half-maximum inhibition, because of the linear relationship, and also indicates the type of reversible inhibition (mixed noncompetitive in Figure 28.4) and whether the inhibitor affects substrate affinity (K_m), the maximum velocity (V_{max}) or both, as seen in Figure 28.4. Examples of K_i values for comparison of potencies of reversible phenothiazine inhibitors are given in Tables 28.1 and 28.2.

A second cholinesterase inhibitor class, pseudo-irreversible inhibitors, consists of molecules that show a time-dependent deactivation of the enzyme, most often by acylation of the catalytic site serine. The most common cholinesterase inhibitors that fall into this category are carbamates, like physostigmine and rivastigmine (Figure 28.5). Such compounds form a covalent carbamylated intermediate that is

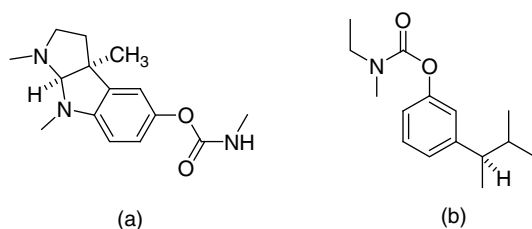
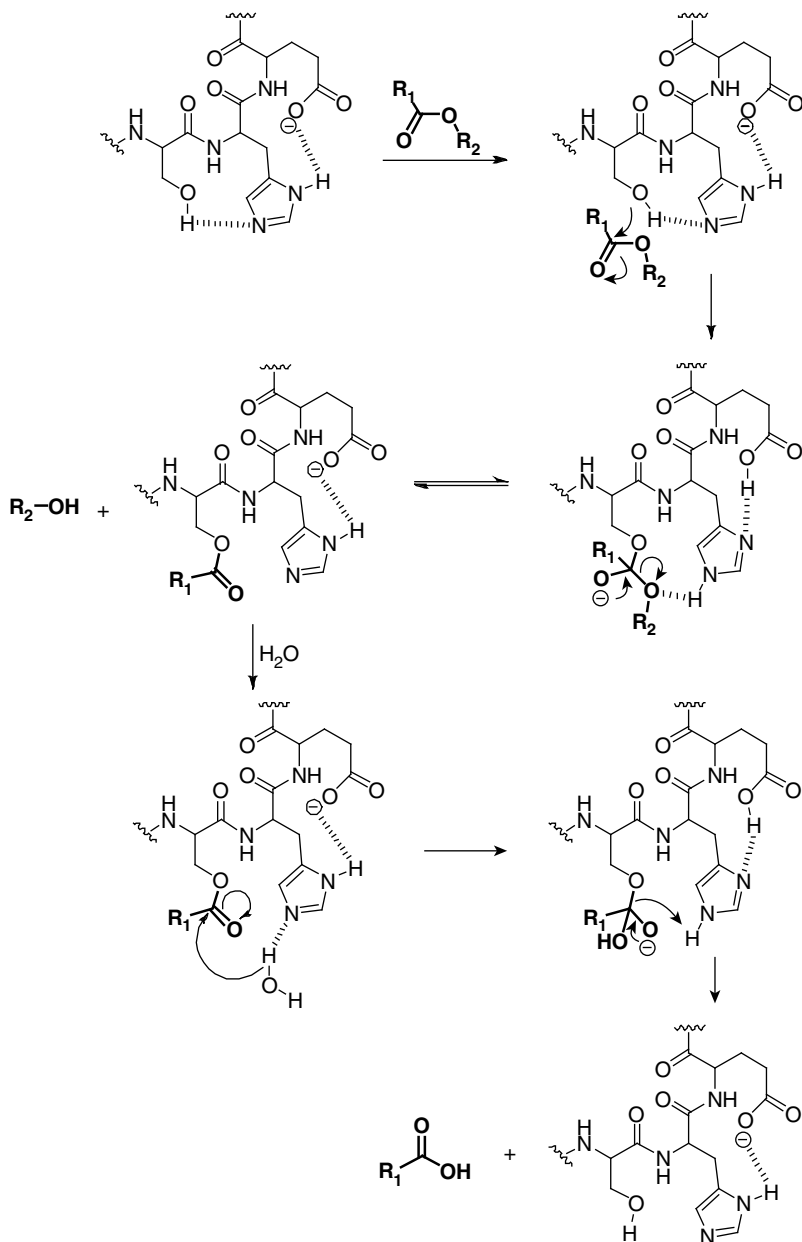


Figure 28.5 Carbamate cholinesterase inhibitors: (a) physostigmine and (b) rivastigmine.

hydrolyzed to regenerate the enzyme active site in a manner comparable to substrate ester hydrolysis (Scheme 28.4). However, with carbamates, deacylation occurs much more slowly so the enzyme is deactivated for an extended period of time.



Scheme 28.4 Acylation and deacylation of the active site serine during catalysis by cholinesterases.

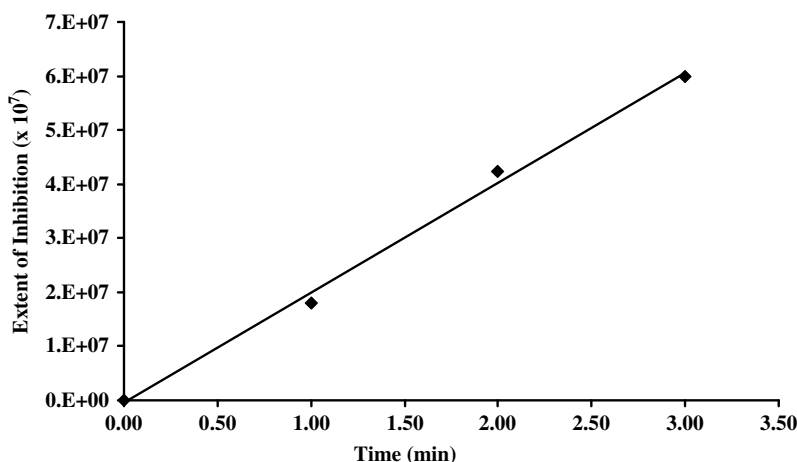


Figure 28.6 Plot to determine the second-order rate constant (k_a) for deactivation of acetylcholinesterase by 3-*N,N*-dimethylphenyl carbamate **22**.

Comparative potencies of pseudo-irreversible inhibitors are also often expressed as relative IC_{50} values. Although, under the same assay conditions, this dose–response approach provides a tool for comparison, it does not incorporate the very important parameter of variable time for enzyme deactivation for individual inhibitors. A more definitive quantification of the action of this carbamate “inhibitor–substrate” on acetylcholinesterase is the calculation of the second-order rate constant for enzyme deactivation as described by Dixon and Webb [25] and Reiner and Radić [26]. In this method, a plot of $\ln(e_0/e_t)/[I]$ against time gives a linear relationship and the slope of this line is the second-order rate constant (k_a) for deactivation of the enzyme over time (Figure 28.6). In this case, a larger k_a indicates more rapid enzyme deactivation and hence a more potent inhibitor. In such a plot, e_0 is the enzyme activity at $t=0$ with no pre-incubation of the enzyme with the inhibitor before initiating reaction, e_t is the enzyme activity after an appropriate time (t) of pre-incubation of the enzyme with inhibitor before initiating reaction and $[I]$ is the concentration of the inhibitor used. Relative strengths of phenothiazine carbamate pseudo-reversible cholinesterase inhibitors can be assessed by comparing k_a values in Table 28.2.

28.5

Computation of Physical Parameters to Interpret Structure–Activity Relationships

To facilitate the interpretation of relationships between biological evaluations and inhibitor structures, several physical parameters were computed for phenothiazine and its derivatives. These characteristics included the shape of the phenothiazine tricyclic ring system, the total volume of the compound, the length and width of

particular substituents attached to the phenothiazine moiety and the juxtaposition of certain side chain groups and the phenothiazine tricycle. The rationale for computing these parameters will be given below.

Phenothiazine (Figure 28.3), the scaffold moiety of the compounds studied here, consists of two aromatic rings linked through two atoms (nitrogen and sulfur) that are sp^3 hybridized and permit flexibility to the tricyclic ring system through these atoms. The overall non-planar nature of the complete tricycle has been described as having the shape of a butterfly [28]. Substitution of the hydrogen atom on the nitrogen of the central ring by other substituents may influence the shape of the phenothiazine moiety and affect its properties as an initiator. Therefore calculations were made to determine the "butterfly angle," defined as the angle of fold between the two aromatic rings of phenothiazine, to see whether this parameter could play a role in cholinesterase inhibition.

Preliminary results were obtained using the MMFF94 force field [29]. However, Hartree–Fock and density functional theory calculations were carried out on a few of the simpler derivatives to assess the efficacy of the force field method. The compounds phenothiazine and acetyl phenothiazine were the derivatives chosen for study at the Hartree–Fock/STO-3G and B3LYP/6-31G(d) level [30–32].

The Gaussian 98 suite of programs was employed [33]. Table 28.3 presents the butterfly angle results at various levels of theory. A butterfly angle of 150° is predicted for phenothiazine at the B3LYP/6-31G(d) level, whereas molecular mechanics (MM) predicts an angle of 180° . The corresponding results for acetyl phenothiazine are 135° and 159° , respectively. The B3LYP/6-31G(d) results obtained for phenothiazine were consistent with the work at the B3LYP level carried out by Palafox *et al.* [34]. While molecular mechanics is not able to reproduce the butterfly angles, it does reproduce the trend of butterfly angle decrease upon substitution, and thus represents a short-term compromise between accuracy and affordability for the larger systems. Butterfly angles for phenothiazine and the analogues studied at the molecular mechanics level varied from approximately 180° for phenothiazine itself to approximately 160° for most *N*-10-amides (Table 28.1), to 146° for the corresponding carbamate derivatives in Table 28.2. The computational results for the carbamate derivatives were obtained using the MMFF force field in Spartan '06 [29b].

Previous observations [27] have suggested that the large molecular volume of the phenothiazine derivative ethopropazine (Figure 28.3) is a significant factor in

Table 28.3 Comparison of butterfly angles calculated using different levels of molecular theory for phenothiazine and acetylphenothiazine.^{a)}

Method	Energy (hartree)	Butterfly angle ($^\circ$)
MM	—	180 (159)
HF/STO-3G	−901.0197(−1050.8336)	165 (147)
B3LYP/6-31G(d)	−915.6436(−1068.2925)	150 (135)

a) Results for acetylphenothiazine in parentheses.

determining its specificity towards BuChE; that is, the active site gorge of BuChE ($V \sim 500 \text{ \AA}^3$) [9], but not of AChE ($V = 300 \text{ \AA}^3$), can accommodate this molecule. Molecular volumes were calculated by selecting a surface of fixed electron density ($0.001 \text{ e bohr}^{-3}$), and computing the volume within that isodensity surface [35]. *Ab initio* methods were employed to obtain an electron density surface. Accordingly, single-point HF/STO-3G calculations were carried out starting from the atomic coordinates obtained at the molecular mechanics level. Molecular volumes were computed as averages of five separate calculations. Molecular volumes computed in this manner, based on an average of five separate calculations, are summarized in Tables 28.1 and 28.2 and range from $212 \text{ \AA}^3$ for phenothiazine itself to $410 \text{ \AA}^3$ for the 9-anthrylcarbonyl amide derivative **12** (Table 28.1). This parameter was found to be a major determinant for BuChE-specificity and inhibitor potency for most derivatives of phenothiazine, as seen with the four aryl amides (**9**–**12**, Table 28.1) in which a 1000-fold increase in inhibitor potency occurs in direct relationship with an inhibitor volume increase from $314 \text{ \AA}^3$ (**9**) to $410 \text{ \AA}^3$ (**12**).

Once in the active site of the enzyme, the compounds are able to form strong π – π interactions between the phenothiazine and aromatic amino acid residues (F329 and Y332; see Figure 28.1) near the mouth of the BuChE active site gorge, reducing the enzyme's ability to hydrolyze substrate. The substituent groups on these inhibitors act like a plug and the end result is to block off the approach of the substrate to the catalytic site. The larger the substituent group (anthryl > naphthyl > phenyl in Figure 28.7) the better its ability to block the active site and prevent the substrate

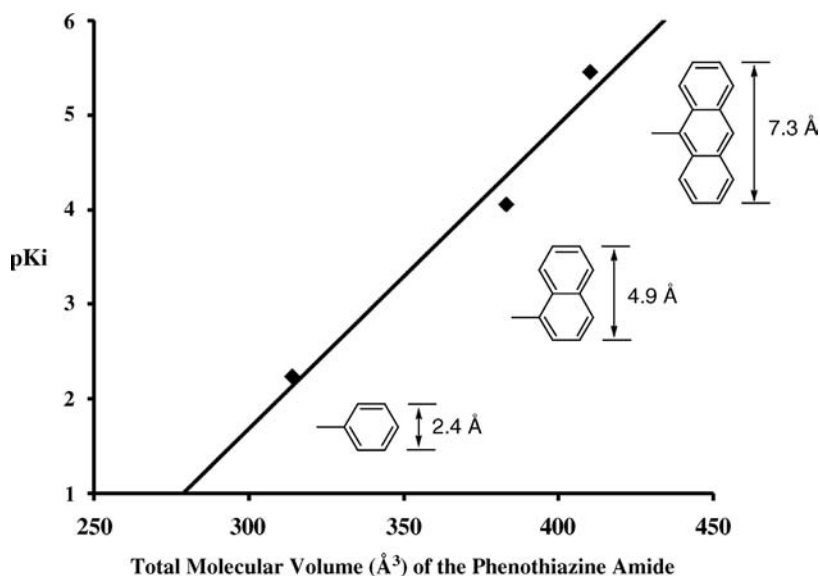


Figure 28.7 Total volume and substituent size for phenothiazine amides as factors governing potency of butyrylcholinesterase inhibition by phenyl (**9**), naphthyl (**10**) and anthryl (**12**) phenothiazine amides.

from reaching the catalytic triad. Thus, the bulky substituent of the 9-anthrylcarbonyl phenothiazine amide derivative produces a very potent inhibition of BuChE ($K_i = 3.5 \times 10^{-9}$ M) [20]. This system was found, in its preferred conformation, to conform quite well to the active site gorge of BuChE [20], its substituent effectively blocking substrate access to the mouth of the gorge.

Measurement of other factors such as length and width of aryl substituents, shape, flexibility and angle between the phenothiazine moiety and substituents were determined from the preferred conformation of the relevant compounds.

28.6

Enzyme–Inhibitor Structure–Activity Relationships

Phenothiazine and small alkyl derivatives, such as *N*-10-methylphenothiazine (Figure 28.3), inhibit BuChE catalysis but have no effect on AChE catalysis, despite the fact that the molecular volumes of these compounds are smaller (Table 28.1) than the estimated active site gorge volume of AChE (302 Å³) [9]. This specificity for binding only to BuChE has been attributed to an interaction involving π – π stacking of the aromatic rings of the phenothiazine tricycle with two aryl side chains (F 329 and Y332) in the BuChE active site gorge (Figure 28.1) [20]. Although comparable residues occur in the AChE gorge (F338 and Y341), another aryl residue, Y337, interferes with phenothiazine binding [9] in this case (Figure 28.1).

In contrast to these *N*-10-alkyl-phenothiazine derivatives, the synthesized *N*-10 amides did show some ability to reversibly inhibit AChE. This modest inhibition (Table 28.1) could be attributed to electrostatic interaction of the substituent carbonyl group, such as through hydrogen bonding within the AChE active site gorge [19]. The involvement of ligand formation within the active site gorge in the inhibition of enzyme activity was suggested by the fact that the ability to inhibit was lost when the total molecular volume of the inhibitor exceeded the AChE gorge volume (~ 300 Å³). As can be seen in Table 28.1, amide derivatives up to and including the pivalyl derivative (**4**) (total molecular volume 308 Å³) inhibited AChE. That this binding to the enzyme was through the substituent carbonyl group, and not the phenothiazine tricycle, and that substrate binding was affected by the unbound phenothiazine tricycle, was indicated by a virtually constant inhibitor potency for AChE until inhibition is lost (Table 28.1) through the molecular volume limitation of the AChE active site gorge.

Antithetic to the restrictions observed for phenothiazine amides and their ability to inhibit AChE, reversible inhibition of BuChE was essentially in direct relationship with the total molecular volume of the alkyl amides (Figure 28.3, Table 28.1) [19, 20], all of which are smaller than the active site gorge volume (≈ 500 Å³) of BuChE. In this case the binding of the phenothiazine moiety to F329 and Y332 produces a very noticeable effect related to the size and orientation of the substituent on the phenothiazine scaffold, in addition to the total molecular volume of the inhibitor. Figure 28.7 illustrates a dramatic 1000-fold increase in inhibitor potency (Table 28.1)

as the substituent linked to the carbonyl group changes from a phenyl group through the bicyclic naphthyl to the tricyclic anthryl substituent. Since all three inhibitors differ only in the substituent ring system, the increasing width of the planar rings (Figure 28.7) from mono- (width = 2.4 Å) to a tricyclic system (width = 7.3 Å) must have a profound effect in blocking substrate access to the catalytic serine. In addition, that orientation and length of the substituent can influence BuChE inhibition is indicated by the superior inhibition exhibited by the 1-naphthoyl (**10**) over the 2-naphthoyl phenothiazine amide derivative (**11**) (Table 28.1), even though their molecular structures and volumes are comparable (Table 28.1). This substituent effect is further emphasized by the complete lack of BuChE inhibition by the biphenyl carbonyl derivative (**13**) (Table 28.1) that is an analog of the potent naphthoyl derivatives. This complete lack of BuChE inhibition by the biphenyl carbonyl amide, despite a total volume comparable to the naphthoyl counterpart, can be attributed to the long and rigid nature of the conjugated biphenyl amide system that is almost perpendicular to the phenothiazine tricycle (Figure 28.8a). This would prevent the usual alignment of the phenothiazine moiety with F329 and Y332 in BuChE, because of interaction of the substituent with the opposite rim (Figure 28.1) of the active site gorge. This notion is supported by the observation that the larger, but more flexible, biphenylacetyl phenothiazine amide (**14**) (Figure 28.8a) provides fairly robust inhibition of this enzyme (Table 28.1).

Also of note, with respect to the putative BuChE requirement for binding the phenothiazine moiety to F329 and Y332, was the effect of positioning the phenyl group in a series (compounds **15**–**17**) of phenylbutanoyl amide derivatives of phenothiazine (Table 28.1). In the 3-position of the alkyl chain, the preferred conformation of the substituent phenyl ring is in close proximity to an aryl ring of the phenothiazine tricycle (Figure 28.8b). The separation between geometric ring centers is 4.2 Å. This leads to a weaker inhibitor strength than when the phenyl group is in the 2- (4.8 Å) or 4- (5.9 Å) position on the alkyl chain (Figure 28.8b; Table 28.1), presumably because of interference by the more favored intramolecular π – π stacking by the phenyl group in the 3-position that weakens the usual enzyme–inhibitor π – π intermolecular interaction.

Investigation of a second series of *N*-10-carbonyl derivatives, the phenothiazine carbamates (Table 28.2) [11], has provided a different perspective and further insight into the inhibition of cholinesterases by derivatives of phenothiazine.

Carbamates are generally known to be pseudo-irreversible inhibitors of cholinesterases, acting as substrate analogs that react with the catalytic serine residue, forming an acylated intermediate that is slow to hydrolyze, compared to choline ester substrate intermediates (Scheme 28.4), and is therefore able to deactivate the enzyme for an extended period of time [11]. This effect is typified by the action of carbamates such as physostigmine and rivastigmine (Figure 28.5), which can produce a time-dependant deactivation of both BuChE and AChE [3]. This type of inhibition was found to occur for AChE in the presence of carbamate derivatives of phenothiazine (Table 28.2, Figure 28.6).

In contrast to this effect on AChE, the carbamate species derived from phenothiazine-10-carbonyl chloride and various alcohols and phenols (Scheme 28.2) [11]

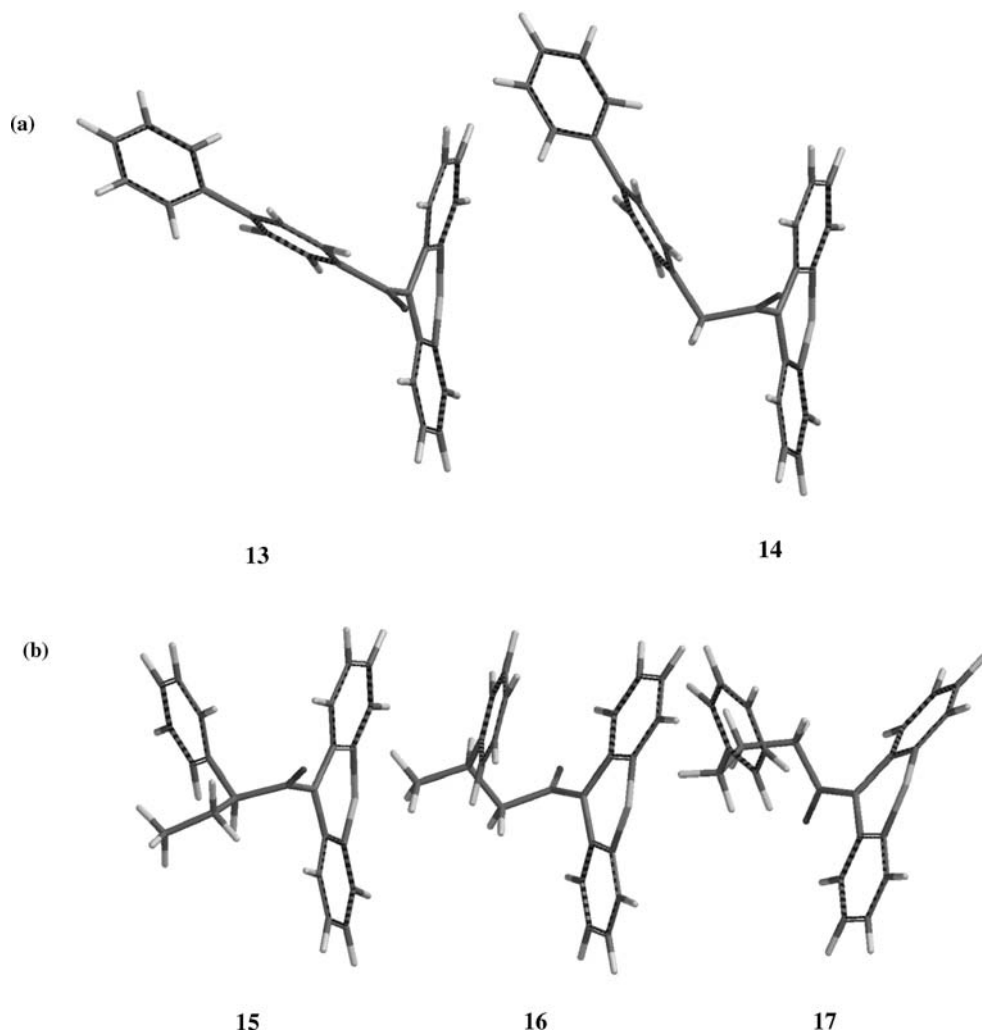


Figure 28.8 (a) Preferred conformations of biphenyl-4-yl(10*H*-phenothiazin-10-yl) methanone (**13**) and 2-(biphenyl-4-yl)-1-(10*H*-phenothiazin-10-yl)ethanone (**14**), indicating the importance of substituent flexibility in governing butyrylcholinesterase inhibition. The rigid nature of **13** prevents inhibition of butyrylcholinesterase while the flexible nature of **14** renders it a potent inhibitor. (b) Preferred conformations of 1-(10*H*-phenothiazin-10-yl)-2-phenylbutan-1-one (**15**), 1-(10*H*-phenothiazin-10-yl)-3-phenylbutan-1-one (**16**) and 1-(10*H*-phenothiazin-10-yl)-4-phenylbutan-1-one (**17**). The proximity of the phenyl substituent in **16** to the phenothiazine tricycle produces intramolecular hydrogen bonding, diminishing inhibitor potency relative to **15** and **17**.

exhibited reversible inhibition of BuChE, as seen with the amide derivatives above, rather than the pseudo-irreversible inhibition of BuChE expected for carbamates. This unusual phenomenon for carbamates has been attributed to a highly preferred π - π interaction between the aromatic rings of the phenothiazine moiety and the

previously described aromatic residues (F329 and Y332) within the BuChE active site gorge. These aryl residues are part of a helical segment of the enzyme (the E-helix) that also contains E325 of the catalytic triad (Figure 28.1) [11]. Converting either of these residues (F329 or Y332) into aliphatic residues by site-directed mutation alters the reversible inhibition seen with wild-type BuChE by phenothiazine carbamate derivatives, such as the 3-*N,N*-dimethylaminophenyl derivative (22), to the more typical pseudo-irreversible inhibition by carbamates (Figure 28.9). Furthermore, BuChE mutations between the two aromatic residues (F329 and Y332), converting A328 into an aryl residue (e.g., A328Y), make this region of BuChE more like AChE (Figure 28.1). Y337 in AChE interferes with reversible binding of phenothiazine to F338 and Y341 in that enzyme [9], and permits delivery of the phenothiazine carbamate group to S203 for time-dependant deactivation of AChE (Table 28.2). Similarly, BuChE mutants such as A328Y permit pseudo-irreversible inhibition of the mutant enzyme (Figure 28.9) by carbamates such as compound (22). To confirm that the binding of phenothiazine carbamates to BuChE involves π – π interaction comparable to the phenothiazine amides, the interaction of the human wild-type BuChE with the potent 9-anthryl carbonyl phenothiazine amide [20] ($K_i = 3.5 \times 10^{-9}$ M, Table 28.1, 12) was compared to the mutant BuChE species, Y332A, which should not be able to bind reversibly to the

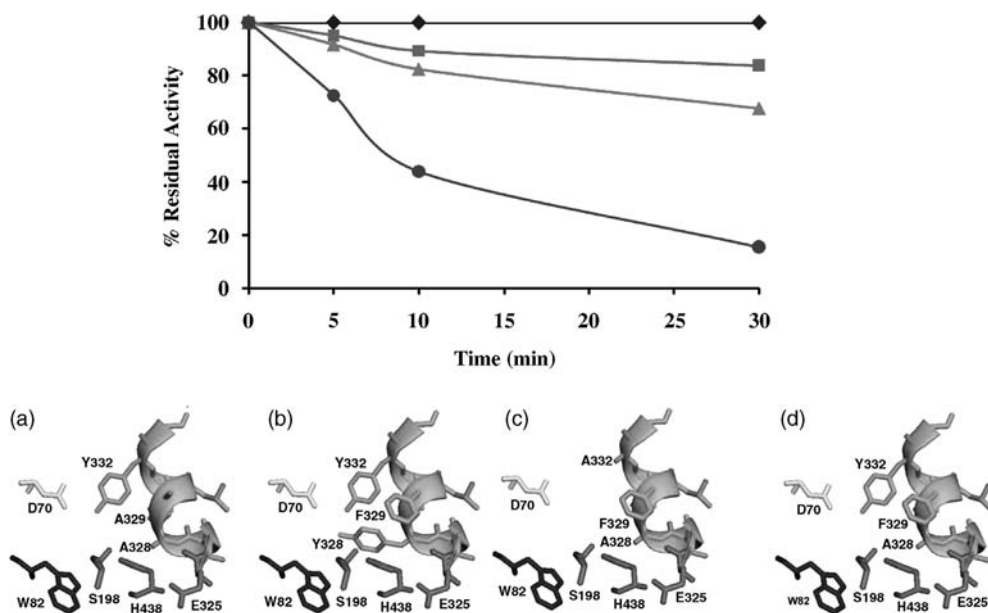


Figure 28.9 Residual enzymatic activity over time for (a) wild-type and mutant butyrylcholinesterases ((b) Y332A, (c) A328Y, (d) F329A) in the presence of 3-*N,N*-dimethylaminophenyl phenothiazine carbamate (22). The figures for the active site gorge were

generated using crystal structure coordinates of butyrylcholinesterase [8] from the Protein Data Bank [36], using The PyMOL Molecular Graphics System [37] (◆) wild-type BuChE, (■) Y332A, (▲) A328Y, (●) F329A.

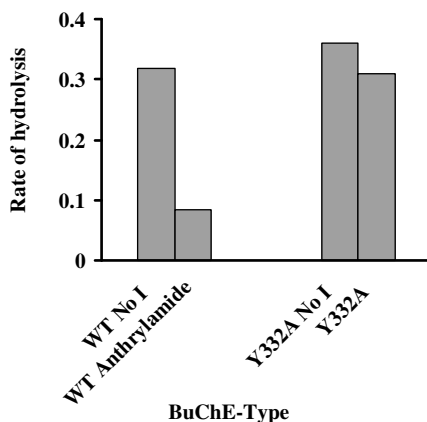


Figure 28.10 Rate of hydrolysis ($\Delta A \text{ min}^{-1}$) of butyrylthiocholine by wild-type and Y332A butyrylcholinesterase in the absence and presence of 9-anthryl carbonyl amide (**12**).

phenothiazine tricycle in an efficient manner. At the same concentration of inhibitor ($2.9 \times 10^{-8} \text{ M}$), wild-type native human BuChE is strongly inhibited by amide **12**, while the mutant enzyme (Y332A) is only slightly inhibited by the same amide (Figure 28.10).

In contrast to the action of phenothiazine carbamates, which produce only reversible inhibition of wild-type human BuChE, the effect of the same compounds on AChE is the more common pseudo-irreversible inhibition expected for carbamate action on cholinesterases (Table 28.2). As stated above, the inability of AChE to reversibly bind the phenothiazine moiety in π - π association has been attributed [9] to interference of this binding by a tyrosine residue (Y337) in AChE that is not present in BuChE (Figure 28.1). The consequence of this is that with AChE the carbamate functionality of the inhibitor can be delivered to the catalytic serine (S203) for carbamylation of the enzyme (Scheme 28.4). Nor does the inhibitor molecular volume limitation ($\sim 310 \text{ \AA}^3$) appear to have the importance observed in the phenothiazine amide inhibition of AChE described earlier [19, 20], and summarized above. For example, 4-biphenyl phenothiazine carbamate (**23**) (volume = $400 \text{ \AA}^3$), like the amide equivalent, shows no BuChE reversible inhibition, ostensibly because the length of the rigid substituent group interferes with π - π binding of the phenothiazine moiety to F329 and Y332 of the enzyme (Figure 28.8a). This is despite the fact that the molecular volume of this compound is less than the BuChE active site gorge volume ($502 \text{ \AA}^3$). Why this carbamate does not show a time-dependant inhibition of BuChE if it cannot effect the usual reversible π - π stacking in BuChE remains unclear since, despite its total molecular volume exceeding the AChE active site gorge volume ($302 \text{ \AA}^3$), it still shows pseudo-irreversible inhibition of that enzyme. The 4-biphenylcarbamate of phenothiazine (**23**) is a specific and highly potent inhibitor of AChE (Table 28.2), with a second-order rate constant for deactivation (k_a value) of $9.1 \times 10^5 \text{ M}^{-1} \text{ min}^{-1}$. This suggests that the entire phenothiazine carbamate need not enter the

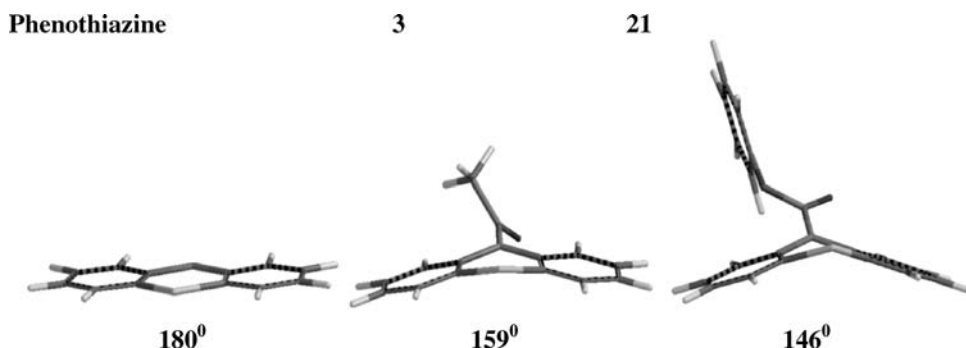


Figure 28.11 Calculated butterfly angles for phenothiazine and its amide derivative **3** and carbamate derivative **21**.

AChE active site gorge to permit carbamylation of the catalytic serine (S203). Another factor that could facilitate, at least partially, entry of the phenothiazine moiety into the AChE gorge is an unusual property related to the carbonyl group of the carbamates relative to that of the comparable amides. In the infrared spectrum, the carbonyl stretch of the phenothiazine carbamates is roughly 1730 cm^{-1} [11], typical of carbonyls. In contrast, phenothiazine amides exhibit a carbonyl stretch at about 1680 cm^{-1} [19, 20], which is characteristic of amide carbonyls that are considered to have more N=C character and less C=O character. A consequence of this difference is reflected in the “butterfly angle” (amides $\sim 160^\circ$, carbamates $\sim 145^\circ$) [11, 19] of the phenothiazine tricycle being changed by the small difference in the hybridization of the nitrogen atom within the phenothiazine scaffold [11, 19]. A smaller butterfly angle produces a more compact phenothiazine ring system in the carbamate derivatives (Figure 28.11) relative to amides and unsubstituted phenothiazine and this could facilitate delivery of the carbamate functionality down the narrow AChE gorge for reaction at S203 of AChE.

28.7

Conclusions

Structure–activity comparison of the inhibitory properties of several phenothiazine amides and carbamates, synthesized through the *N*-10 position of the phenothiazine scaffold, has provided insights into how the structures of such molecules effect selective inhibition of the cholinesterases. Most of the phenothiazine amides examined proved to be selective reversible inhibitors of BuChE. While none of the phenothiazine carbamates were pseudo-irreversible inhibitors of wild-type BuChE, because of a favored electrostatic interaction with the enzyme, a number of these derivatives were found to be selective pseudo-irreversible inhibitors of AChE. On the other hand, the absence of serum BuChE hydrolysis of phenothiazine carbamates

promises abundant intact inhibitor concentration reaching the brain for AD treatment with these reversible inhibitors.

AChE inhibition by phenothiazine amides exhibits strict size limits that are closely related to the volume of the active site gorge ($\sim 300 \text{ \AA}^3$) for this enzyme. The mechanism of binding between inhibitor and AChE has not been clearly defined but appears to involve the carbonyl function of the substituent, possibly hydrogen bonded to the enzyme.

In contrast, BuChE inhibition involves π - π interaction of the aromatic rings of phenothiazine to F329 and Y332 within the wild-type BuChE active site gorge. This binding not only blocks substrate access to the catalytic site, through substituents attached to the phenothiazine-*N*-10 position, but may also produce conformational changes at the catalytic site since E325 of the triad is on the same flexible helical segment (E-helix). Displacement of E325 in this way could directly affect the efficiency of ester hydrolysis. The importance of the E-helix in the reversible inhibition of BuChE by phenothiazine can be illustrated by the loss of such inhibition with site-directed mutations of BuChE (e.g., Y332A). Phenothiazine derivatives can be designed to be specific inhibitors of BuChE, such as compound **12** (Table 28.1) or AChE, such as compound **23** (Table 28.2).

Structure-activity relationships of phenothiazine derivatives have shown that molecules with this scaffold can be designed to have differential inhibition effects on BuChE and AChE. This has the potential for the development of disease-modifying drugs for the treatment of Alzheimer's disease.

Acknowledgments

The authors would like to thank the Canadian Institutes of Health Research, Vascular Health and Dementia Initiative (DOV-78 344) (through partnership of Canadian Institutes of Health Research, Heart & Stroke Foundation of Canada, the Alzheimer Society of Canada and Pfizer Canada Inc.), the Natural Sciences and Engineering Research Council of Canada, the Canadian Foundation for Innovation, Capital District Health Authority Research Fund, the Nova Scotia Health Research Foundation, the Alzheimer Society of Nova Scotia, the Brain Tumor Foundation of Canada, Mount Saint Vincent University, and the MS Society of Canada for funding. Purified wild-type and mutant BuChE samples were a gift from Dr Oksana Lockridge (University of Nebraska Medical Center). We thank Andrea LeBlanc, Andrew Reid, Jillian Soh and Ryan Walsh for valuable technical assistance.

References

- 1 Massoulie, J. (2002) The origin of the molecular diversity and functional anchoring of cholinesterases. *Neurosignals*, **11**, 130–143.
- 2 Giacobini, E. (2000) Cholinesterase inhibitors: from the Calabar bean to Alzheimer therapy. *Cholinesterases Cholinesterase Inhibitors*, Martin Dunitz, Ltd., London, 181–226.
- 3 Darvesh, S., Hopkins, D.A., and Geula, C. (2003) Neurobiology of butyrylcholinesterase. *Nat. Rev. Neurosci.*, **4**, 131–138.

- 4 Glick, D. (1942) Specificity studies on enzymes hydrolyzing esters of substituted amino and nitrogen heterocyclic alcohols. *J. Am. Chem. Soc.*, **64**, 564–567.
- 5 Gomori, G. (1948) Histochemical demonstration of sites of choline esterase activity. *Proc. Soc. Exp. Biol. Med.*, **68**, 354–358.
- 6 Darvesh, S., McDonald, R.S., Darvesh, K.V., Mataija, D., Mothana, S., Cook, H., Carneiro, K.M., Richard, N., Walsh, R., and Martin, E. (2006) On the active site for hydrolysis of aryl amides and choline esters by human cholinesterases. *Bioorg. Med. Chem.*, **14**, 4586–4599.
- 7 Sussman, J.L., Harel, M., Frolov, F., Oefner, C., Goldman, A., Tokar, L., and Silman, I. (1991) Atomic structure of acetylcholinesterase from *Torpedo Californica*: a prototypic acetylcholine-binding protein. *Science*, **253**, 872–879.
- 8 Nicolet, Y., Lockridge, O., Masson, P., Fontecilla-Camps, J.C., and Nachon, F. (2003) Crystal structure of human butyrylcholinesterase and of its complexes with substrate and products. *J. Biol. Chem.*, **278**, 41141–41147.
- 9 Saxena, A., Redman, A.M.G., Jiang, X., Lockridge, O., and Doctor, B.P. (1997) Differences in active site gorge dimensions of cholinesterases revealed by binding of inhibitors to human butyrylcholinesterase. *Biochemistry*, **36**, 14642–14651.
- 10 Silver, A. (1974) *The Biology of Cholinesterases*, *Frontiers of Biology*, vol 36, Elsevier, Amsterdam.
- 11 Darvesh, S., Darvesh, K.V., McDonald, R.S., Mataija, D., Walsh, R., Mothana, S., Lockridge, O., and Martin, E. (2008) Carbamates with differential mechanism of inhibition toward acetylcholinesterase and butyrylcholinesterase. *J. Med. Chem.*, **51**, 4200–4212.
- 12 Birks, J. (2006) Cholinesterase inhibitors for Alzheimer's disease. *Cochrane Database of Systematic Reviews*, Issue 1, Art. No. CD005593, DOI: 10.1002/14651858.CD005593.
- 13 Darvesh, S., MacKnight, C., and Rockwood, K. (2001) Butyrylcholinesterase and cognitive function. *Int. Psychogeriatr.*, **13**, 461–464.
- 14 Darvesh, S., Grantham, D.L., and Hopkins, D.A. (1998) Distribution of butyrylcholinesterase in the human amygdala and hippocampal formation. *J. Comp. Neurol.*, **393**, 374–390.
- 15 Mesulam, M., Guillozet, A., Shaw, P., Levey, A., Duysen, E.G., and Lockridge, O. (2002) Acetylcholinesterase knockouts establish central cholinergic pathways and can use butyrylcholinesterase to hydrolyze acetylcholine. *Neuroscience*, **110**, 627–639.
- 16 Mesulam, M.M. and Geula, C. (1994) Butyrylcholinesterase reactivity differentiates the amyloid plaques of aging from those of dementia. *Ann. Neurol.*, **36**, 722–727.
- 17 Guillozet, A.L., Smiley, J.F., Mash, D.C., and Mesulam, M.M. (1997) Butyrylcholinesterase in the life cycle of amyloid plaques. *Ann. Neurol.*, **42**, 909–918.
- 18 Radić, Z., Pickering, N.A., Vellom, D.C., Camp, S., and Taylor, P. (1993) Three distinct domains in the cholinesterase molecule confer selectivity for acetyl- and butyrylcholinesterase inhibitors. *Biochemistry*, **32**, 12074–12084.
- 19 Darvesh, S., McDonald, R.S., Penwell, A., Conrad, S., Darvesh, K.V., Mataija, D., Gomez, G., Caines, A., Walsh, R., and Martin, E. (2005) Structure-activity relationships for inhibition of human cholinesterases by alkyl amide phenothiazine derivatives. *Bioorg. Med. Chem.*, **13**, 211–222.
- 20 Darvesh, S., McDonald, R.S., Darvesh, K.V., Mataija, D., Conrad, S., Gomez, G., Walsh, R., and Martin, E. (2007) Selective reversible inhibition of human butyrylcholinesterase by aryl amide derivatives of phenothiazine. *Bioorg. Med. Chem.*, **15**, 6367–6378.
- 21 Ellman, G.L., Courtney, K.D., Andres, V. Jr, and Featherstone, R.M. (1961) A new and rapid colorimetric determination of acetylcholinesterase activity. *Biochem. Pharmacol.*, **7**, 88–95.
- 22 Copeland, R.A. (2000) *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis*, 2nd edn, Wiley-VCH Verlag GmbH, Weinheim.
- 23 Michaelis, L. and Menten, M.L. (1913) Kinetics of invertase action. *Biochem. Z.*, **49**, 333–369.

- 24 Lineweaver, H. and Burk, D. (1934) Determination of enzyme dissociation constants. *J. Am. Chem. Soc.*, **56**, 658–666.
- 25 Dixon, M., Webb, E.C., Thorne, C.J.R., and Tipton, K.F. (1979) *Enzymes*, 3rd edn, Academic Press, New York, 1115.
- 26 Reiner, E. and Radić, Z. (2000) Mechanism of action of cholinesterase inhibitors, in *Cholinesterases and Cholinesterase Inhibitors* (ed E. Giacobini), Martin Dunitz, Ltd, London, pp. 103–119.
- 27 Debord, J., Merle, L., Bollinger, J., and Dantoine, T. (2002) Inhibition of butyrylcholinesterase by phenothiazine derivatives. *J. Enzyme Inhib. Med. Chem.*, **17**, 197–202.
- 28 Ragg, E., Fronza, G., Mondelli, R., and Scapini, G. (1983) Carbon-13 nuclear magnetic resonance spectroscopy of nitrogen heterocycles. Part 4. Intra-extra configuration of the N-acetyl group in phenothiazine and related systems with a butterfly shape. *J. Chem. Soc., Perkin Trans. 2*, 1289–1292.
- 29 (a) Molecular modeling was carried out using the MMFF94 force field method, as part of the PC Spartan Pro Software: PC Spartan Pro, Wavefunction, Inc., 18401 Von Karman, Suite 370, Irvine, California, 92612; (b) (2006) *Spartan '06*, Wavefunction, Inc.: Irvine, CA.
- 30 Becke, A.D. (1993) A new mixing of Hartree-Fock and local-density-functional theories. *J. Chem. Phys.*, **98**, 1372–1377.
- 31 Becke, A.D. (1993) Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, **98**, 5648–5652.
- 32 Lee, C., Yang, W., and Parr, R.G. (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B: Condens. Matter*, **37**, 785–789.
- 33 Frisch, M.J. *et al.* (2001) *Gaussian 98, Revision A.11.2*. Gaussian, Inc., Pittsburgh PA.
- 34 Palafox, M.A., Gil, M., Nunez, J.L., and Tardajos, G. (2002) Study of phenothiazine and N-methyl phenothiazine by infrared, Raman, ¹H-, and ¹³C-NMR spectroscopies. *Int. J. Quantum Chem.*, **89**, 147–171.
- 35 Frisch, A.E. and Frisch, M.J. (1999) *Gaussian 98, User's reference*, 2nd edn, Gaussian Inc., Pittsburg, PA.
- 36 Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Research*, **28**, 235–242 <http://www.pdb.org/>.
- 37 DeLano, W.L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA, USA <http://www.pymol.org>.

29

Reduction Potentials of Peptide-Bound Copper (II) – Relevance for Alzheimer's Disease and Prion Diseases

Arvi Rauk

29.1

Introduction

Copper is a ubiquitous, redox-active metal in biological systems. It is the third most abundance transition element in the human body after iron and zinc. Up to 90% of it is strongly bound to ceruloplasmin, a metalloprotein whose principal function it is to oxidize Fe(II) to Fe(III) so that it can be transported by transferrin. Most of the remaining copper is loosely bound to albumin, and a smaller fraction still to small peptides, copper transporters and other metalloproteins.

Inorganic copper [1] in aqueous solution exists in predominantly two oxidation states, the more stable of which is Cu(II) with four covalently bound ligands arranged in a distorted square-planar arrangement, six ligands in a distorted octahedron or, less commonly, five ligands in a tetragonal pyramid. Cu(II) has a d^9 electronic configuration and is paramagnetic (open shell). The closed shell d^{10} oxidation state, Cu(I), is appreciably less stable in aqueous solution and occurs with two ligands in a linear arrangement, three ligands in a planar arrangement or four ligands in a tetrahedral arrangement. Aqueous Cu(II) is a mild oxidant, with standard a reduction potential $E^\circ(\text{Cu}_{(\text{aq})}^{2+}/\text{Cu}_{(\text{aq})}^{1+})$ of 0.159 V relative to the standard hydrogen electrode (SHE) [2].

Biological copper exists in predominantly two coordination environments, labeled Type 1 and Type 2 [3]. Type 1 copper occurs in the blue copper proteins, of which ceruloplasmin is one. Type 1 Copper has 3–5 coordination ligands. Three of the ligands are always a covalently bonded thiolate of cysteine, and the hetero-ring N atoms of two histidines, disposed in a distorted trigonal arrangement. If there is a fourth ligand, it is usually a methionine or glutamine side chain in an axial position. If there is a fifth ligand, it is usually a carbonyl group of the backbone. Reduction potentials of the blue copper proteins encompass a wide range (in V vs SHE): stellacyanin, 0.18; plastocyanin, 0.30; azurin, 0.38; rusticyanin, 0.68; ceruloplasmin, 1.00 [4].

Type 2 copper has four ligands in a tetragonal (square planar) arrangement and possibly a fifth ligand in an axial position. At least one and possibly all four of the equatorial ligands are histidine residues coordinated through one of the hetero-ring N atoms. Type 2 copper occurs in albumin, numerous oxidases and copper-zinc superoxide dismutase (CuZnSOD). Human serum albumin (HSA) complexed to Cu (II) has an extremely negative reduction potential, $E^\circ = -0.80$ V vs SHE, indicating that it is very difficult to reduce to Cu(I)/HSA [5]. In contrast, CuZnSOD has a much higher reduction potential, $E^\circ = 0.31$ V vs SHE. In the latter case, the reduction potential is dependent on pH and the symbol, E°' , indicates the value at pH 7.

Loss of copper homeostasis is associated with two well-defined disease states, Wilson's disease, caused by hepatic accumulation and toxicity of copper, and Menkes' disease, a severe form of mental retardation caused by a deficiency of copper. Copper is also implicated in several other neurodegenerative diseases, including familial amyotrophic lateral sclerosis, prion diseases and Alzheimer's disease.

For the purpose of discussion of the various influences on the reduction potential of "Cu(II)"/"Cu(I)" couples, we adopt as reference the experimental value for the simple aquated ions, $\text{Cu}_{(\text{aq})}^{2+}$ and $\text{Cu}_{(\text{aq})}^{1+}$. In aqueous solution Cu(II), exists as an aquated ion with an average of between five and six water molecules in a distorted octagonal arrangement. Four ligands form a distorted square planar arrangement while the fifth and sixth water molecules are labile with elongated approximately axial bonds. Computationally, the best representation is a distorted square planar coordination pattern of four water molecules. At the B3LYP/6-311+G(2df,2p) level, the fifth and sixth water ligands are released into the bulk aqueous medium. At the same level of theory, the aquated Cu(I) ion has two coordinated water molecules in a linear arrangement. Thus the reduction of Cu(II) to Cu(I) is accompanied by the release of two water molecules into bulk water and a calculated entropic component, $-T\Delta S = -63$ kJ mol⁻¹ at 298 K. The entropic contribution alone raises the reduction potential by $63/F = 0.65$ V, where F is the Faraday constant ($F = 96.485$ kJ mol⁻¹ V⁻¹). The experimental value for aqueous Cu^{2+} is $E^\circ = 0.159$ V vs SHE [2]. Thus, if the entropic component is zero (i.e., the water ligands could not escape), the reduction potential would be negative, -0.48 V. We note also that $\Delta G^{\text{Cu}} = -434$ kJ mol⁻¹, of which $\Delta H^{\text{Cu}} = -371$ kJ mol⁻¹. The enthalpic component arises from a complex interaction of several influences that are listed below.

- 1) **Nature of the ligands:** A good ligand for Cu(II) may not be a good ligand for Cu(I), thereby destabilizing Cu(I) and reducing the reduction potential, and vice versa. An *ab initio* study of small Cu(II) complexes [6] has revealed that of the biologically available ligand types, at pH 7, the best for Cu(II) are the imidazole ring (histidine) and the thiolate ion (cysteine), followed by the amino group (N-terminus, lysine), carboxylate (aspartate, glutamate), phenolate (tyrosine) and dialkyl sulfide (methionine). Neither the phenolate nor the dialkyl sulfide are predicted to be bound to Cu(II) in aqueous solution at physiological pH [6]. In a related study (unpublished) of small Cu(I) complexes, it was found that

the imidazole and amino groups were also good ligands for Cu(I), as is a dialkyl sulfide.

- 2) **Charge of the ligands:** Cu(II) coordinated to one or more negatively charged ligands would be stabilized more than Cu(I), thereby lowering the reduction potential.
- 3) **Geometry of the coordination sphere:** for instance a tetrahedral arrangement of ligands is ideal for Cu(I) but unfavorable for Cu(II), thereby destabilizing Cu(II) and raising the reduction potential.
- 4) **Polarity of the medium:** Although biological processes, including redox processes, take place in water, the local environment of the metal site may be shielded from bulk water by the surrounding ligands and the wider protein environment. A local environment with a lower effective dielectric constant than water would destabilize Cu(II) more than Cu(I), thereby raising the reduction potential.
- 5) **Electrostatic effects:** The presence of metals or charged residues can significantly affect the ability of Cu(II) to be reduced. Thus the proximity of a second metal, such as the zinc ion in CuZnSOD, undoubtedly contributes to raising its reduction potential above that of a typical Type 2 copper ion. Conversely, a nearby carboxylate group (of Asp or Glu) would lower the reduction potential.

A complete description of the procedures used to calculate reduction potentials is given in the Appendix.

Before we address the specific cases of Alzheimer's disease and the prion diseases, namely copper bound to the amyloid beta peptide and to the sequence HG₃GG, respectively, we consider briefly the two extreme cases, namely albumin, which has a very negative reduction potential, and ceruloplasmin, which is at the opposite end of the scale with a very positive reduction potential. These two proteins, both present in plasma, account for the vast majority of the approximately 100 mg of copper found in the human body.

29.2

Copper Binding in Albumin – Type 2

There is a specific Cu(II) binding site at the N-terminus of albumin (DAHK...), involving coordination of the free N-terminus (of Asp1), two deprotonated amide groups (of Ala2 and His3) and N_π of the imidazole ring in a Type 2 binding pattern. The copper binding site can be adequately modeled by the tripeptide, GGH, in which the N-terminus is protonated, and the C-terminus is in the form of the N-methyl amide in order to simulate continuation of the protein. Figure 29.1 shows the structures and thermodynamics of the Cu(II) and Cu(I) complexes. The most stable Cu(II) complex, Cu(II)·Alb, at physiological pH is formed after proton loss from no fewer than three nitrogen atoms. The binding affinity is predicted to be -101 kJ mol^{-1} . With two negatively charged ligands, the Cu(II) complex, Cu(II)·Alb, is electrically neutral. The reduced complex, Cu(I)·Alb, has the N-terminal amino

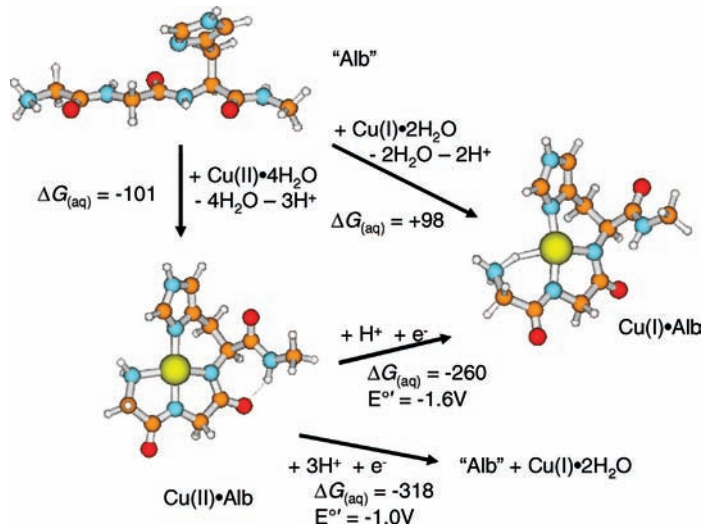


Figure 29.1 Type 2 copper binding to the N-terminus of albumin, modeled by GlyGlyHisNHCH₃ ("Alb"). Large yellow spheres are Cu, medium orange, blue and red spheres are C, N and O, respectively, and small white balls are H. Free energy changes are in kJ mol⁻¹ at pH 7. Standard reduction potentials versus SHE are also at pH 7.

group protonated and hydrogen bonded to the Cu(I) center. The Cu(II) center of Cu(II)·Alb is electron-rich and has little tendency to accept an electron. The predicted reduction potential at pH 7 is $E^\circ[\text{Cu(II)} \cdot \text{Alb}/\text{Cu(I)} \cdot \text{Alb}] = -1.6\text{ V}$, which is substantially lower than the experimental value for Cu(II)/albumin, $E^\circ = -0.8\text{ V}$ [5]. The discrepancy arises because the product of reduction is most likely *not* Cu(I)·Alb, since Cu(I)·Alb is predicted to be very unstable in aqueous solution, by 98 kJ mol⁻¹ relative to "Alb," that is, GGH and Cu(I)·2H₂O. The instability ensues from the fact that, unlike Cu(II), Cu(I) has no tendency to acidify an amide group. Thus, reduction of the Cu(II) complex would be accompanied by dissociation of the copper from the protein. If the redox couple includes the dissociated Cu(I)_(aq) + protein, then the reduction potential is raised to $E^\circ[\text{Cu(II)} \cdot \text{Alb} + 3\text{H}^+/\text{Cu(I)}_{\text{(aq)}} \cdot \text{"Alb"}] = -1.0\text{ V}$. This last value is in satisfactory agreement with the experimental value for albumin, and represents the extreme case in which the reduced cuprous ion is released into solution. In any case, here, as in other peptides in which the Cu(II) ion is bound to one or more deprotonated amide groups, the reduction potential is negative. In summary, of the five factors that influence the enthalpic part of the reduction potential of albumin to the greatest extent, factors 1 and 2, the nature and charge of the ligands, are probably the most important and result in a value that is more negative than observed. Since the binding site in albumin is at the N-terminus, it is likely that the metal is exposed to solvent and that reduction is accompanied by dissociation. The associated entropic contribution serves to raise the reduction potential closer to the experimental value.

29.3

Copper Binding to Ceruloplasmin – Type 1

At the other end of the redox scale is the blue copper protein ceruloplasmin. Ceruloplasmin is a fascinating multifunctional, multicopper oxidase [7]. From the point of view of the present chapter, our interest is centered on the origin of the unusually high reduction potential, about 1 V vs SHE. Human ceruloplasmin (M_r 132 000) consists of a single polypeptide chain (1046 amino acid residues, with about 8% carbohydrate content), divided into three contiguous homology units [8] with three different Type 1 copper sites, and a trinuclear copper cluster, in which one of the three copper atoms is in a typical Type 2 site and the other two are spin-coupled into an EPR-silent electronic configuration (Type 3 site) [9]. All three of the Type 1 sites have the characteristic Cys, His, His coordinating ligands. Two also have a fourth (Met) residue, while the third site is tricoordinate, having a non-coordinating Leu residue in place of the Met. It is this last site to which the high reduction potential is attributed [9]. A high reduction potential for a Type 1 site is problematic by factors 1 and 2, the nature and charge of the ligands. Besides the two His ligands, our theoretical modeling mentioned in connection with factor 1 above shows that a thiolate is also a very good ligand for Cu(II). In fact, the geometry of a thiolate-substituted Cu(II) complex is more typical of a Cu(I) complex, indicating substantial reduction of the copper and oxidation of the thiolate. On this basis, the shortage of ligands notwithstanding, one would expect a *negative* reduction potential rather than a high positive one. We have examined a model of the basic tricoordinated Type 1 site with the $\text{Cu}(\text{Im})_2(\text{SCH}_3)$ complex, labeled Cu(II)·Cp2 and Cu(I)·Cp2 in Figure 29.2a, where the Cp2 part identifies it as the Type 1 site in domain 2 of ceruloplasmin. As expected, in the absence of geometry constraints and in aqueous solution, this complex is predicted to have a negative reduction potential, $E^\circ = -0.3$ V vs SHE, far below the value attributed to this site. We can use the vacuum phase free-energy change to calculate the reduction potential for the extreme case of a low polarity medium. In the absence of solvent, $E^\circ = +0.80$ V. The difference, -1.1 V, or about 110 kJ mol^{-1} , reflects the amount of extra stabilization that the positively charged oxidized form experiences over the neutral reduced form in the presence of water. The calculated value in the gaseous (as a model for hydrophobic) phase is still lower than observed for the Type 1 site in domain 2 of ceruloplasmin. Additional geometry constraints that are unfavorable for Cu(II) may account for the rest. Computationally, if the geometry of the Cu(II)·Cp2 complex is constrained to correspond to that of Cu(I)·Cp2, its “gaseous phase” reduction potential is predicted to be $+1.0$ V.

The reduction potentials of the other Type 1 sites (in domains 4 and 6) of ceruloplasmin, which also have an axial Met residue, have been determined to be 0.45 V. This configuration of copper binding, which is typical of most Type 1 sites, is modeled by $\text{Cu}(\text{Im})_2(\text{SCH}_3)(\text{CH}_3\text{SCH}_3)$ [Cu(II)·Cp4 and Cu(I)·Cp4 in Figure 29.2b]. The Met residue (modeled by dimethyl sulfide) forms a fourth ligand for the coordinatively unsaturated Cu(II) site, but in a distorted tetrahedral geometry more indicative of Cu(I). As in Figure 29.2a, extensive spin delocalization to the thiolate sulfur atom is predicted. One-electron reduction yields Cu(I)·Cp4 in which

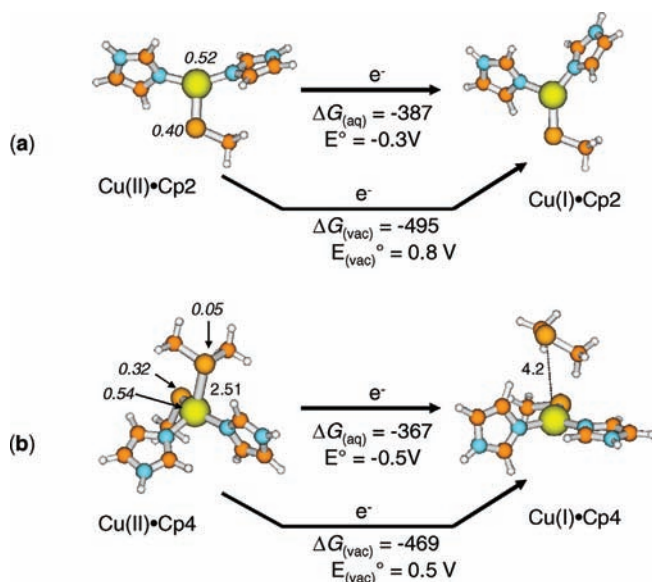


Figure 29.2 Models of Type 1 copper binding sites in ceruloplasmin: (a) the high reduction potential site in domain 2; (b) a redox active site in domain 4. Large yellow spheres are Cu, larger orange spheres are S, the smaller orange and blue spheres are C and N, respectively, and the white spheres are H. Free energy changes are in kJ mol⁻¹ at pH 7. Standard reduction potentials versus SHE are also at pH 7.

the copper is coordinatively saturated with three ligands and does not add the Met residue as a fourth ligand. In the gaseous phase, the Cu–S separation is 4.2 Å. In enzymes with this Type 1 copper site, the Met side chain is not free to move so far away. In ceruloplasmin, the domain 4 and 6 Type 1 sites have Cu–S distances of 2.9 Å and 3.3 Å, respectively [10]. Thus our model systems indicate that the Cu(II) is stabilized by the presence of Met but the Cu(I) is not, suggesting that the reduction potential should be lower than that of the domain 2 site where the Met is absent. Earlier computational modeling has already shown that the nature and position of axial ligands has little influence on the reduction potential of the blue copper proteins [4]. Experimentally, the presence of the Met residue has been shown to lower the reduction potential by 0.10 V [11]. We find that if the site is fully solvated, a negative reduction potential, $E^\circ = -0.5$ V vs SHE, would result. As for the domain 2 site, a positive value can only be achieved in a hydrophobic environment, the extreme case being $E_{(vac)}^\circ = 0.5$ V. We conclude that the observed positive reduction potentials for the blue copper proteins (all have Type 1 sites) must be due to the fact that these sites exist in a hydrophobic, water-free, low dielectric environment. This conclusion directly contradicts an early study of fungal laccase in which the high reduction potential, 0.78 V, was attributed to stabilization of Cu(I), with solvent accessibility playing a minor role [12]. The factors that determine the *relative* reduction potentials

of various blue copper proteins have been shown computationally to include axial ligand interactions, hydrogen bonding to the S_{Cys} and protein constraint on the inner sphere ligand orientations [13].

As well as being at the opposite ends of the reduction scale spectrum, we have seen that albumin and ceruloplasmin are also at two extremes with respect to the environment in which the copper finds itself. In fact, were it not for the extreme shielding of the latter Type 1 site from the aqueous environment, the two systems would have had similar, negative, reduction potentials. We ask now what factors might determine the reduction potential in the case of a small peptide, or a structureless segment of protein, that is exposed to the aqueous environment. The descriptors “small” and “structureless” are meant to imply that the peptide/protein is unable to impose severe geometrical constraints and that both the oxidized and reduced forms of the metal/peptide complex are as stable as possible. Such a situation exists in albumin since the N-terminus is exposed to the solvent, and we have seen that it is responsible for the negative reduction potential. We examine whether additional factors come into play in two copper-binding peptides/proteins of interest in connection with neurological diseases, the prion protein in the case of transmissible spongiform encephalopathies (TSEs) and the amyloid beta peptide of Alzheimer’s disease.

29.4

The Prion Protein Octarepeat Region

The prion protein, PrP^C, is present in all mammalian and avian tissues but its precise function is not known. A refolded form, PrP^{Sc}, is the infectious agent in rapidly degenerating and incurable neurological diseases collectively known as transmissible spongiform encephalopathies (TSEs), including scrapie in sheep, “mad cow disease” in cattle, chronic wasting disease in elk and deer, and Creutzfeldt–Jakob disease (CJD) and kuru in humans. At the N-terminal region are normally four octarepeats, PHGGGWGQ, spanning PrP^C(60–91), that can bind a single cupric ion each in a Type 2 binding environment that, at pH 7, has an N3O1 coordination pattern. While other regions of PrP^C can also bind copper, expansion of the octarepeat segment has been directly linked to conversion of PrP^C into PrP^{Sc} and development of CJD [14]. Sporadic CJD is a very rare condition. However, there has been much research directed to TSEs because the infection can jump species barriers in some cases, for example sheep to cattle to humans through feed, and the 10–15 year induction period for the development, and the subsequent fear of an epidemic of the variant form of CJD. Our interest here is directed only toward the redox chemistry of copper/prion complexes in the octarepeat region.

In vitro and murine experiments indicate that copper-loaded prion protein undergoes catalytic redox cycling in the presence of reducing agents such as superoxide, ascorbate and catecholamines (including the neurotransmitter dopamine) and is itself damaged under these conditions [15]. The reduction potential of the prion protein is poorly defined experimentally, $-0.16\text{ V} < E(\text{PrP}^{\text{C}}) < +0.53\text{ V}$ [15].

The reduction potentials of the octarepeat segment, Cu(II)/PHGGGWGQ, and the shorter unit, Cu(II)/HGGG, which have the same copper binding pattern, fall outside of this range, $E^\circ[\text{Cu(II)/PHGGGWGQ}] = -0.31 \text{ V}$ and $E^\circ[\text{Cu(II)/HGGG}] = -0.289 \text{ V}$ [16]. This is an indication that the redox activity may reside in another binding site, for example, at H96 and H111.

Detailed information about the structures of some Cu(II)/peptide complex models of the octarepeat region is available from crystallography, but information about structures in solution comes from electron paramagnetic resonance (EPR), circular dichroism (CD), infrared (IR) and UV/Vis spectroscopies, and is much less detailed. Millhauser and coworkers were able to partially clarify the binding of Cu(II) to full PrP^C from their use of X-ray crystallography, CD and EPR spectroscopy studies [17–19]. They determined the prominent binding modes in both solid and solvated forms, at various pH. At pH 7.4, two binding modes existed, called component 1 and component 2. Component 1, the major component, has a square planar binding environment about the Cu center consisting of three N ligands and an oxygen atom (i.e., N3O1). It was established that one of the N atoms is N_π of His and the other two are deprotonated amides of the adjacent two Gly residues. The oxygen is the carbonyl oxygen of the second Gly. In the absence of the Trp residue, that is, HGGG or HGG, component 2 is dominant. Its binding stoichiometry was tentatively identified as N2O2. These results supported previous experimentally determined binding environments at varied pH values [20, 21]. It was found that the minimum peptide sequence required to model the observed N3O1 and N2O2 environments of the full length Cu(II)-bound octarepeat region is the fragment “HGGG” [17, 19].

We have examined computationally the binding of the Cu(II) to the shorter segment, N-AcHGGG(NH₂), which we will refer to simply as HGGG [22]. At physiological pH, two forms of Cu(II)/HGGG, labeled as Cu(II)·PrA and Cu(II)·PrB in Figure 29.3, are predicted to coexist. Cu(II)·PrA has N2O2 coordination and corresponds to Millhauser’s component 2. One of the oxygen atoms of Cu(II)·PrA is part of the π-type coordination of the His carbonyl group to the copper. Cu(II)·PrB has N3O1 coordination and corresponds to Millhauser’s component 1 [17–19]. The calculated effective pK_a for Cu(II)·PrA, pK_a = 8.6, confirms that Cu(II) can acidify an amide group by 6 or 7 orders of magnitude.

The corresponding reduced Cu(I) structures are labeled Cu(I)·PrA and Cu(I)·PrB in Figure 29.3, respectively. The gaseous- or solution-phase-optimized geometry of Cu(I)·PrA is very similar to Cu(II)·PrA except that the Cu(I) center has moved away from the carbonyl oxygen of the His and appears to be undertaking a nucleophilic addition to the carbon end of the carbonyl group. This is a consequence of the nucleophilic character of Cu(I) bound to electron-rich ligands that we have already seen in Cu(I)·Alb (Figure 29.1) in which the protonated N-terminus was H-bonded to the Cu(I). At pH 7, structure Cu(I)·PrA is unstable by 58 kJ mol^{−1} relative to Cu(I)·PrC in which the amide group has been reprotonated and the Cu(I) is only attached to the peptide by N_τ of the His residue.

The structure of Cu(I)·PrB, the direct product of reduction of Cu(II)·PrB, has N3 coordination. The carbonyl oxygen of the second Gly residue has moved away from the metal. Cu(I)·PrB, with two deprotonated amide groups, is less stable than

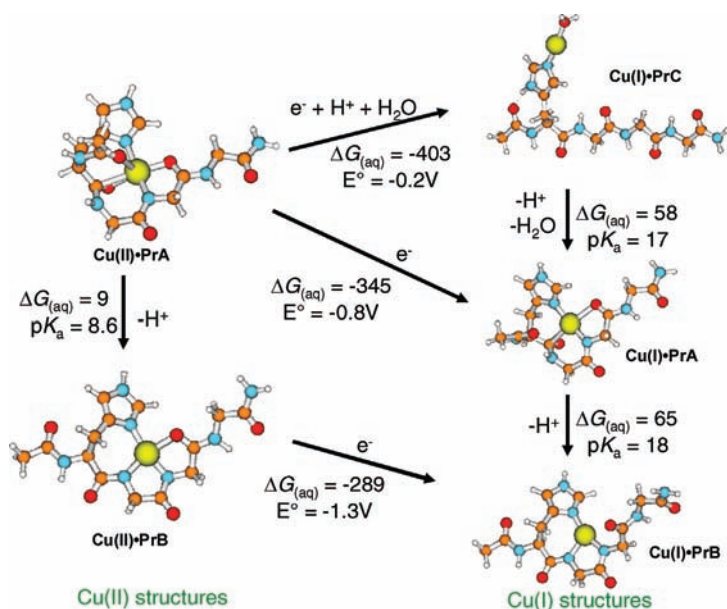


Figure 29.3 Type 2 copper binding to the octarepeat region of the prion protein, modeled by N-AcHisGlyGlyGlyNH₂. Large yellow spheres are Cu, medium orange, blue, and red spheres are C, N, and O, respectively, and small white balls are H. Free energy changes are in kJ mol⁻¹ at pH 7. Standard reduction potentials versus SHE are also at pH 7.

Cu(I)·PrA, which has one, by 65 kJ mol⁻¹ at pH 7. The calculated effective pK_as for Cu(I)·PrC and Cu(I)·PrA, 17 and 18, respectively, indicate that Cu(I) has no power to acidify an amide group.

The calculated reduction potentials corresponding to no ligand loss, $E^\circ[\text{Cu(II)·PrA}/\text{Cu(I)·PrA}] = -0.8\text{ V}$ and $E^\circ[\text{Cu(II)·PrB}/\text{Cu(I)·PrB}] = -1.3\text{ V}$ (Figure 29.3), are clearly too low compared to the experimental value, -0.3 V [16]. However, if reduction is accompanied by release of the amide nitrogen ligands and reprotonation of the amide groups, then satisfactory agreement with experiment is obtained: $E^\circ[\text{Cu(II)·PrA} + \text{H}^+/\text{Cu(I)·PrC}] = -0.2\text{ V}$ and $E^\circ[\text{Cu(II)·PrB} + 2\text{H}^+/\text{Cu(I)·PrC}] = -0.3\text{ V}$.

29.5

Copper and the Amyloid Beta Peptide ($A\beta$) of Alzheimer's Disease

Alzheimer's disease (AD) is of particular interest because it is primarily a disease of old age with only a minor genetic component. With the general life expectancy steadily increasing, AD has become a problem of epidemic proportions that will continue to escalate, afflicting about 5 million Americans in 2008. The progression of the disease is slow. The average period of survival is eight years – some can survive as

long as 20 years. There is no known cause and no cure. Physically, AD is characterized by massive loss of neurons and disruption of synaptic function throughout the brain, beginning in the hippocampus, an area of the cortex that plays a key role in formation of new memories. Currently approved drugs ameliorate symptoms for a short time by boosting levels of neurotransmitters, but do not alter the general progression or outcome of the disease. Genetics plays a small role, about 5% of the total [23]. As for the rest, only one risk gene, apolipoprotein E- ϵ 4 (ApoE- ϵ 4) has been identified with certainty. All of the genetic mutations and risk factors are associated with abnormal production or clearance of a small peptide, the amyloid β -peptide ($A\beta$), which is the major constituent of the senile plaques that are diagnostic of AD. The case for the amyloid beta peptide ($A\beta$) as a causative agent in AD, first enunciated as the Amyloid Hypothesis [24, 25], is now widely accepted [26–33]. The last decade has seen significant advances in understanding the mechanisms of $A\beta$ neurotoxicity and this understanding has spawned a new generation of drug candidates that should lead to prevention of the disease.

The role of copper in AD is circumstantial but compelling [34]. The AD brain is characterized by extensive oxidative stress [35–37]. This manifests itself as significantly reduced levels of antioxidants (e.g., vitamins E and C, and glutathione) and elevated levels of products of oxidative damage to proteins, to DNA (e.g., 8-hydroxy-2'-deoxyguanosine) [38] and to lipids (e.g., 4-hydroxynonenal, HNE). The last, lipid peroxidation, is the beginning of a complex cascade of events that results in cell death, the result of the accumulated damage, and/or by apoptosis [28]. Lipid peroxidation is correlated with brain degeneration [39]. $A\beta$ has been reported to have a high affinity for copper [40, 41]. $A\beta$, in combination with copper ions and oxygen, generates reactive oxygen species, especially hydrogen peroxide [42, 43], and causes lipid peroxidation and protein oxidation [44]. The chemical toxicity due to radical formation in the brain is aggravated by the fact that neuronal membranes are enriched in polyunsaturated fatty acids (PUFAs) that are particularly susceptible to lipid peroxidation [45]. $A\beta$ toxicity is ameliorated by antioxidants [46–50], including vitamin E [51] and glutathione [52]. $A\beta$ is itself damaged by Cu(II)- and Fe(III)-catalyzed oxidation [53]. This chapter is concerned with the primary mechanism of radical production in AD, namely initiation by redox-active peptide-bound copper(II).

The $A\beta$ peptide is a normally soluble 4.3-kDa peptide found in all biological fluids, but it accumulates as the major constituent of the extracellular deposits in the brain that are the pathological hallmarks of Alzheimer's disease (AD) [54, 55]. $A\beta$ is generated as a mixture of polypeptides manifesting carboxyl terminal heterogeneity. The two main isoforms are $A\beta$ 1–40 and $A\beta$ 1–42. The $A\beta$ 1–40 isoform is the predominant soluble species in biological fluids [56, 57], while $A\beta$ 1–42 is the predominant species found in senile plaques (SP) deposits [58]:



Metabolic signs of oxidative stress in the AD-affected neocortex include increased glucose-6-phosphate dehydrogenase activity [59] and increased heme oxygenase-1 levels [60]. Signs of oxygen radical-mediated chemical attack include increased free protein carbonyls [61–63], lipid peroxidation adducts [64, 65] protein nitration [66] and mitochondrial and nuclear DNA oxidation adducts [67]. The generation of reactive oxygen species by A β coordinating Cu(II) is well documented [41]. A β 1–42, as well as A β 1–40, ascorbate and other peptides, is known to reduce the A β -bound Cu(II) to Cu(I) [68–72]. The reduced Cu-A β 1–42 complex was shown to catalytically reduce O₂ to neurotoxic H₂O₂ [68, 69, 71]. Copper chelators abolish the H₂O₂ production, indicating that it is a metal dependent reaction. A modified TBARS (thiobarbituric acid-reactive substance) assay that detected the presence of hydroxyl radicals (OH[•]) suggested Fenton-like chemistry [69].

The electrochemical behavior of Cu(II) was assessed in the presence and absence of A β by cyclic voltammetry. The voltammetric response has been reported to have a formal reduction potential of +0.72–0.77 V versus SHE in phosphate-buffered solution. More recently, this value has been questioned on the basis that reduction potentials of Cu(II)/A β (1–16) and Cu(II)/A β (1–28) were measured in the range $E^{\text{red}} = +0.33\text{--}0.34$ V versus SHE [73]. A slightly lower value was found for Cu(II)/A β (1–42), $E^{\text{red}} = 0.28$ V, when care was taken to ensure measurements on the monomeric species [74]. These properties directly correlate with the copper-mediated potentiation of A β neurotoxicity in cell culture [69].

29.6

Cu(II)/Cu(I) Reduction Potentials in Cu/A β

The reduction of Cu(II) to Cu(I) in the Cu(II)/A β complex is an important step for neurotoxicity of A β . The one-electron transfer generates radicals that can lead to lipid peroxidation [75]. Although the identity of these radicals is still in doubt, we have shown computationally that “C-centered backbone radicals are capable of initiating lipid peroxidation [76, 77], and that these could be generated by methionyl sulfur radical cations that were hypothesized to be generated from oxidation of the Met35 residue by Cu(II) bound to A β [78]. This hypothetical step has been problematic because of the wide disparity of reduction potentials of Cu(II) bound to proteins in a Type 2 pattern (normally <0.0 V) and Met sulfur radical cations (≈ 1.4 V) [79, 80]. As previously stated, there is a report of an experimental measurement of the reduction potential of Cu(II)/A β that yielded a value ~ 0.7 V (versus SHE) [69]. The reason for the large value has yet to be determined, especially in light of more recent measurements on the monomeric species that yielded a lower value, ~ 0.3 V (versus SHE) [74]. The nature of the copper binding environment in both the oxidized and reduced form of the Cu/A β complex still requires clarification. To gain some insight into how the copper coordination sphere in A β may affect the reduction potential, model Cu(II) [81] and Cu(I) species have been used to calculate Cu(II)/Cu(I) reduction potentials [82]. These are described below.

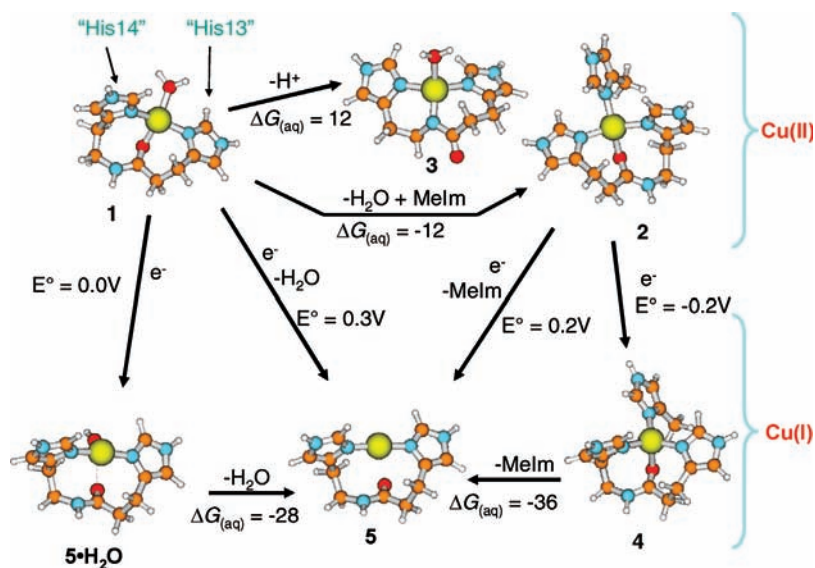


Figure 29.4 Model of Type 2 copper binding to His13His14 of the amyloid beta peptide. 1, 2, and 3 are Cu(II) species; 5•H₂O, 5, and 4 are Cu(I) species. Large yellow spheres are Cu, medium orange, blue and red spheres are C, N, and O, respectively, and small white balls are H. Free energy changes are in kJ mol⁻¹; standard reduction potentials versus SHE.

In human blood plasma, more than 98% of the amino acid bound Cu(II) occurs in histidine (His) complexes [83, 84]. A β has three His residues and binds Cu(II) and other metals with high affinity [40]. ¹H NMR spectra [85] Raman [86] and selective oxidation [87] experiments indicate that His13 and His14 are the two most firmly established ligands in a Type 2 square-planar coordination sphere of Cu(II) [85], bound to A β . As a model of the His-His sequence, we adopted 3-(5-imidazolyl)-propionylhistamine, and found computationally that the three most stable Cu(II) complexes at pH 7 are 1–3 (Figure 29.4). The computed values are slightly different because of a different treatment of the free energy of solvation and the use of a different basis set. In 1 the oxygen of the backbone carbonyl (Oc) is involved in the Cu(II) coordination sphere along with both imidazoles, plus a water molecule used to fill the fourth coordination site of the Cu(II). His6 has also been shown as a ligand for Cu(II) in monomeric A β [85], but there is no evidence for the coordination of Met35, and Tyr10 may [70] or may not [85] be a ligand.

Copper-induced aggregation of A β (1–40) or A β (1–42) is associated with enhanced neurotoxicity, at least *in vitro* [88–90]. Aggregation may take place by either or both of two mechanisms. By analogy with the structure of the CuZn superoxide dismutase, one of the Cu-bound His residues may act as a bridge between two Cu ions bound to separate A β . Evidence suggests that such a bridge occurs at Cu(II) : A β (1–42) ratios in

excess of 0.5 : 1 [88, 89]. The study of binuclear complexes is beyond the scope of the present work, and we do not discuss this binding mode further.

Copper(II) has also been proposed to be the link atom bridging A β molecules through His residues, thus promoting aggregation [70, 86]. Both possibilities are considered by replacing the water molecule of **1** by 4-methylimidazole. Thus species **2** (Figure 29.4) models the involvement of His6 or a His from a second A β , in addition to His13 and His14 in the Cu(II) coordination sphere. In **3**, the Cu(II) is coordinated to the deprotonated nitrogen of the backbone (Nb). Previous theoretical work found that species **1** and **3**¹⁾ (Figure 29.4) are approximately equally populated at pH 7. In the present treatment of solvation (see the Appendix), **3** is predicted to be significantly populated only at alkaline pH 9. Displacement of the water ligand of **1** by the methylimidazole group, MeIm, is found to be exergonic by 12 kJ mol⁻¹.

To the best of our knowledge there are no experimental data in the literature about the binding of Cu(I) to A β . In our group, we have modeled computationally the binding of Cu(I) to the Met35 region [91] and to the His-His regions[82]. The former was found to be of low affinity, while the latter complex, **5** in Figure 29.4, was strongly bound, by 93 kJ mol⁻¹, relative to the copper-free peptide and Cu(I)(H₂O)₂. In the present work, we have reexamined, by means of high level quantum chemical calculations and a better treatment of solvation, a number of the Cu(I)-bound species from the earlier study [82]. The dicoordinated species, **5**, is still the most stable, but by a lesser amount, 64 kJ mol⁻¹, relative to the separated species. As shown in Figure 29.4, this species may be produced by reduction of either **1** (with the loss of water) or **2** (with loss of MeIm). The reduction potentials of both paths are positive, 0.3 and 0.2 V, respectively, and are in very good agreement with the most recent experimental value [74]. Loss of a ligand is required to achieve a positive reduction potential. Reduction of **1** or **2** without loss of a ligand to yield the corresponding tetracoordinated Cu(I) species, **5-H₂O** and **4**, respectively, entails potentials of 0.0 V and -0.2 V, respectively. The calculated values are short of the high value, 0.7 V, measured for Cu/A β (1-42). Thus the present results reinforce the earlier conclusion that, at least within the limitations of the peptide model for a monomeric Cu/A β complex, it is not possible to achieve a reduction potential as high as observed. It is reasonable that this conclusion applies to the full Cu/A β complex as well. Within the set of ligands available for coordination of Cu(I) to A β , the Cu(I) complex will be as stable as possible. A thiol group would be a better ligand than an imidazole group, but there are no thiol groups available within A β , unless demethylation of an oxidized Met35 could take place (yielding homocysteine). Such an eventuality was shown to be feasible computationally by nucleophilic transfer to a second Met residue [92], but there is only one Met in monomeric A β and such a transfer could only take place in an oligomeric form of A β . The same considerations apply to Cu(II)/A β complexes. The best ligand for Cu(II) is an imidazole group. Chelation by two such groups, His13 and His14, is well modeled by **1**, **2** or **3**. The adjacency of the two His residues ensures

1) Species **1** is species **A** in Reference [82] and **20_x** in Reference [81]; species **3** is species **C** in Reference [82] and species **35_x** in Reference [81]. Species **2** and **4** correspond to species **B** and **12**, respectively in Reference [82].

that there is only the possibility of deprotonating one adjacent amide group, namely the amide bridge between the two, yielding structure **3**, which is marginally less stable than **1** at pH 7. Which Cu(II)/peptide complex is most representative of monomeric Cu(II)/A β (1–42)? In agreement with the imidazole group being a better ligand for Cu(II) than water, the computations indicate that **2** is slightly more stable than **1**. However, molecular dynamics simulations on Cu(II)/A β (1–42) suggest that this may not be the case if the third imidazole is that of His6 because of internal strain in the macrocyclic structure [93]. NMR experiments on Cu(II)/A β (1–28) indicate that His6 is only transiently coordinated to the Cu(II), as is the N-terminus. Thus the Cu(II)/A β (1–42) complex, modeled by structures **1** or **2**, is as stable as possible, and the calculated reduction potentials, $E(1/5) = 0.3$ V and $E(2/5) = 0.2$ V are probably an accurate indication of the magnitude of the values that may be expected in a monomeric Cu/A β complex.

This work has indicated that the most likely description of a Cu(I) binding site for a His-His peptide in aqueous solution involves a dicoordinated Cu(I) ligated to the N $_{\pi}$ of both histidines. The Oc of the backbone is farther away than in the Cu(II) complex. The relatively high affinity of this site for Cu(I) makes it fairly stable toward ligand replacement (understood as substitution of other amino acid side chain groups). The non-reversible redox behavior of Cu(II)/Cu(I) is a well-known phenomenon, sometimes described as gated electron transfer [94]. Therefore, the reoxidation of the lower coordinated Cu(I) species will depend on the nature of the ligands available to satisfy the higher coordination of the Cu(II) species.

29.7

Concluding Remarks

The computational studies described herein provide valuable insights into the factors that determine the reduction potentials of copper in different peptide/protein environments. Redox activity requires that the reduction potential be positive versus SHE. In copper–peptide complexes, both the oxidized and reduced copper will be as stable as the available ligand set permits since the small peptide units lack rigid tertiary structure that can impose geometric constraints to destabilize either or both forms. In such a case, the reduction potential will tend to be lower. However, since Cu(I) prefers fewer coordinating ligands than does Cu(II), the entropic contribution to the free energy change that ensues from the shedding of ligands upon reduction from Cu(II) to Cu(I) serves to drive the reduction potential to more positive values. In copper–protein complexes, this contribution is usually missing since the protein superstructure prevents ligands from moving away.

The typical Type 1 environment of the blue copper proteins, in which the copper ion is ligated to a cysteine residue, two histidine residues and a methionine residue, would have a negative reduction potential (relative to SHE) if the coordinated metal is exposed to the solvent. Positive values for the redox-active blue copper proteins require that the complexes reside in regions from which water is excluded.

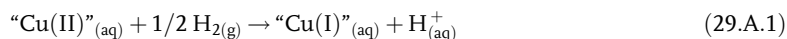
In a Type 2 environment in which the copper is bonded to a single histidine residue, it will complete its coordination sphere by attaching one or two deprotonated amide residues if it is exposed to the aqueous environment at physiological pH. This is the case for albumin and the octarepeat region of the prion protein. Under such conditions, the cupric ion will also have a strongly negative reduction potential versus SHE.

Reduction potentials calculated for model copper/A β complexes tend to have low positive values, largely due to the entropic effect of ligand shedding. The high positive value observed in one experiment is most likely evidence that the measurement was carried out on an aggregated form of A β . It is significant that the neurotoxic form of A β is an oligomeric form, and part of the neurotoxicity is due to its ability to induce oxidative stress [95].

29.A Appendix

29.A.1 Calculation of Reduction Potentials, E° , of Copper/Peptide Complexes

The standard reduction potentials (E°), relative to the standard hydrogen electrode (SHE), can be calculated from the free-energy change of the one-electron transfer reaction:



by the relation:

$$E^\circ(\text{“Cu(II)”}/\text{“Cu(I)”}) = -\Delta G_{(\text{aq})}/F \quad (29.A.2)$$

Here it is understood that “Cu(II)” and “Cu(I)” represent all species that take part in the half-cell reaction involving the oxidized and reduced forms of the copper complexes:



for which the free energy change is denoted ΔG^{Cu} , ignoring the electron. In the case of the Cu(II)/Cu(I) couple especially, it is important to consider all species because Cu(II) tends to have a higher coordination number, typically 4–6, than Cu(I), typically 2–4, and the shedding of ligands during the reduction process results in an important contribution to ΔG^{Cu} . For the same reason, reduction and reoxidation are not necessarily the microscopic reverses of each other since ligands shed during the reduction may be replaced by others during the reoxidation step. Cyclic voltammetry of copper enzymes usually is reversible because the enzyme superstructure prevents the ligands from drifting away.

The standard reduction potential, E° , has all species at 1-M concentration. If this is not the case, the actual electrochemical potential may be calculated from the Nernst equation:

$$E = E^\circ - (RT/F)\ln Q \quad (29.A.4)$$

where Q is the reaction coefficient, $Q = [\text{“Cu(II)”}]/[\text{“Cu(I)”}]$, which incorporates the actual concentrations. A special consideration arises if protons are liberated or consumed during the reduction, for example, $\text{“Cu(II)”} + e^- \rightarrow \text{“Cu(I)”} + n\text{H}^+$. Since the reactions we are primarily interested in occur under physiological conditions in which pH is buffered close to 7, it is convenient to define, again taking all concentrations as 1 M, except $[\text{H}^+]$:

$$E' = E^\circ - (RT/F)\ln[\text{H}^+]^n \quad (29.A.5.1)$$

$$\begin{aligned} &= E^\circ - (RT/F)\ln(10^{-7n}) \\ &= E^\circ + 0.414n \end{aligned} \quad (29.A.5.2)$$

The last equality applies to $T = 298$ K. Thus, the reduction potential is increased by about 0.4 V for each proton that is liberated, and reduced by the same amount for each proton consumed (i.e., n is negative).

The free energy change, $\Delta G_{\text{(aq)}}$, of Reaction A.1 is the sum of $\Delta G^{\text{Cu}} + \Delta G^{\text{SHE}}$, where ΔG^{SHE} is the free energy change of the hydrogen half-cell reaction, $1/2\text{H}_{2(\text{g})} \rightarrow \text{H}^+_{\text{(aq)}} + e^-$, again ignoring the electron. The experimental free energy of the half-cell reaction for the standard hydrogen electrode, using the electron convention, is $\Delta G^{\text{SHE}} = +418 \text{ kJ mol}^{-1}$, and has been adopted for calculating the Cu(II)/Cu(I) reduction potentials. The 418 kJ mol^{-1} is the Gibbs free energy for the half-reaction [96]: $1/2\text{H}_{2(\text{g})} \rightarrow \text{H}^+_{\text{(aq)}} + e^-$. It can be obtained by adding $\Delta_f G^\circ_{\text{(g)}}(\text{H}^+) = 1517 \text{ kJ mol}^{-1}$, $\Delta G_{\text{solv}}(\text{H}^+) = -1107 \text{ kJ mol}^{-1}$ [39] and, to change the $\text{H}^+_{\text{(aq)}}$ reference state to 1 M, $-RT\ln(1/24.46) = 8 \text{ kJ mol}^{-1}$. $\Delta_f G^\circ_{\text{(g)}}(\text{H}^+) = 1517 \text{ kJ mol}^{-1}$ is computed from $\Delta_f G^\circ_{\text{(g)}}(\text{H}) = 203 \text{ kJ mol}^{-1}$ [53] plus $\Delta_r G^\circ_{\text{(g)}}(\text{H}^+ + e^- \rightarrow \text{H}) = 1314 \text{ kJ mol}^{-1}$.

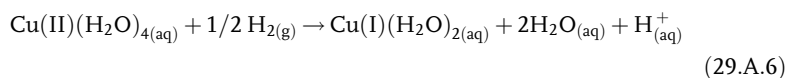
29.A.2 Computational Methodology

All-electron calculations have been carried out using the Gaussian 98 [97] and Gaussian 03 [98] suites of electronic structure codes. The Molden 4.3 [99] visualization program was employed to generate the figures. All geometry optimizations were performed without constraints, using the B3LYP hybrid density functional method [100] as implemented by Gaussian, with the 6-31G(d) basis set (denoted SB for small basis set). Harmonic frequency analysis of each structure, calculated at the B3LYP/SB level, was used to verify that it is at a stationary point on the potential energy surface. The harmonic frequency data were scaled by 0.9806 [101] to derive zero-point energies, but were not scaled for entropies and thermal corrections to the enthalpy. To provide more accurate relative enthalpies, single-point energies have been calculated at the B3LYP/LB//B3LYP/SB level of theory, where LB denotes

the large basis set, 6-311+G(2df,2p), and // denotes “at the geometry of.” In previous work, a very similar level of theory was shown to provide reasonably accurate structures and energetics for complexes of water, ammonia [102], imidazole [103], acetonitrile [104] and dimethyl ether [105] with Cu(I). The difference between experiment and theory ranged from 8 to 12 kJ mol⁻¹, and we assume this range as the accuracy of the present results. For reference, an error of 12 kJ mol⁻¹ in the free energies corresponds to ± 2 units in pK_a values and ± 0.1 V in reduction potentials (E°).

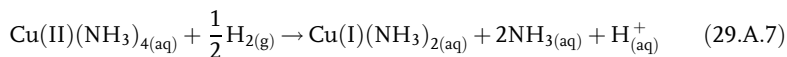
As indicated above at the B3LYP/LB//B3LYP/SB level, residual errors in the calculation of *absolute* values will cancel to a good approximation, yielding reliable *relative* energies. This is not the case for the calculation of free energy changes for reactions like Equation 29.A.1. Since a transition element is involved and the number of electrons changes, the enthalpy change will be less accurately described at this theoretical level than expected for lighter elements. For instance, the ionization potential of Cu⁺ (i.e., the second ionization potential of atomic copper) is calculated to be $IE_2^{\text{calc}} = 2008 \text{ kJ mol}^{-1}$, whereas experiment gives $IE_2^{\text{exp}} = 1958 \text{ kJ mol}^{-1}$ [106]. The discrepancy, 50 kJ mol⁻¹, is probably due to the unequal treatment of electron correlation (an enthalpic term). The error in the ionization potential of Cu⁺ will be present in the reduction potentials, $E^\circ[\text{“Cu(II)”/“Cu(I)”}]$, irrespective of the metal environment, since they will all involve the change in copper oxidation state from +2 to +1. Consequently, it is appropriate to add 50 kJ mol⁻¹ to the gas-phase $\Delta H_{(\text{g})}$ of reaction 29.A.3. Therefore, it is part of the gaseous free energy change, and after addition of the change in free energies of solvation, $\Delta\Delta G_{\text{solv}}$ in ΔG^{Cu} for computing the reduction potentials.

The problem is compounded by the change in charge in Equation 29.A.3, which aggravates errors inherent in the calculation of $\Delta\Delta G_{\text{solv}}$ values due to deficiencies in the implicit solvation model. In partial compensation, we use the experimental value for the free energy of solvation of the water, $\Delta G_{\text{solv}}(\text{H}_2\text{O}) = -26.4 \text{ kJ mol}^{-1}$, and proton, $\Delta G_{\text{solv}}(\text{H}^+) = -1107 \text{ kJ mol}^{-1}$ [107]. [$\Delta G_{\text{solv}}(\text{H}_2\text{O})$ was calculated as the difference between $\Delta_f G_{(\text{g})}(\text{H}_2\text{O})$ and $\Delta_f G_{(\text{l})}(\text{H}_2\text{O})$, corrected to the standard state of 1 M. The CPCM calculated value is $\Delta_f G_{\text{solv}}(\text{H}_2\text{O}) = -28.7 \text{ kJ mol}^{-1}$.] For the calculation of ΔG_{solv} of all other species mentioned in this chapter, we have adopted the integral equation following polarizable continuum (IEFPCM) model in which the cavity is defined in terms of united atom Hartree-Fock (UAHF) radii, and the wavefunction is defined at the level of B3LYP/LB. The default scale factor used in Gaussian03 is 1.20 and this was used throughout:



Using the procedures described above, the calculation of the reduction potential of aqueous Cu²⁺ to Cu⁺ 29.A.6 yields $E^\circ = 0.38 \text{ V}$. This case corresponds to one extreme of the copper environments studied. The 22 kJ mol⁻¹ discrepancy arises from errors in the calculation of the free energies of solvation of the copper ions and may be specific to the particular reaction 29.A.6, as these errors will be dependent on

the nature of the copper environment. To determine if this is the case the reduction potential has been calculated for the other extreme of the copper environments studied, where the copper coordination sphere is filled by ammonias:



The experimental $E^\circ = 0.01 \text{ V}$ [108] was measured at pH 10; consequently, the ammonias released during the reactions would not be protonated (NH_4^+ $\text{p}K_a = 9.24$). The calculated $E^\circ[\text{Cu(II)(NH}_3)_4/\text{Cu(I)(NH}_3)_2 + 2\text{NH}_3]$ is 0.05 V. This includes the correction for the ionization of copper, and is very close to the experimental value. Therefore, in this case the application of a solvation correction is not appropriate. In summary, the $E^\circ[\text{Cu(II)}_{(\text{aq})}/\text{Cu(I)}_{(\text{aq})}] = 0.38 \text{ V}$ should represent the situation where the relative errors in solvation are greatest. The calculated $E^\circ[\text{Cu(II)(NH}_3)_4/\text{Cu(I)(NH}_3)_2 + 2\text{NH}_3]$ close to that of experiment suggests that the magnitude of the error in E° will decrease as the copper environment changes with the replacement of water by nitrogen ligands X:. Therefore, the correction arising from the ionization energy of the copper has been applied, but a correction due to the solvation of the aqueous copper complexes has not been included. We expect that the calculated reduction potentials reported in the main body of the chapter will be overestimated by ~ 0 to 0.2 V.

Acknowledgments

The work described in this chapter would not have happened without the collaboration of many excellent students and coworkers: M. Jake Pushie, Duilio F. Raffa, Gail A. Rickard, Belquis Mothana, Nadine Hewitt, Patrick Brunelle, Rodolfo Gomez-Balderas, Meilan Huang and the late David A. Armstrong. The financial support of the Natural Sciences and Engineering Council (NSERC) of Canada is gratefully appreciated.

References

- 1 Cotton, F.A. and Wilkinson, G. (1988) *Advanced Inorganic Chemistry*, 5th edn, Wiley-Interscience, New York, pp. 752–757.
- 2 Bard, A.J., Parsons, R., and Jordan, J. (1985) *Standard Potentials in Aqueous Solutions*, IUPAC (Marcel Dekker), New York, USA.
- 3 For a concise review of copper coordination environments in biological systems, see the PROMISE database hosted by the Scripps Institute: <http://metallo.scripps.edu/PROMISE/CUMAIN.html>.
- 4 Olsson, M.H.M. and Ryde, U. (1999) The influence of axial ligands on the reduction potential of blue copper proteins. *J. Biol. Inorg. Chem.*, **4**, 654–663.
- 5 Greenaway, F.T., Hahn, J.J., Xi, N., and Sorenson, J.R.J. (1998) Interaction of Cu(II) 3,5-diisopropylsalicylate with human serum albumin – an evaluation of spectroscopic data. *BioMetals*, **11**, 21–26.
- 6 Rickard, G.A., Gómez-Balderas, R., Brunelle, P., Raffa, D.F., and Rauk, A. (2005) Binding affinities for models of

- biologically available potential Cu(II) ligands relevant to Alzheimer's disease: an ab initio study. *J. Phys. Chem. A*, **109**, 8361–8370.
- 7 Bielli, P. and Calabrese, L. (2002) Structure to function relationships in ceruloplasmin: a “moonlighting” protein. *Cell. Mol. Life Sci.*, **59**, 1413–1427.
 - 8 Orte, T.L., Takahashi, N., and Putnam, F.W. (1984) Structural model of human ceruloplasmin based on internal triplication, hydrophilic/hydrophobic character, and secondary structure of domains. *Proc. Natl. Acad. Sci. U.S.A.*, **81**, 4761–4765.
 - 9 Machonkin, T.E., Zhang, H.H., Hedman, B., Hodgson, K.O., and Solomon, E.I. (1998) Spectroscopic and magnetic studies of human ceruloplasmin: identification of a redox-inactive reduced Type 1 copper site. *Biochemistry*, **37**, 9570–9578.
 - 10 Bento, I., Peixoto, C., Zaitsev, V.N., and Lindley, P.F. (2007) Ceruloplasmin revisited: structural and functional roles of various metal cation-binding sites. *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, **63**, 240–248.
 - 11 Quintanar, L., Stoj, C., Taylor, A.B., Hart, P.J., Kosman, D.J., and Solomon, E.I. (2007) Shall we dance? How a multicopper oxidase chooses its electron transfer partner. *Acc. Chem. Res.*, **40**, 445–452.
 - 12 Taniguchi, V.T., Malmstrom, B.G., Anson, F.C., and Gray, H.B. (1982) Temperature dependence of the reduction potential of blue copper in fungal laccase. *Proc. Natl. Acad. Sci. U.S.A.*, **79**, 3387–3389.
 - 13 Li, H., Webb, S.P., Ivanic, J., and Jensen, J.H. (2004) Determinants of the relative reduction potentials of Type-1 copper sites in proteins. *J. Am. Chem. Soc.*, **126**, 8010–8019.
 - 14 Leliveld, S.R., Dame, R.T., Wuite, G.J.L., Stitz, L., and Korth, C. (2006) The expanded octarepeat domain selectively binds prions and disrupts homomeric prion protein interactions. *J. Biol. Chem.*, **281**, 3268–3275.
 - 15 Shiraishi, N., Inai, Y., Bi, W., and Nishikimi, M. (2005) Fragmentation of copper-loaded prion protein by copper-catalyzed oxidation. *Biochem. J.*, **387**, 247–255.
 - 16 Bonomo, R.P., Imperlizzeri, G., Pappalardo, G., Rizzarelli, E., and Tabbi, G. (2000) Copper(II) binding modes in the prion octapeptide PHGGGWGQ: a spectroscopic and voltammetric study. *Chemistry*, **6**, 4195–4202.
 - 17 Aronoff-Spencer, E., Burns, C.S., Avdievich, N.I., Gerfen, G.J., Peisach, J., Antholine, W.E., Ball, H.L., Cohen, F.E., Prusiner, S.B., and Millhauser, G.L. (2000) Identification of the Cu²⁺ binding sites in the N-terminal domain of the prion protein by EPR and CD spectroscopy. *Biochemistry*, **39**, 13760–13771.
 - 18 Burns, C.S., Aronoff-Spencer, E., Dunham, C.M., Lario, P., Avdievich, N.I., Antholine, W.E., ? Olmstead, M.M., Vrielink, A., Gerfen, G.J., Peisach, J., Scott, W.G., and Millhauser, G.L. (2002) Molecular features of the copper binding sites in the octarepeat domain of the prion protein. *Biochemistry*, **41**, 3991–4001.
 - 19 Burns, C.S., Aronoff-Spencer, E., Legname, G., Prusiner, S.B., Antholine, W.E., Gerfen, G.J., Peisach, J., and Millhauser, G.L. (2003) Copper coordination in the full-length, recombinant prion protein. *Biochemistry*, **42**, 6794–6803.
 - 20 Miura, T., Hori-i, A., Mototani, H., and Takeuchi, H. (1999) Raman spectroscopic study on the copper(II) binding mode of prion octapeptide and its pH dependence. *Biochemistry*, **38**, 11560–11569.
 - 21 Stöckel, J., Safar, J., Wallace, A.C., Cohen, F.E., and Prusiner, S.B. (1998) Prion protein selectively binds copper(II) ions. *Biochemistry*, **37**, 7185–7193.
 - 22 Barry, S.D., Rickard, G.A., Pushie, M.J., and Rauk, A. (2009) The Affinity of HGCG, GHGG, GGGH, and GGGH Peptides for Copper(II) and the Structures of their Complexes. An Ab Initio Study, *Can. J. Chem.*, **87**, 942–953.
 - 23 Selkoe, D.J. and Podlisny, M.B. (2002) Deciphering the genetic basis of Alzheimer's disease. *Annu. Rev. Genomics Hum. Genet.*, **3**, 67–99.

- 24 Glenner, G.G. and Wong, C.W. (1984) Alzheimer's disease: initial report for the purification and characterization of a novel cerebrovascular amyloid protein. *Biochem. Biophys. Res. Commun.*, **120**, 885–890.
- 25 Glenner, G.G. and Wong, C.W. (1984) Alzheimer's disease and Down's syndrome sharing of a unique cerebrovascular amyloid fibril protein. *Biochem. Biophys. Res. Commun.*, **122**, 1131–1135.
- 26 Hardy, J. and Selkoe, D.J. (2002) The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science*, **297**, 353–356.
- 27 Suh, Y.H. and Checler, F. (2002) Amyloid precursor protein, presenilins, and alpha-synuclein: molecular pathogenesis and pharmacological applications in Alzheimer's disease. *Pharmacol. Rev.*, **54**, 469–525.
- 28 Martin, G.M. (2000) Molecular mechanisms of late life dementias. *Exp. Gerontol.*, **35**, 439–443.
- 29 Del Rio, M.J. and Velez-Pardo, C. (2001) Apoptosis in neurodegenerative diseases: facts and controversies. *Rev. Neurol.*, **32**, 851–860.
- 30 Sayre, L.M., Zagorski, M.G., Surewicz, W.K., Krafft, G.A., and Perry, G. (1997) Mechanisms of amyloid beta deposition and the role of free radicals in the pathogenesis of Alzheimer's disease: a critical appraisal. *Chem. Res. Toxicol.*, **10**, 518–526.
- 31 Dragunow, M., MacGibbon, G.A., Lawlor, P., Butterworth, N., Connor, B., Henderson, C., Walton, M., Woodgate, A., Hughes, P., and Faull, R.L.M. (1998) Apoptosis, neurotrophic factors, and neurodegeneration. *Rev. Neurosci.*, **8**, 223–265.
- 32 Manelli, A.M. and Puttfarcken, P.S. (1995) β -Amyloid-induced toxicity in rat hippocampal cells: *in vitro* evidence for the involvement of free radicals. *Brain Res. Bull.*, **38**, 569–576.
- 33 Harrigan, M.R., Kunkel, D.D., Nguyen, L.B., and Malouf, A.T. (1995) Beta-amyloid is neurotoxic in hippocampal slice cultures. *Neurobiol. Aging*, **16**, 779–789.
- 34 Kowalik-Jankowska, T., Ruta-Dolejsz, M., Wisniewska, K., Lankiewicz, L., and Kozłowski, H. (2002) Possible involvement of copper(II) in Alzheimer disease. *Environ. Health Perspect.*, **110**, 869–870.
- 35 Lynch, T., Cherny, R.A., and Bush, A.I. (2000) Oxidative processes in Alzheimer's disease: the role of A β -metal interactions. *Exp. Gerontol.*, **35**, 445–451.
- 36 Markesbury, W.R. and Carney, J.M. (1999) Oxidative alterations in Alzheimer's disease. *Brain Pathol.*, **9**, 133–146.
- 37 Nunomura, A., Castellani, R.J., Zhu, X., Moreira, P.I., Perry, G., and Smith, M.A. (2006) Involvement of oxidative stress in Alzheimer disease. *J. Neuropathol.*, **65**, 631–641.
- 38 Mecocci, P., Polidori, M.C., Cherubini, A., Ingegni, T., Mattioli, P., Catani, M., Rinaldi, P., Cecchetti, R., Stahl, W., Senin, U., and Beal, M.F. (2002) Lymphocyte oxidative DNA damage and plasma antioxidants in Alzheimer disease. *Arch. Neurol.*, **59**, 794–798.
- 39 Montine, T.J., Markesbery, W.R., Zackert, W., Sanchez, S.C., Roberts, L.J. II, and Morrow, J.D. (1999) The Magnitude of brain lipid peroxidation correlates with the extent of degeneration but not with neuritic plaques or neurofibrillary tangles or with APOE genotype in Alzheimer's disease patients. *Am. J. Path.*, **155**, 863–868.
- 40 Atwood, C.S., Scarpa, R.C., Huang, X., Moir, R.D., Jones, W.D., Fairlie, D.P., Tanzi, R.E., and Bush, A.I. (2000) Characterization of copper interactions with Alzheimer amyloid β peptides. Identification of an attomolar-affinity copper binding site on amyloid β 1-42. *J. Neurochem.*, **75**, 1219–1233.
- 41 Bush, A.I. (2003) The metallobiology of Alzheimer's disease. *Trends Neurosci.*, **26**, 207–214.
- 42 Huang, X., Atwood, C.S., Hartshorn, M.A., Multhaup, G., Goldstein, L.E., Scarpa, R.C., Cuajungco, M.P., Gray, D.N., Lim, J., Tanzi, R.E., and Bush, A.I. (1999) The A β peptide of Alzheimer's disease directly produces hydrogen peroxide through metal ion reduction. *Biochemistry*, **38**, 7609–7617.

- 43 Bondy, S.C., Guo-Ross, S.X., and Truong, A.T. (1998) Promotion of transition metal-induced reactive oxygen species formation by beta-amyloid. *Brain Res.*, **799**, 91–96.
- 44 Butterfield, D.A. and Lauderback, C.M. (2002) Lipid peroxidation and protein oxidation in Alzheimer's disease brain: potential causes and consequences involving amyloid beta-peptide-associated free radical oxidative stress. *Free Radical Biol. Med.*, **32**, 1050–1060.
- 45 Kim, H.-Y. (2007) Novel metabolism of docosahexaenoic acid in neural cells. *J. Biol. Chem.*, **282**, 18661–18665.
- 46 Subramaniam, R., Koppal, T., Green, M., Yatin, S., Jordan, B., Drake, J., and Butterfield, D.A. (1998) The free radical antioxidant vitamin E protects cortical synaptosomal membranes from amyloid β (25–35) toxicity but not from hydroxynonenal toxicity: relevance to the free radical hypothesis of Alzheimer's disease. *Neurochem. Res.*, **23**, 1403–1410.
- 47 Kanski, J., Aksenova, M., Stoyanova, A., and Butterfield, D.A. (2002) Ferulic acid antioxidant protection against hydroxyl and peroxy radical oxidation in synaptosomal and neuronal cell culture systems *in vitro*: structure-activity studies. *J. Nutr. Biochem.*, **13**, 273–281.
- 48 Romero, F.J., Bosch-Morell, F.E., Jareno, J., Romero, B., Marin, N., and Roma, J. (1998) Lipid peroxidation products and antioxidants in human disease. *Environ. Health. Perspect.*, **106**, 1229–1234.
- 49 Suo, Z.M., Crawford, F., Fang, C.H., Paris, D., Parker, T., Placzek, A., Humphrey, J., and Mullan, M. (1998) Phenyl-N-tert-butyl nitron neutralizes the activities of beta-amyloid peptides in both cultured cells and isolated vessels. *Alzheimer's Rep.*, **1**, 381–387.
- 50 Richardson, J.S., Zhou, Y., and Kumar, U. (1996) Free radicals in the neurotoxic actions of β -amyloid. *Ann. New York Acad. Sci.*, **777**, 362–367.
- 51 Dhitavat, S., Rivera, E.R., Rogers, E., and Shea, T.B. (2001) Differential efficacy of lipophilic and cytosolic antioxidants on generation of reactive oxygen species by amyloid-beta. *J. Alzheimer's Dis.*, **3**, 525–529.
- 52 White, A.R., Bush, A.I., Beyreuther, K., Masters, C.L., and Cappai, R. (1999) Exacerbation of copper toxicity in primary neuronal cultures depleted of cellular glutathione. *J. Neurochem.*, **72**, 2092–2098.
- 53 Schöneich, C. and Williams, T.D. (2002) β -amyloid peptide targets His13 and His14 over His6: detection of 2-oxo-histidine by HPLC-MS/MS. *Chem. Res. Toxicol.*, **15**, 717–722.
- 54 Masters, C.L., Simms, G., Weinman, N.A., Multhaup, G., McDonald, B.L., and Beyreuther, K. (1985) Amyloid plaque core protein in Alzheimer disease and Down syndrome. *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 4245–4249.
- 55 Roher, A.E., Chaney, M.O., Kuo, Y.M., Webster, S.D., Stine, W.B., Haverkamp, L.J., Woods, A.S., Cotter, R.J., Tuohy, J.M., Krafft, G.A., Bonnell, B.S., and Emmerling, M.R. (1996) Morphology and toxicity of A β -(1–42) dimer derived from neuritic and vascular amyloid deposits of Alzheimer's disease. *J. Biol. Chem.*, **271**, 20631–20635.
- 56 Shoji, M., Golde, T.E., Ghiso, J., Cheung, T.T., Estus, S., Shaffer, L.M., Cai, X.D., McKay, D.M., Tintner, R., Frangione, B., and Younkin, S.G. (1992) Production of the Alzheimer amyloid β protein by normal proteolytic processing. *Science*, **258**, 126–129.
- 57 Vigo-Pelfrey, C., Lee, D., Keim, P., Lieberburg, I., and Schenk, D.B. (1993) Characterization of beta-amyloid peptide from human cerebrospinal fluid. *J. Neurochem.*, **61**, 1965–1968.
- 58 Iwatsubo, T., Odaka, A., Suzuki, N., Mizusawa, H., Nukina, N., and Ihara, Y. (1994) Visualization of A β 42(43) and A β 40 in senile plaques with end-specific A β monoclonals: evidence that an initially deposited species is A β 42(43). *Neuro*, **13**, 45–53.
- 59 Martins, R.N., Harper, C.G., Stokes, G.B., and Masters, C.L. (1986) Increased cerebral glucose-6-phosphate dehydrogenase activity in Alzheimer's disease may reflect oxidative stress. *J. Neurochem.*, **46**, 1042–1045.

- 60 Smith, M.A., Kutty, R.K., Richey, P.L., Yan, S.D., Stern, D., Chader, G.J., Wiggert, B., Petersen, R.B., and Perry, G. (1994) Heme oxygenase-1 is associated with the neurofibrillary pathology of Alzheimer's disease. *Am. J. Pathol.*, **145**, 42–47.
- 61 Smith, C.D., Carney, J.M., Starke-Reed, P.E., Oliver, C.N., Stadtman, E.R., Floyd, R.A., and Markesbery, W.R. (1991) Excess brain protein oxidation and enzyme dysfunction in normal aging and in Alzheimer's disease. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 10540–10543.
- 62 Hensley, K., Hall, N., Subramanian, R., Cole, P., Harris, M., Aksenov, M., Aksenova, M., Gabbita, S.P., Wu, J.F., Carney, J.M., Lovell, M., Markesbery, W.R., and Butterfield, D.A. (1995) Brain regional correspondence between Alzheimer's disease histopathology and biomarkers of protein oxidation. *J. Neurochem.*, **65**, 2146–2156.
- 63 Smith, M.A., Perry, G., Richey, P.L., Sayre, L.M., Anderson, V.E., Beal, M.F., and Kowall, N. (1996) Oxidative damage in Alzheimer's disease. *Nature*, **382**, 120–121.
- 64 Palmer, A.M. and Burns, M.A. (1994) Selective increase in lipid peroxidation in the inferior temporal cortex in Alzheimer's disease. *Brain Res.*, **645**, 338–342.
- 65 Sayre, L.M., Zelasko, D.A., Harris, P.L.R., Perry, G., Salomon, R.G., and Smith, M.A. (1997) 4-Hydroxynonenal-derived advanced lipid peroxidation end products are increased in Alzheimer's disease. *J. Neurochem.*, **68**, 2092–2097.
- 66 Smith, M.A., Richey Harris, P.L., Sayre, L.M., Beckman, J.S., and Perry, G. (1997) Widespread peroxynitrite-mediated damage in Alzheimer's disease. *J. Neurosci.*, **17**, 2653–2657.
- 67 Mecocci, P., MacGarvey, U., and Beal, M.F. (1994) Oxidative damage to mitochondrial DNA is increased in Alzheimer's disease. *Ann. Neurol.*, **36**, 747–751.
- 68 Huang, X., Atwood, C.S., Hartshorn, M.A., Multhaup, G., Goldstein, L.E., Scarpa, R.C., Cuajungco, M.P., Gray, D.N., Lim, J., Moir, R.D., Tanzi, R.E., and Bush, A.I. (1999) The A β peptide of Alzheimer's disease directly produces hydrogen peroxide through metal ion reduction. *Biochemistry*, **38**, 7609–7616.
- 69 Huang, X., Cuajungco, M.P., Atwood, C.S., Hartshorn, M.A., Tyndall, J.D.A., Hanson, G.R., Stokes, K.C., Leopold, M., Multhaup, G., Goldstein, L.E., Scarpa, R.C., Saunders, A.J., Lim, J., Moir, R.D., Glabe, C., Bowden, E.F., Masters, C.L., Fairlie, D.P., Tanzi, R.E., and Bush, A.I. (1999) Correlation with cell-free hydrogen peroxide production and metal reduction. *J. Biol. Chem.*, **274**, 37111–37116.
- 70 Curtain, C.C., Ali, F., Volitakis, I., Cherny, R.A., Raymond, S., Norton, R.S., Beyreuther, K., Barrow, C.J., Masters, C.L., Bush, A.I., and Barnham, K.J. (2001) Alzheimer's disease amyloid- β binds copper and zinc to generate an allosterically ordered membrane-penetrating structure containing superoxide dismutase-like subunits. *J. Biol. Chem.*, **276**, 20466–20473.
- 71 Opazo, C., Huang, X., Cherny, R.A., Moir, R.D., Roher, A.E., White, A.R., Cappai, R., Masters, C.L., Tanzi, R.E., Inestrosa, N.C., and Bush, A.I. (2002) Metalloenzyme-like activity of Alzheimer's disease beta-amyloid. Cu-dependent catalytic conversion of dopamine, cholesterol, and biological reducing agents to neurotoxic H₂O₂. *J. Biol. Chem.*, **277**, 40302–40308.
- 72 Opazo, C., Barriá, M.I., Ruiz, F.H., and Inestrosa, N.C. (2003) Copper reduction by copper binding proteins and its relation to neurodegenerative diseases. *BioMetals*, **16**, 91–98.
- 73 Guilloreau, L., Combalbert, S., Sournia-Saquet, A., Mazarguil, H., and Faller, P. (2007) Redox chemistry of copper-amyloid- β : the generation of hydroxyl radical in the presence of ascorbate is linked to redox-potentials and aggregation state. *ChemBioChem*, **8**, 1317–1325.
- 74 Jiang, D., Men, L., Wang, J., Zhang, Y., Chikenyan, S., Wang, Y., and Zhou, F. (2007) Redox reactions of copper complexes formed with different β -amyloid peptides and their neuropathological relevance. *Biochemistry*, **46**, 9270–9282.
- 75 For a review, see: Brunelle, P., Raffa, D.F., Rickard, G.A., Gómez-Balderas, R.,

- Armstrong, D.A., and Rauk, A. (2006) The "theoretical" chemistry of Alzheimer's disease: the radical model, in *Modelling Molecular Structure and Reactivity in Biological Systems* (Series: Special Publication no. 304), (eds K.J. Naidoo, M. Hann, J. Gao, M. Field, and J. Brady), Royal Society of Chemistry, Cambridge.
- 76 Rauk, A., Yu, D., Taylor, J., Shustov, G.V., Block, D.A., and Armstrong, D.A. (1999) Effects of structure on the alpha-C-H bond enthalpies of amino acid residues: relevance to H transfers in enzyme mechanisms and in protein oxidation. *Biochemistry*, **38**, 9089–9096.
- 77 Rauk, A. and Armstrong, D.A. (2000) Influence of beta-sheet structure on the susceptibility of proteins to backbone oxidative damage: preference for alpha-C-centered radical formation at glycine in antiparallel beta-sheets. *J. Am. Chem. Soc.*, **122**, 4185–4192.
- 78 Rauk, A., Armstrong, D.A., and Fairlie, D.P. (2000) Is oxidative damage by beta amyloid and prion peptides mediated by hydrogen atom transfer from glycine alpha-carbon to methionine sulfur within beta-sheets? *J. Am. Chem. Soc.*, **122**, 9761–9767.
- 79 Brunelle, P., Schoeneich, C., and Rauk, A. (2006) One-electron oxidation of methionine peptides: stability of the three-electron SN(amide) bond. *Can. J. Chem.*, **84**, 893–904.
- 80 Brunelle, P. and Rauk, A. (2004) One-electron oxidation of methionine in peptide environments: the effect of three-electron bonding on the reduction potential of the radical cation. *J. Phys. Chem. A*, **108**, 11032–11041.
- 81 Raffa, D.F., Gómez-Balderas, R., Brunelle, P., Rickard, G.A., and Rauk, A. (2005) Ab initio model studies of copper binding to peptides containing a His–His sequence: relevance to the β -amyloid peptide of Alzheimer's disease. *J. Biol. Inorg. Chem.*, **10**, 887–902.
- 82 Raffa, D., Rickard, G.A., and Rauk, A. (2007) Ab initio modelling of the structure and redox behaviour of copper(I) bound to a His–His model peptide: relevance to the beta amyloid peptide of Alzheimer's disease. *J. Biol. Inorg. Chem.*, **12**, 147–164.
- 83 Hallman, P.S., Perrin, D.D., and Watt, A.E. (1971) The computed distribution of copper(II) and zinc(II) ions among seventeen amino acids present in human blood plasma. *Biochem. J.*, **121**, 549–555.
- 84 Neumann, P.Z. and Sass-Kortsak, A.J. (1967) The state of copper in human serum: evidence for an amino acid-bound fraction. *J. Clin. Invest.*, **46**, 646–658.
- 85 Syme, C.D., Nadal, R.C., Rigby, S.E.J., and Viles, J.H. (2004) Copper binding to the amyloid-beta (A beta) peptide associated with Alzheimer's diseases – Folding, coordination geometry, pH dependence, stoichiometry, and affinity of A beta- (1–28): insights from a range of complementary spectroscopic techniques. *J. Biol. Chem.*, **279**, 18169–18177.
- 86 Miura, T., Suzuki, K., Kohata, N., and Takeuchi, H. (2000) Metal binding modes of Alzheimer's amyloid β -peptide in insoluble aggregates and soluble complexes. *Biochemistry*, **39**, 7024–7031.
- 87 Schöneich, C. and Williams, T.D. (2000) Cu(II)-catalyzed oxidation of β -amyloid peptide targets His13 and His14 over His6: detection of 2-oxo-histidine by HPLC-MS/MS. *Chem. Res. Toxicol.*, **15**, 717–722.
- 88 Tickler, A.K., Smith, D.G., Ciccotosto, G.D., Tew, D.J., Curtain, C.C., Carrington, D., Masters, C.L., Bush, A.I., Cherney, R.A., Cappai, R., Wade, J.D., and Barnham, K.J. (2005) Methylation of the imidazole side chains of the Alzheimer disease amyloid-beta peptide results in abolition of superoxide dismutase-like structures and inhibition of neurotoxicity. *J. Biol. Chem.*, **280**, 13355–13363.
- 89 Smith, D.P., Smith, D.G., Curtain, C.C., Boas, J.F., Pilbrow, J.R., Ciccotosto, G.D., Lau, T.-L., Tew, D.J., Perez, K., Wade, J.D., Bush, A.I., Drew, S.C., Separovic, F., Masters, C.L., Cappai, R., and Barnham, K.J. (2006) Copper-mediated amyloid- β toxicity is associated with an intermolecular histidine bridge. *J. Biol. Chem.*, **281**, 15145–15154.
- 90 Dai, X.-L., Sun, Y.-X., and Jiang, Z.-F. (2006) Cu(II) potentiation of Alzheimer A β 1-40 cytotoxicity and transition on its secondary structure. *Acta Biochim. Biophys. Sin.*, **38**, 765–772.

- 91 Gómez-Balderas, R., Raffa, D.F., Rickard, G.A., Brunelle, P., and Rauk, A. (2005) Binding of copper ions to methionine peptide models: relevance to Alzheimer's disease. *J. Phys. Chem. A*, **109**, 5498–5508.
- 92 Leung, B.O. and Rauk, A. (2005) Dialkyl sulfur radical cations: competition between proton and methyl cation transfers to sulfur nucleophiles: an ab initio study. *Mol. Phys.*, **103**, 1201–1209.
- 93 Raffa, D.F. and Rauk, A. (2007) Molecular dynamics study of the beta amyloid peptide of Alzheimer's disease and its divalent copper complexes. *J. Phys. Chem. B*, **111**, 3789–3799.
- 94 Martin, M.J., Endicott, J.F., Ochrymowycz, L.A., and Rorabacher, D.B. (1987) Structure-reactivity relationships in copper(II)/copper(I) electron-transfer kinetics: evaluation of self-exchange rate constants for copper polythia ether complexes. *Inorg. Chem.*, **26**, 3012–3022.
- 95 Rauk, A. (2006) Why is the amyloid beta peptide of Alzheimer's disease neurotoxic?. *Dalton Trans.*, 1273–1282.
- 96 Wagman, D.D., Evans, W.H., Parker, V.B., Schumm, R.H., Halow, I., Bailey, S.M., Churney, K.L., and Nuttall, R.L., (1982) *J. Phys. Chem. Ref. Data*, (Supl. 1), 11.
- 97 Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E. et al. (2001) *Gaussian 98, Revision A11*, Gaussian, Inc., Pittsburgh PA.
- 98 Frisch, M.J., Trucks, G.W., Schlegel, H.B. Scuseria, G.E. et al. (2003) *Gaussian 03, Revision B04*, Gaussian, Inc., Pittsburgh PA.
- 99 Schaftenaar, G. and Noordik, J.H. (2000) Molden 4.3. *J. Comput.-Aided Mol. Design*, **14**, 123–134.
- 100 Becke, A.D. (1993) Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, **98**, 5648–5652.
- 101 Scott, A.P. and Radom, L. (1996) Harmonic vibrational frequencies: an evaluation of Hartree-Fock, Moller-Plesset, quadratic configuration interaction, density functional theory, and semiempirical scale factors. *J. Phys. Chem.*, **100**, 16502–16513.
- 102 Ducéré, J.M., Goursot, A., and Berthomieu, D. (2005) Comparative density functional theory study of the binding of ligands to Cu^+ and Cu^{2+} : influence of the coordination and oxidation state. *J. Phys. Chem. A*, **109**, 400–408.
- 103 Rannulu, N.S. and Rodgers, M.T. (2005) Solvation of copper ions by imidazole: structures and sequential binding energies of Cu^+ (imidazole) $_x$, $x = 1-4$. Competition between ion solvation and hydrogen bonding. *Phys. Chem. Chem. Phys.*, **7**, 1014–1025.
- 104 Vitale, G., Valina, A.B., Huang, H., Amunugama, R., and Rodgers, M.T. (2001) Solvation of copper Ions by acetonitrile. Structures and sequential binding energies of Cu^+ (CH₃CN) $_x$, $x = 1-5$ from collision-induced dissociation and theoretical studies. *J. Phys. Chem. A*, **105**, 11351–11364.
- 105 Feller, D. and Dixon, D.A. (2002) Metal ion binding: an electronic structure study of M^+ (dimethyl ether) $_n$, $\text{M} = \text{Cu}, \text{Ag}$ and Au , and ($n = 1-4$) complexes. *J. Phys. Chem. A*, **106**, 5136–5143.
- 106 Weast, R.C. (ed.) (1977–1978) *Handbook of Chemistry and Physics*, 58th edn, CRC Press.
- 107 Liptak, M.D., Gross, K.C., Seybold, P.G., Feldus, S., and Shields, G.C. (2002) Absolute pKa determination for substituted phenols. *J. Am. Chem. Soc.*, **124**, 6421–6427.
- 108 James, B.R. and Williams, R.J.P. (1961) The oxidation–reduction potentials of some copper complexes. *J. Chem. Soc.*, 2007–2019.

30

Theoretical Investigation of NSAID Photodegradation Mechanisms

Klefah A.K. Musa and Leif A. Eriksson

30.1

Drug Safety

Any drug, in general, whether prescription or over-the-counter (OTC), must be safe and have beneficial effects on the body by curing, controlling, preventing disease or relieving symptoms before it can be dispensed or administered. Drug safety evaluation is obtained by looking at its side effects or risks (most are minor, but some can be serious or in rare events even life-threatening), manufacturing processes, results of preclinical testing and clinical trials, and more [1–3].

Normally, the common side effects or risks are clearly exposed before a drug is approved. However, clinical trials, which typically involve several thousand persons, do not reveal all of the side effects or risks. Rare side effects may appear only after the drug is used in large quantities among the population. Side effects associated with the medicine may upon recognition help in identifying risk factors for a particular adverse reaction or ways in which a medicine can be used more safely. The adverse reactions are of two types: (i) Severe, which refers to the degree of harm, disability or effect on quality of life. They are not life-threatening or result in hospital admission. (ii) Serious reactions, which may be fatal, life-threatening, disabling or incapacitating, and cause hospital admission or prolonged hospital stay. Cardiotonic glycosides, adrenal corticosteroidal, anticoagulants, opiates and related narcotics, antibiotics, analgesics, and antineoplastic and immunosuppressive drugs are examples of compounds associated with a high risk of adverse drug reactions [1].

In 2000, approximately 25 000 adverse reaction reports were included in the Irish Medicines Board, Adverse Drug Reaction (IMBs ADR) database, provided by health care professionals and pharmaceutical companies in Ireland since 1968 [4]. In the United States, adverse drug reactions are considered a significant public health problem. For the 12 261 737 Medicare admitted to 6238 hospitals in USA during 1998, adverse drug reactions were projected to cause 2976 deaths and 118 200 patient-days with \$516 034 829 in total charges, \$37 611 868 in drug charges and \$9 456 698 in laboratory charges. Such data is helpful when evaluating the safety profile of

medicinal products, and when dealing with enquiries from healthcare professionals [5].

Drug side effects can be very diverse, like gastrointestinal [6], CNS disorders [7], nephrotoxicity [8], hematotoxicity [9] and hepatotoxicity [10]. In addition to these common side effects, the drugs can also affect eyes and vision. These include eye drops, other topical eye medications, pills and more [11]. Some drugs, for example, NSAIDs, psoralen and antibiotics have the ability to cause local reactions that are usually described as “phototoxic” or “photoallergic” responses in the skin. Both types of reactions are referred to as “photosensitivity” [12–21]. In this chapter we discuss in more detail photosensitivity reactions based on results obtained from theoretical perspectives to get an overview of this type of side effects and provide information that may be helpful in designing new drugs with less or no adverse side effects of this kind.

30.2

Drug Photosensitivity

Photosensitization is referred to as the reaction between artificial or normal sunlight and a chemical substance that has the capability to absorb light at an appropriate wavelength. It requires the presence of the compound, known as photosensitizer, in the biological medium. The photosensitizer is a drug, food, cosmetic or other chemical that increases the photosensitivity of the organism. Pharmacologically, it is considered an adverse side effect of many drugs, or a desired effect in photodynamic therapy of cancer and various dermatological conditions [22, 23]. In photochemistry a photosensitizer is defined as a chemical compound that readily undergoes photoexcitation and then transfers its energy to other molecules, subsequently making the reaction mixture more sensitive to light [23]. The effect of the photosensitization process varies from simple rashes to severe cutaneous afflictions that may even lead to damage of internal organs.

The appropriate wavelength for many pharmaceutical drugs is located in the UV/Vis region of the spectrum. UV radiation is an inherent part of our everyday life, characterized by its short wavelength (high energy) and capacity to pass through the different layers of the atmosphere and reach the Earth. The wavelength range of UV-B radiation is 290–320 nm and for UV-A it is 320–400 nm. UV-B frequencies are absorbed in the first layers of the skin, composed of dead cells of the stratum corneum. However, wavelengths in the UV-A region can reach the blood system. In addition to direct responses of UV-A and UV-B exposure, the human system can be subjected to sunlight-mediated effects involving endogenous and exogenous photosensitizers [24–26]. The exogenous (xenobiotic) species, such as pharmaceutical products transported through the blood system, will frequently reach superficial areas in the body, where they can readily absorb the incident radiation. These xenobiotic–incident sunlight interactions can be very detrimental for living tissues since they can result in photoallergic, photophobic, and phototoxic responses.

The photosensitization reactions can be divided according to four different pathways: (i) energy transfer, (ii) electron or hydrogen transfer, (iii) covalent photobinding

and (iv) decomposition [26]. In general, these reactions are classified into main reactions known as Type I and Type II; in Type I, transfer of an electron between a photosensitizer and substrate takes place, resulting in the creation of highly reactive radical species, molecules that have an uneven number of electrons. The radicals can react with biological substrates producing changes in structure and/or function. Type II reactions are mediated by transfer of energy from the photosensitizer to molecular oxygen to produce an excited singlet state of oxygen, which in turn might participate in lipid peroxidation, protein oxidation or induce DNA damage. In biological systems, distinguishing between Type I and Type II reaction mechanisms is not always easy. However, the generation of singlet oxygen by a photosensitization reaction can be confirmed by detecting light emission at 1270 nm produced by the spontaneous decay of singlet oxygen to its ground state. Direct detection of Type I (radical) mechanisms is more technically demanding, requiring electron spin resonance measurements to trap and detect radical species [26, 27].

There are furthermore different modes of photosensitivity disorders caused by pharmaceutical products, generally classified into three main categories as follows.

30.2.1

Photoallergies

In photoallergic reactions [28], which generally occur due to medication applied to the skin, UV-light may change the drug structurally, causing the skin to produce antibodies. The result is an allergic reaction. Symptoms can appear within 20 seconds after sun exposure but can sometimes also be delayed (e.g., in the case of suprofen) [29], producing eczema-like skin conditions that can spread to non-exposed parts of the body. Cosmetics can also cause photoallergic reactions, especially if containing musk ambrette or sandalwood oil. Other products, like some quinolone antibacterial drugs and the OTC NSAID pain relievers containing ibuprofen and naproxen sodium are good examples of photoallergic substances.

30.2.2

Photophobia

Some drugs can cause photophobia, “fear of light.” In photophobic photosensitivity disorders, patients avoid light because their eyes are painfully sensitive to it. Examples of medications that induce photophobia are Crystodigin (digitoxin) and Duraquine (quinidine) and several drugs for diabetes, such as Tolinase (tolazamide) and Orinase (tolbutamide) [30].

30.2.3

Phototoxicity

Many drugs can cause phototoxic reactions [15, 25, 26, 28, 29]. These do not affect the body’s immune system but arise from damaging effects of highly activated compounds on membranes and, in some instances, DNA. They are more common

than photoallergic reactions, and occur in response to injected, oral or topically applied medications. This category will be explored in more detail in the current chapter.

30.3

Non-Steroid Anti-Inflammatory Drugs (NSAIDs)

30.3.1

NSAID: Definition and Classification

In 1829, the isolation of salicin from the folk remedy willow bark was first performed and opened the stage for non-steroidal anti-inflammatory drug (NSAID) therapy [40]. These agents are also sometimes referred as non-steroidal anti-inflammatory agents/analgesics (NSAIAs). NSAIDs or NSAIAs have long been used worldwide to treat inflammation and to relieve pain in both humans and animals. In addition to the traditional members of this drug family such as aspirin and ibuprofen, a huge number of new drug members have been introduced to the community by pharmaceutical companies. Moreover, a new generation of drugs in this family has been designed based on enzymatic selectivity and introduced to the market as celecoxib, rofecoxib, valdecoxib and paracoxib sodium [31–35].

NSAIDs can be classified from various criteria such as chemical structure and selectivity. Based on their selectivity towards their cyclooxygenase enzyme targets (COX; see below), they are classified into Group 1: selective COX-1 inhibitors, for example, low dose aspirin; Group 2: non-selective COX inhibitors, for example, high dose aspirin, diclofenac, ibuprofen, naproxen, paracetamol and piroxicam; Group 3: selective COX-2 inhibitors, for example, ketorolac, meloxicam, nabumetone and nimuselide; Group 4: highly selective COX-2 inhibitors, for example, celecoxib and rofecoxib [36–38]. Considering their chemical structures, the NSAIDs are instead classified into the groups I: salicylates (aspirin), II: indoles (indomethacin), III: pyrazoles (phenyl butazone), IV: fenamates (mefenamic acid), V: propionic acids (ibuprofen, ketoprofen, flurbiprofen, naproxen), VI: phenylacetic acids (diclofenac, aceclofenac), VII: oxicams (piroxicam, tenoxicam, meloxicam), VIII: sulfonanilides (nimesulide) and IX: coxibs (celecoxib, rofecoxib, valdecoxib, parecoxib, etoricoxib, lumiracoxib) [38]. According to their chemical structures, nearly all of them share an acidic character with a pK_a of 3–5, and tightly bind to plasma proteins, usually albumin [39, 40].

30.3.2

Pharmacological Action

NSAIDs function by inhibiting an endogenous enzyme known as cyclooxygenase (COX), which catalyzes the first committed step in the conversion of arachidonic acid into prostaglandins and thromboxanes. These molecules mediate different important functions in the hematic, renal and gastric systems and regulate processes such

as inflammation and body temperature [41, 42]. NSAID action results in their anti-inflammatory, analgesic, antipyretic and antithrombotic activity [43]. Some NSAIDs exhibit a time-independent inhibition, in which the ligand competes reversibly with the natural substrate to form an enzyme–inhibitor complex. An exception is acetylsalicylic acid, which has an irreversible inhibition mode [44]. Others display a time-dependent inhibition, including an initial reversible binding that progresses to a tight irreversible one to form a new enzyme–inhibitor complex.

In the early 1990s, two isoforms of the COX enzyme were identified, and named COX-1 and COX-2. [43, 45, 46]. The COX-1 isozyme plays an important role in the maintenance of normal physiologic states in many tissues, including the platelets, kidney and gastrointestinal tract; for example, COX-1 activation in the gastric mucosa leads to prostaglandin PGI₂ (prostacyclin) production, which is cytoprotective. Thromboxane A₂, which is primarily synthesized in platelets through COX-1 activity, causes platelet aggregation, vasoconstriction and smooth muscle proliferation. This implies that inhibition of this isozyme by, for example, NSAIDs will lead to various side effects, including gastrointestinal damage [31, 45, 47].

The COX-2 isozyme was discovered in 1991. One year later, the first lead inhibitors were described and within seven years the first member of selective inhibitors was on the market [45, 55]. COX-2 is less widely expressed, but is induced by pro-inflammatory stimuli, and catalyses the production of prostaglandins that mediate inflammation. The anti-inflammatory efficacy of NSAID is thus believed to result from inhibition of COX-2. At therapeutic doses, COX-2 inhibitors affect COX-2 but not COX-1, which is why it was postulated that COX-2 inhibitors relieve inflammation with less gastrointestinal toxicity than conventional NSAIDs. The COX-2 isozyme has many other functions besides its role in inflammation, such as mediating production of prostaglandin PGI₂ (prostacyclin) in the vascular endothelium [48]. The COX-2 gene is highly inducible, and thus the expression of COX-2 and prostaglandins was observed to be increased in intestinal tumors in rodents [49, 50] and humans [51]. From a molecular biology perspective, COX-1 and COX-2 are of similar molecular weight (70 and 72 kDa, respectively), have nearly identical catalytic sites and about 65% amino acid sequence homology [52]. The most significant difference between the two isozymes, which allows for selective inhibition, is the substitution of isoleucine at position 523 in COX-1 with valine in COX-2. The smaller Val₅₂₃ residue in COX-2 allows access to a hydrophobic side-pocket in the active site (which Ile₅₂₃ sterically hinders) [53]. Despite the overall high sequence similarity, the active site cavity contains a large number of nonconserved residues between the two isozymes, including (residues in parenthesis correspond to COX-2): Leu-357 (Phe), Val-119 (Ser), Leu-115 (Tyr), Ile-112 (Leu), Leu-92 (Ile) and Thr-89 (Val) [54].

30.3.3

NSAID Uses

NSAIDs are mainly used to reduce pain, fever and inflammation, that is, they are analgesic, antipyretic and anti-inflammatory agents, respectively [40]. Their specific uses include the treatment of headaches, sports injuries, arthritis (rheumatoid

arthritis or the more severe forms of osteoarthritis), post-operative pain, pain from kidney stones (renal colic), dental pain, migraine, menstrual cramps and dysmenorrhoea. Acetylsalicylic acid (also a NSAID) is used to prevent strokes and heart attacks in individuals at high risk by inhibiting blood clotting. More recently, the use of NSAIDs has become more interesting due to research into other areas of treatments by this drugs family. For example, the literature shows that NSAIDs may protect against the development of Alzheimer's disease [56–58]. However, other studies indicate that rofecoxib or low-dose naproxen does not slow cognitive decline in patients with mild-to-moderate Alzheimer's disease [59, 60]. Very recent work shows that acetylsalicylic acid decreases the hepatitis C viral expression *in vitro*. The inhibition of HCV-RNA and protein expression occurs during the COX-2 signaling pathways. Thus, it could be considered an excellent adjuvant in the treatment of chronic HCV infection if confirmed *in vivo* [61]. Various NSAIDs, furthermore, have antitumor potentials, such as amide derivatives of fenoprofen and ketoprofen, which display modest antiproliferative activity against tumor *in vitro*. A stronger cytostatic activity of these agents is expected due to greater lipophilicity and/or better cell uptake. The major mechanisms of NSAID antitumor activity is apoptosis and cell cycle arrest [62]. Along this line, the NSAID diclofenac has protective effects against ER-stress-induced apoptosis, which may indicate a new application of NSAIDs as anti-neurodegenerative agents being possible in the future [63].

30.3.4

Side Effects

The deleterious side effects of NSAIDs vary from minor to severe in different parts of the body. Continuous administration of these drugs will normally effect the kidney and gastrointestinal (GI) systems, causing nephrotoxicity and gastric ulcerations, respectively [64, 65]. The serious GI complications due to NSAIDs have been well documented [66]. NSAIDs have also been associated with severe cardiovascular risks, and for this reason some of them have been withdrawn from the market [67]. In this respect, the Food and Drugs Administration (FDA) in the US and the European Medicines Agency (EMA) have issued special caution and restrictions regarding prescription of COX-2 inhibitors, particularly those showing increased cardiovascular risk, for long-term use. In addition, the FDA issued cardiovascular warnings regarding nonselective NSAIDs. The main recommendation from both the EMA and the FDA is the lowest effective dose of NSAIDs for the shortest duration [66]. Some NSAIDs such as indomethacin have an adverse effect on the central nervous system [68, 69]. A remarkably high frequency of unwanted renal side effects causing clinical syndromes such as acute renal failure, acute interstitial nephritis, worsening of chronic kidney disease, salt and water retention and hypertension have also been associated with NSAID usage [70]. Bleeding, platelet dysfunction and severe thromboembolic events have also been noted [71]. Moreover, after some surgical procedures, NSAIDs are considered a risk factor of severe postoperative bleeding requiring surgical hemostasis, while after other surgical procedures the selective COX-2 inhibitor (coxibs) have been shown to promote arterial thrombosis [72]. In addition

to the effects on the above-mentioned systems, the liver and ear are also affected by these drugs, leading to hepatic insufficiency [73] and ototoxicity [74], respectively. NSAIDs do not cause side effects only on the endogenous systems but they also affect the exogenous one, leading to photoallergy, especially if their pharmaceutical dosage forms are as topical medications. For example, in a total of 139 contact reactions to topical NSAIDs ketoprofen was found being responsible for 28% of the allergies and 82% of the contact photoallergies [75], and this and other drugs from this family have been found to cause photosensitivity reactions, including both photoallergic and phototoxic reactions [76]. The theoretical mechanisms of a set of NSAIDs causing the latter side effect will be discussed in more detail in this chapter.

30.4

NSAID Phototoxicity

Phototoxic reactions are a common result of the interaction of sunlight with pharmaceutical agents transported in the blood or applied topically, and occur as side effect of various different drug families, including NSAIDs. The reactions depend on the wavelength of the light and the chemical structure of drug. The phenomenon has been extensively studied experimentally and is well documented for different NSAIDs, for instance, ketoprofen [76–80], ibuprofen [81–85], naproxen [85–90], nabumetone active form MNAA¹⁾ [91–94], suprofen [95–97], flurbiprofen [82, 98] and diclofenac [99–101]. Experimentally most studies are made on cell systems and biological targets such as cell membranes and DNA. The drugs have been found to cause cell damage by membrane lysis. Red blood cell photosensitized lysis has been employed as an indicator of membrane damage. The experimental findings also indicate that some NSAIDs are capable of photoinduced lipid peroxidation processes by photoperoxidation of linoleic acid [76].

In the last two decades, a considerable amount of research has been carried out to understand both the unimolecular deactivation pathways of photoexcited pharmaceutical products and their photosensitizing capabilities in the presence of substrates. Figure 30.1 summarizes the different molecular mechanisms of photosensitizing drugs leading to phototoxic response [102].

Absorption of radiation at appropriate wavelength by the photosensitizing drug leads to an excited state. The life time of an excited singlet state is very short (10^{-10} – 10^{-9} s). Radiative (fluorescence) or non-radiative processes (internal conversion or intersystem crossing, ISC) are the monomolecular deactivation mechanisms of the excited electronic singlet state. Through ISC, an excited triplet state with a much longer lifetime (10^{-6} – 10^{-3} s) will be formed. Subsequently, the formation of excited singlet oxygen can occur, a pathway known as a Type II reaction. Another pathway leading to phototoxic response is through formation of free radicals (Type I

1) Nabumetone (NB) is transformed into the pharmacologically active form MNAA by extensive first-pass metabolism. This drug has a half-life of approximately 24 h, and is known to cause photosensitivity and skin lesions in patients.

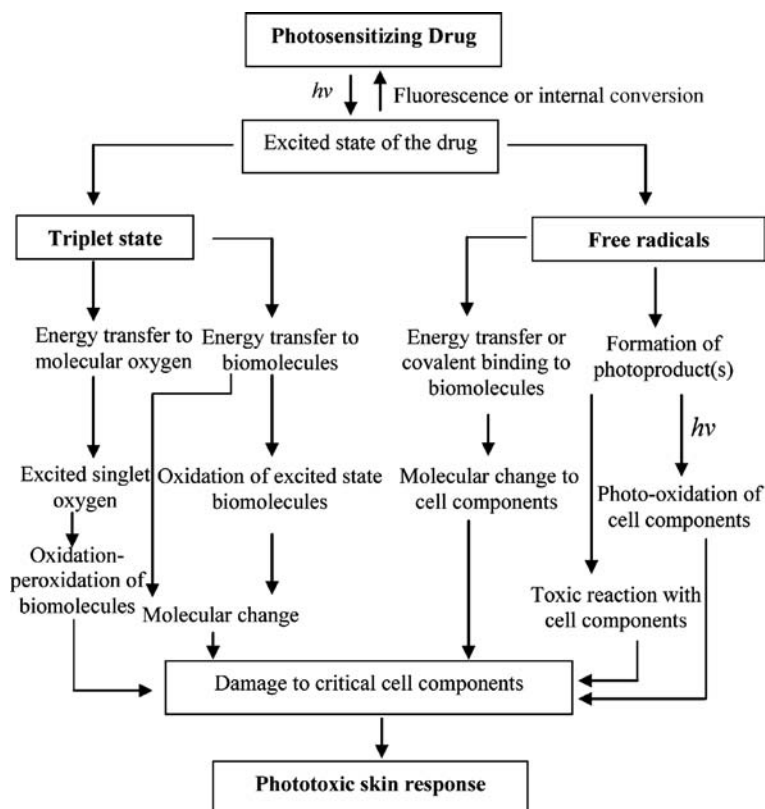


Figure 30.1 Different pathways leading to phototoxic response [102].

reaction). The photosensitizer may bind to a particular macromolecular cell component by covalent photobinding, inducing cell damage, or undergo decomposition resulting in toxic photoproducts or new photosensitizers [26].

30.5

Theoretical Studies

30.5.1

Overview

As mentioned earlier, numerous experimental studies have investigated the photo-decomposition mechanisms of NSAIDs resulting in phototoxic side effects [76–101]. NSAID photodegradation mechanisms can also be explored in detail by using computational tools. Unfortunately, the number of theoretical studies investigating this problem is far less, primarily due to the lack of appropriate and sufficiently accurate methodologies in the past. The advent of methods based on DFT (density

functional theory) for excited states has, however, made such studies possible. In this chapter we summarize our recent theoretical work related to this problem, to understand in more detail the NSAID photodegradation mechanisms. This may, in turn, provide new strategies to reduce or prevent this side effect, and to develop new safer pharmacological drugs with less or no phototoxic side effect.

The NSAIDs ketoprofen (KP) [103], ibuprofen (IBU) [104], flurbiprofen (FBP) [105], suprofen (SUP) [106], naproxen (NP) [107], nabumetone active form (6-methoxynaphthylacetic acid; MNAA) [107] and diclofenac (DF) [108] were studied theoretically using quantum computational tools. The conclusions drawn from the different studies are well and in line with experimental findings. Theoretical results will be discussed in the sections below for such as electron affinity, ionization potential, proton affinity, singlet–triplet gaps, orbital configuration, computed UV-spectra, possible pathways of decarboxylation (the main critical step in those photodegradation mechanisms), expected ways for forming different reactive oxygen species, and possible reactions between radicals formed and macromolecules such as lipids. Figure 30.2 shows the chemical structures of the series of NSAIDs investigated.

An example of NSAID photodegradation mechanisms explored is outlined for KP in Figure 30.3. At physiological pH, the drug is predominantly present in its deprotonated form (1^-). Once this species absorbs UV-radiation and through ISC reaches the first excited triplet state, it will undergo spontaneous decarboxylation to give the decarboxylated triplet anion (**3**). The latter is followed by several possible pathways; based on the intermediate energies we conclude the following:

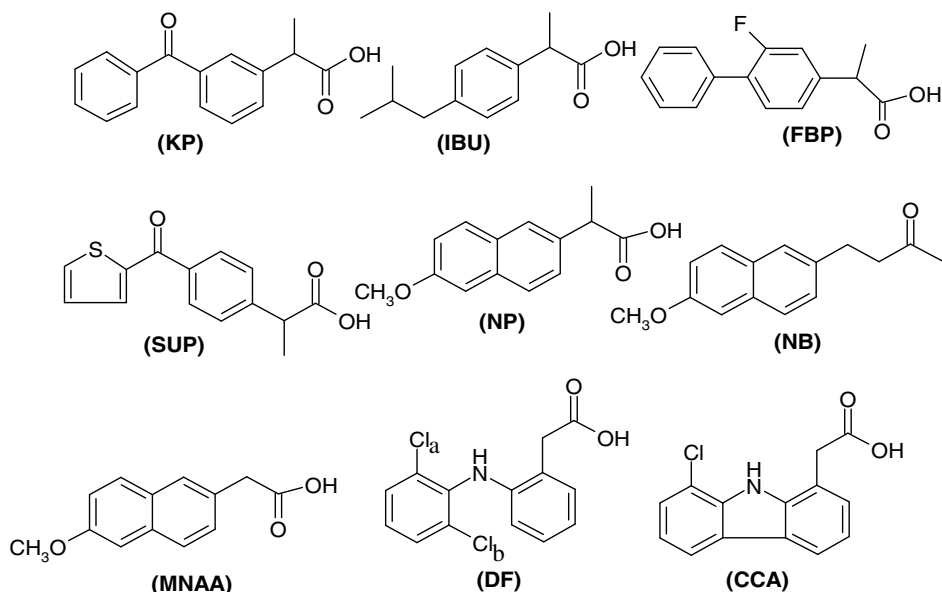


Figure 30.2 Chemical structures of the NSAIDs under study (NB is transformed into the pharmacologically active form MNAA; CCA is the main photoproduct of DF).

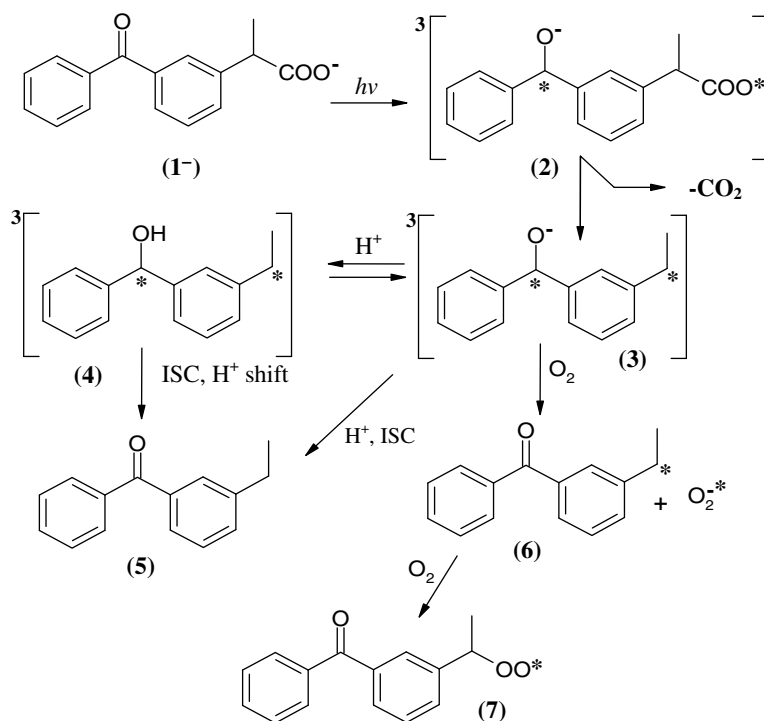


Figure 30.3 Proposed KP photodegradation mechanism.

(i) Protonation of the carboxylic oxygen (the main negative site in **3**) yields the triplet form of compound **4**. There is, however, energetically no obvious route from the **4** triplet by ISC or hydrogen/proton shift to reach the singlet state of **5**, which explains the long lifetime of **4** ($4\ \mu\text{s}$) observed experimentally. (ii) The route from the triplet radical anion **3** to ground singlet state **5** is suggested to proceed by way of the less stable triplet state of **5**, followed by ISC and decay. In this step, there is a possible generation of singlet oxygen. (iii) Electron transfer from **3** is possible in the presence of molecular oxygen, leading to the formation of species **6** and superoxide. (iv) A second molecule of oxygen can be added to radical intermediate **6** in a strictly exothermic reaction to form peroxyl radical **7**, which may either undergo further oxidation to the corresponding alcohol or keto forms or initiate lipid peroxidation.

30.5.2

Methodology

All ground-state singlet and doublet radical states, as well as excited triplet states, along the reaction pathways of each drug mentioned above were optimized at the hybrid Hartree–Fock DFT functional B3LYP level of theory [109–111] in conjunction with the 6-31G(d,p) basis set. In addition, the adiabatic electron affinities and

ionization potentials of the neutral ground states of the parent compounds were computed. Frequency calculations were performed on the optimized geometries at the same level of theory to ensure that the systems are local minima (no imaginary vibrational frequencies) and to extract zero-point vibrational energies (ZPE) and thermal corrections to the Gibbs free energies at 298 K. Solvent effects were taken into consideration implicitly, through single-point calculations on the optimized geometries of each drug at the same level of theory, including the integral equation formulation of the polarized continuum model (IEFPCM) [112–114]. Water was used as solvent, through the value 78.31 for the dielectric constant in the IEFPCM calculations. To explore in more detail the effect of the solvent on geometric structures and on the triplet state decarboxylation reaction, optimizations of certain species from each drug were also performed within the IEFPCM environment. Vertical singlet and triplet excitation energies of the protonated and deprotonated forms of each drug were determined using the time-dependent (TD) formalism [115–117], at the B3LYP/6-31G(d,p) level of theory. A scanning approach was employed to investigate the decarboxylation processes of neutral and deprotonated forms of each drug in their excited singlet and triplet states. To investigate the decarboxylation processes from the singlet excited states, the ground state structures were optimized in each step, and at each optimized point computing the lowest excited singlet states. All calculations were performed using the Gaussian03 program package [118].

30.6 Redox Chemistry

As first characterization of the various NSAIDs under study, we investigated the redox chemistry and structural features of each drug as these play a prime role in the different mechanisms. The key differences in optimized geometric structures of the neutral singlet ground states, the radical anion and cation, the deprotonated species and the first excited triplet state of the neutral and deprotonated forms of each drug are as follows:

- 1) The C–C bond responsible for decarboxylation for both KP and SUP is elongated from 1.524 Å in their ground neutral states to 1.621 and 1.624 Å in their singlet deprotonated forms, respectively, whereas in the first excited triplet state of their deprotonated species this bond is completely dissociated and measures 3.222 and 3.185 Å, respectively [103, 106]. For IBU, FBP, NP and MNAA, the corresponding bond length is elongated from 1.523, 1.524, 1.523 and 1.520 Å in their ground neutral states to 1.618, 1.607, 1.601 and 1.603 Å in their corresponding singlet deprotonated species, respectively, while in their optimized first excited triplet states the corresponding decarboxylation bond lengths are close to dissociation and measure 1.765, 1.779 1.733 and 1.675 Å respectively (Table 30.1) [104, 105, 107]. For DF, finally, there is very little change in this bond length in the neutral, singlet and triplet deprotonated species. This is

Table 30.1 C–C and C–Cl bond lengths responsible for decarboxylation and dechlorination processes in various optimized geometries of NSAIDs under study.

NSAID	Bond length responsible for decarboxylation (Å)			Reference
	Singlet neutral	Singlet deprotonated	Triplet deprotonated	
KP	1.524	1.621	3.222	[103]
IBU	1.523	1.618	1.765	[104]
FBP	1.524	1.607	1.779	[105]
SUP	1.524	1.624	3.185	[106]
NP	1.523	1.601	1.733	[107]
MNAA	1.520	1.603	1.675	[107]
DF	1.516	1.569	1.537	[108]
DF ^{a)}	1.755	1.763	2.425	[108]

a) C–Cl_b bond length.

because the deprotonated carboxyl moiety of DF takes a proton from the diphenylamino-nitrogen and thus becomes neutralized [108].

- Owing to steric repulsion between the two phenyl rings in KP, these attain a dihedral angle of 50–55° for all species explored. This induced steric torsion efficiently reduces the delocalized conjugation over the molecule and is also reflected in the elongated C–C bonds to the central carbonyl group (1.50 Å in the neutral form), compared with the C–C bond lengths of 1.39–1.41 Å in the phenyl rings [103]. For SUP, the dihedral angle of the neutral, radical anion, radical cation and deprotonated forms is –33.5°, –15.9°, –54.7° and –27.7°, respectively. Similarly to KP, the steric torsion is reflected in elongated C–C bonds to the central carbonyl group (1.42–1.50 Å in the neutral form), compared with the C–C bond lengths of 1.37–1.40 Å in the phenyl and thiophene rings [106].
- For DF, there is C–Cl_a bond dissociation in the optimized structure of the radical anion. In addition, the proton of the diphenylamino group is taken by the deprotonated carboxylic group of DF once the molecule is optimized in the deprotonated triplet state. The C–Cl_b bond length is elongated from 1.755 Å in the neutral form to 2.001 Å in the neutral triplet state, and completely dissociates in the deprotonated triplet state [108], thus suggesting a different initial step in the photodegradation.

The adiabatic electron affinity (EA) and ionization potential (IP) of each drug is calculated as the energy difference between the neutral molecule and the anion or cation at their respective relaxed geometries. A negative EA indicates that the anion is higher in energy than the corresponding neutral molecule and hence unstable, and vice versa. Table 30.2 lists the EA and the IP of different compounds. The stability order of the anions as obtained from their EA decreases in the series SUP > KP > DF > FBP > MNAA > NP > IBU. Applying bulk solvation through the IEFPCM method, the negatively charged species are stabilized and the EA of all drug molecules become positive with only small changes in the stability sequence. The molecules can be divided into three distinct groups depending on their EA. The

Table 30.2 Computed electron affinities (EAs) and ionization potentials (IPs).

NSAID	EA (kcal mol ⁻¹)		IP (kcal mol ⁻¹)		Reference
	Gas phase	Solvent phase	Gas phase	Solvent phase	
	$\Delta E_{\text{(ZPE)}}$	$\Delta\Delta G_{\text{aq}}^{298}$	$\Delta E_{\text{(ZPE)}}$	$\Delta\Delta G_{\text{aq}}^{298}$	
KP	10.30	55.70	185.20	150.60	[103]
IBU	-22.90	17.70	182.30	141.10	[104]
FBP	-3.70	34.30	173.50	135.90	[105]
SUP	14.70	59.40	185.80	149.50	[106]
NP	-8.10	33.43	163.47	124.91	[107]
MNAA	-6.96	33.77	160.59	122.75	[107]
DF	8.27	58.12	158.40	122.04	[108]

carbonyl or amino linked diphenyls KP, SUP and DF are most prone to electron uptake, the naphthyl and biphenyl moieties NP, MNAA and FBP are intermediate, and the benzylic IBU is the least easily reduced. Interestingly, the reactivities of the deprotonated species follow the same trend, as will be discussed below. The IPs, on the other hand, lies in the range 172 ± 14 kcal mol⁻¹ (1 kcal = 4.184 kJ) in the gas phase, and 136 ± 14 kcal mol⁻¹ in the solvent phase. For the IPs the trends are quite different, with DF, NP and MNAA having the lowest and KP and SUP the highest values.

30.7

NSAID Orbital Structures

As mentioned in the introduction, the 2-arylpropionic acids are weak acids having pK_a values of 3–5. Hence, all the NSAIDs are present predominantly in their deprotonated forms at physiological pH. To provide a setting for the photochemistry of each drug investigated and to determine if there is any difference between the neutral and deprotonated forms, the computed highest occupied and lowest unoccupied molecular orbitals (HOMOs and LUMOs) of the neutral and deprotonated species were analyzed for each drug. As expected, all compounds display a marked difference between their neutral and deprotonated orbitals configurations. The HOMO, HOMO-1 and HOMO-2 of the deprotonated forms are almost always localized to the carboxylic moieties, whereas the HOMO, HOMO-1 and HOMO-2 of the neutral species are localized on the phenyl or aromatic ring(s). In contrast, the LUMO, LUMO + 1 and LUMO + 2 of both the neutral and deprotonated species are found on the phenyl or aromatic ring(s), the substituents thereof or delocalized over the entire molecules. Naturally there is some variation from one species to another. The large difference between the neutral and the deprotonated species is manifested on looking at the Mulliken charge distributions on different atoms or groups of each molecule. The main negative charge of the deprotonated

Table 30.3 Mulliken charges (e^-) on the carboxylic moieties of the neutral and deprotonated species of each drug [B3LYP/6-31G(d,p) level].

NSAID	Neutral species	Deprotonated species	Reference
KP	−0.041	−0.650	[103]
IBU	−0.050	−0.675	[104]
FBP	−0.043	−0.929	[105]
SUP	−0.071	−0.640	[106]
NP	−0.049	−0.690	[107]
MNAA	−0.038	−0.679	[107]
DF	−0.025	−0.630	[108]

species is found on the carboxylic moiety whereas in the neutral form this is essentially uncharged. For instance, in the neutral form of IBU, the carboxylic moiety holds only $-0.050 e$, compared to $-0.675 e$ in the deprotonated species (Table 30.3).

On the basis of the computed orbital configurations we can conclude that: (i) once the drug molecules are deprotonated, charge redistribution takes place leading to a different orbital configuration pattern; (ii) adding an electron to the LUMO, or removing an electron from the HOMO, of the neutral species will generally lead to small structural changes in the drug molecule. For instance in KP we observe an elongation of the C=O bond and slight reduction in the C–CO bonds [103]; (iii) the different MO distribution will also have a considerable impact on the photochemical behavior of the neutral versus acidic form of each drug; (iv) using the neutral species to rationalize the energetics and photochemistry of the deprotonated form of each drug may thus lead to wrong conclusions regarding the actual mechanism involved.

Table 30.4 lists the relative ZPE-corrected energies in the gas phase and relative Gibbs free energies in aqueous solution, for the different KP, IBU, FBP, SUP, NP, MNAA and DF species investigated. We note that in aqueous phase, the anionic species $X^{\bullet-}$ is more stable than the corresponding neutral form X of each drug, whereas in vacuum the neutral form is more stable than the corresponding anionic species only in the case of IBU, FBP, NP and MNAA. Also for the IPs we note a considerable stabilization of the charged species (reduced IP) in aqueous solution relative to vacuum. The free energy differences between the neutral and the deprotonated forms of these drugs are in aqueous solution in the range $297 \pm 2.5 \text{ kcal mol}^{-1}$.

Table 30.5 shows the computed dipole moments in aqueous solution of the different forms of each drug. The dipole moments of the radical anions and radical cations change by only a few Debye relative to their corresponding neutral forms. The above-mentioned localization of charge on the carboxylic groups of the deprotonated species (cf. Table 30.3) is well reflected in the computed dipole moments. For the deprotonated species these increase by more than 16 ± 3 Debye compared with their corresponding neutral forms.

Table 30.4 B3LYP/6-31G(d,p) ZPE corrected electronic energies in the gas phase, and IEFPCM-B3LYP/6-31G(d,p) Gibbs free energies in aqueous solution for a set of NSAIDs. Relative energies in kcal mol⁻¹; X, singlet ground state neutral form; X⁻, radical anion; X^{•+} radical cation; X⁻, singlet ground state deprotonated species; ³X, first excited triplet state of neutral form, ³X⁻, first excited triplet state of deprotonated species.

System	KP [103]			IBU [104]			FBP [105]			SUP [106]			NP [107]			MNA [107]			DF [108]		
	$\Delta E_{(ZPE)}$	$\Delta\Delta G_{aq}^{298}$		$\Delta E_{(ZPE)}$	$\Delta\Delta G_{aq}^{298}$		$\Delta E_{(ZPE)}$	$\Delta\Delta G_{aq}^{298}$		$\Delta E_{(ZPE)}$	$\Delta\Delta G_{aq}^{298}$		$\Delta E_{(ZPE)}$	$\Delta\Delta G_{aq}^{298}$		$\Delta E_{(ZPE)}$	$\Delta\Delta G_{aq}^{298}$		$\Delta E_{(ZPE)}$	$\Delta\Delta G_{aq}^{298}$	
X	0.0	0.0		0.0	0.0		0.0	0.0		0.0	0.0		0.0	0.0		0.0	0.0		0.0	0.0	
X ^{•-}	-10.3	-55.7		22.9	-17.7		3.7	-34.3		-14.7	-59.4		8.10	-33.43		6.96	-33.77		-8.27	-58.12	
X ^{•+}	185.2	150.6		182.3	141.1		173.5	135.9		185.8	149.5		163.47	124.91		160.59	122.75		158.42	122.04	
X ⁻	345.9	299.2		350.9	297.1		346.4	295.1		344.1	298.2		349.43	295.45		349.20	295.93		337.10	296.75	
³ X	61.3	62.7		79.8	78.4		66.6	64.9		58.5	55.8		59.85	58.31		55.00	55.22		62.02	57.69	
³ X ⁻	376.4	332.0		419.2	374.4		400.8	361.7		376.9	333.2		400.04	355.49		397.49	351.44		380.53	335.72	

Table 30.5 Computed dipole moment (debye) of various NSAIDs in aqueous solution.

System	KP [103] μ_{aq}	IBU [104] μ_{aq}	FBP [105] μ_{aq}	SUP [106] μ_{aq}	NP [107] μ_{aq}	MNAA [107] μ_{aq}	DF [108] μ_{aq}
X	5.92	1.9	0.85	6.87	1.39	3.30	0.60
X ^{•-}	8.60	7.5	4.96	5.28	2.28	3.36	10.09
X ^{•+}	4.56	2.6	4.64	5.27	6.26	6.58	2.69
X ⁻	20.33	18.9	19.44	19.76	20.10	22.32	13.88
³ X	4.00	2.0	0.65	8.34	1.83	1.83	4.74
³ X ⁻	6.21	15.0	11.38	13.67	15.33	18.56	6.34

30.8

NSAID Absorption Spectra

Using time-dependent density functional theory, the absorption spectrum of each drug was computed. The methodology employed is well-known to render reaction energies accurate to within $\pm 2 \text{ kcal mol}^{-1}$ ($\sim 0.1 \text{ eV}$), whereas excitation energies tend to be overestimated by $3\text{--}5 \text{ kcal mol}^{-1}$ ($\sim 0.2 \text{ eV}$). Hence, this means that the computed absorption peaks will be blue-shifted relative to experiment, by approximately 10 nm at $\lambda = 250 \text{ nm}$ and by 15 nm at $\lambda = 300 \text{ nm}$. The blue-shift of current TD-DFT (time-dependent DFT) methodology has previously been investigated in great detail, as has the effect of bulk solvation and inclusion of explicit water molecules on the absorption spectra of neutral and charged species [119]. The overall conclusion is that explicit as well as implicit solvents have a very small (within a few kcal mol^{-1}) influence on the calculated spectra. Such effects are thus neglected in the current theoretical work. The computed spectra of the neutral and deprotonated forms of each NSAID were found to be in overall good agreement with experimental findings. The wavelengths of the main peaks for each drug and their corresponding data in previous experimental studies are summarized below. Figure 30.4 displays the computed absorption spectra of the neutral (Figure 30.4a) and deprotonated (Figure 30.4b) species. A general observation is that the excitations of the deprotonated species are of lower probability than the protonated forms. This may be explained by their charge-transfer (CT) nature and thus low overlap of the involved MOs, as we transfer an electron from the carboxylic moiety into the ring systems.

For the KP spectrum: (i) For the neutral form, the main absorption peak occurs at 261 nm followed by a small shoulder at 220 nm and at wavelengths shorter than 200 nm a large number of strong absorptions were also noted. At 277 and 269 nm, the lowest lying absorptions were found but with essentially negligible oscillator strengths. These findings agree with Lhiaubet *et al.* in their TD-LDA analysis [103, 120]. (ii) For the deprotonated species, the main peaks obtained are at 242, 315 and 341 nm. This matches very well with the experimental absorption spectrum with a large peak at 250–260 nm and a shoulder in the 300–350 nm region, attributed

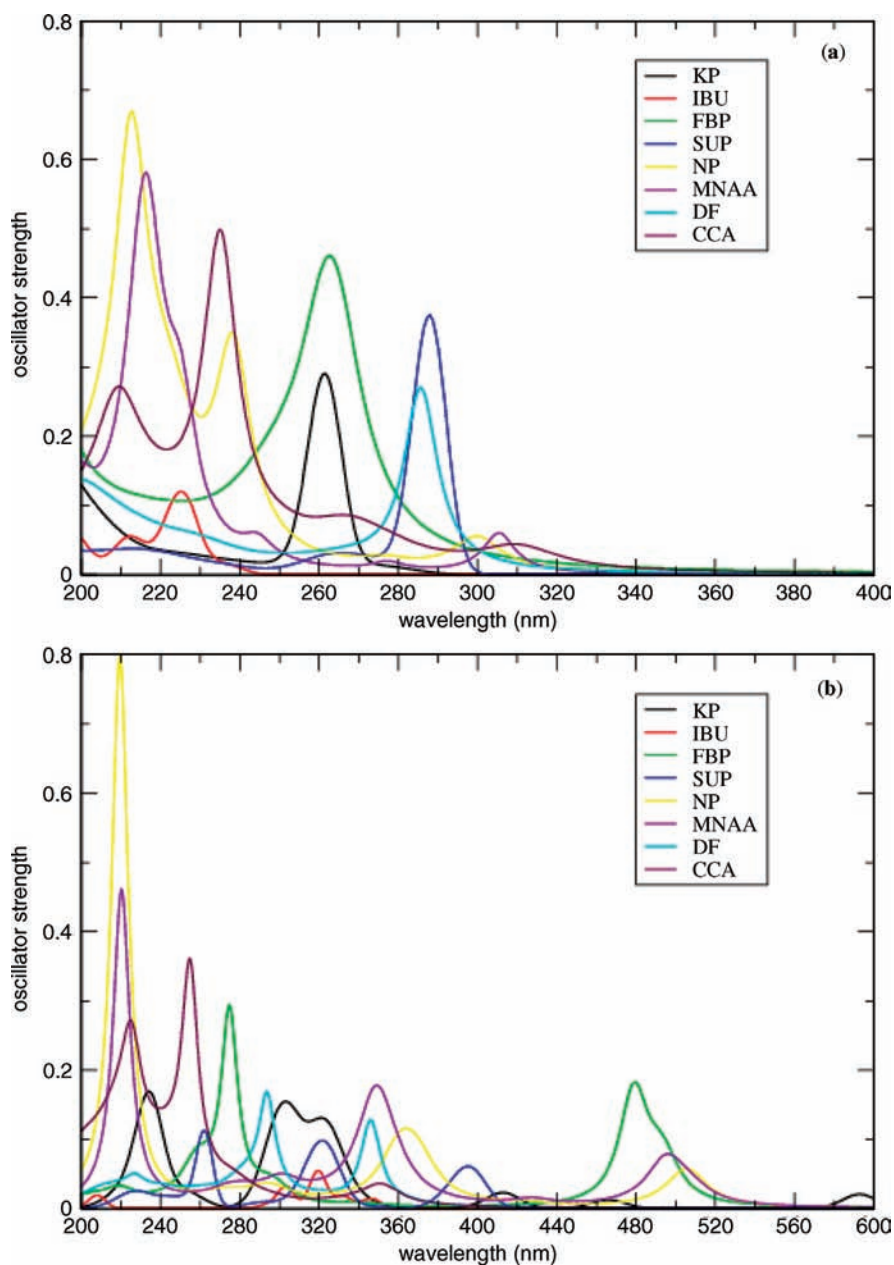


Figure 30.4 Absorption spectra of the neutral (a) and deprotonated (b) species of the set of NSAIDs investigated.

to the $S_0 \rightarrow S_2 \pi, \pi^*$ transition and the forbidden $S_0 \rightarrow S_1 n, \pi^*$ transition, respectively. The same observations were made in pure ethanol, isopentane and phosphate buffer at pH 7.4 [103, 120, 121].

In the IBU spectrum: (i) For the protonated species, the main peak is found at $\lambda = 224$ nm with non-negligible oscillator strength. Absorption peaks were also noted at shorter wavelengths, $\lambda = 211$ and 177 nm, with oscillator strengths 0.061 and 0.600, respectively, these, however are too high in energy to be photochemically relevant. (ii) For the deprotonated form, the main absorption peak is found at 319 nm. Another peak found at lower wavelength ($\lambda = 208$ nm) relates to experimental data with one absorption at approximately 220 nm [85, 104, 122]. Notably, however, the absorptions of IBU are all of very low intensity relative to the other compounds.

In the spectrum of FBP: (i) For the neutral species, the main spectral peak is at 262 nm. It has significant probability ($f = 0.436$) and matches well with the experimental data [105, 123, 124]. It is followed by an absorption at $\lambda = 247$ nm with lower probability ($f = 0.120$). (ii) For the deprotonated form, the main peak is found at $\lambda = 274$ nm ($f = 0.294$).

In terms of the SUP spectrum: (i) The protonated species shows a main peak obtained in the computed spectrum at 278 nm ($f = 0.375$), followed by a small shoulder at 262 nm and several low-intensity excitations below 250 nm. (ii) The main absorptions for the deprotonated form are at 396, 322, 318 and 262 nm, with oscillator strengths of 0.072, 0.095, 0.082 and 0.163, respectively. In contrast, the experimental data shows a broad absorption band between 360 and 250 nm, with a peak at 300 nm and a shoulder at 270 nm [106, 125].

For the NP and MNAA spectra: (i) The neutral form of NP has a main absorption peak at $\lambda = 212$ nm with significant probability ($f = 0.67$), followed by small shoulder at 238 nm that also has a relatively high oscillator strength ($f = 0.35$). Other strong absorption peaks are found at 220 nm ($f = 0.21$) and 215 nm ($f = 0.14$). (ii) For the deprotonated species, the main absorption peak is found at 220 nm ($f = 0.8$). (iii) For the neutral form of MNAA, the main absorption peak is found at 217 nm with significant oscillator strength (0.58), followed by a small shoulder at 199 nm. Other absorptions peaks are at 225 nm ($f = 0.34$) and 215 nm ($f = 0.46$). (iv) For the MNAA deprotonated form, the main absorptions occur at 495, 348, 220 and 215 nm with oscillator strengths of 0.03, 0.16, 0.46 and 0.03, respectively. The computed spectra matches well the experimental data obtained by laser flash photolysis in acetonitrile and PBS solutions, showing four bands with maxima at 220, 270, 320 and 330 nm [94, 107]. Comparing computed NP and MNAA spectra, those of the deprotonated forms are more or less similar except that the probabilities of the MNAA absorptions are roughly half those observed for the NP deprotonated species.

In the DF spectrum: (i) The protonated form shows a main absorption peak at 285 nm ($f = 0.271$). Additional peaks are noted in the computed spectrum at shorter wavelength (< 220 nm); however, they are at too high an energy to be photochemically relevant. These findings are in a good agreement with DF spectra obtained from previous experimental work [108, 126–129]. (ii) For the deprotonated form of DF, the peaks with highest computed probabilities, $f = 0.129$ and 0.168, are found at $\lambda = 346$ and 293 nm, respectively.

In the CCA (the main photoproduct of DF) spectrum: (i) the neutral species shows the first vertical excitation (HOMO $\rightarrow$ LUMO) at $\lambda = 309$ nm ($f = 0.056$). The main peaks for this form are found at $\lambda = 234$ nm ($f = 0.499$) and $\lambda = 209$ nm ($f = 0.267$). (ii) For the deprotonated species, the first excitation to S_1 is found already at $\lambda = 350$ nm ($f = 0.049$). The peaks with high probabilities are found at $\lambda = 254$ and 225 nm with $f = 0.361$ and 0.265, respectively. Compared with the parent compound DF, the spectra are shifted to shorter wavelengths (higher energy) but with higher probabilities.

30.9

Excited State Reactions

The initial step in the photodegradation of NSAIDs is the excitation of the parent compound, in its neutral or deprotonated form, radiationless decay to the first excited singlet state S_1 , followed by intersystem crossing (ISC) to the first excited triplet state. Because these drugs are weak acids with pK_a s in the range 3–5, they will predominantly be in their corresponding deprotonated forms at physiological pH. This should be taken into consideration when discussing the main routes of the excitation and photodegradation pathways. Figure 30.5 outlines the important

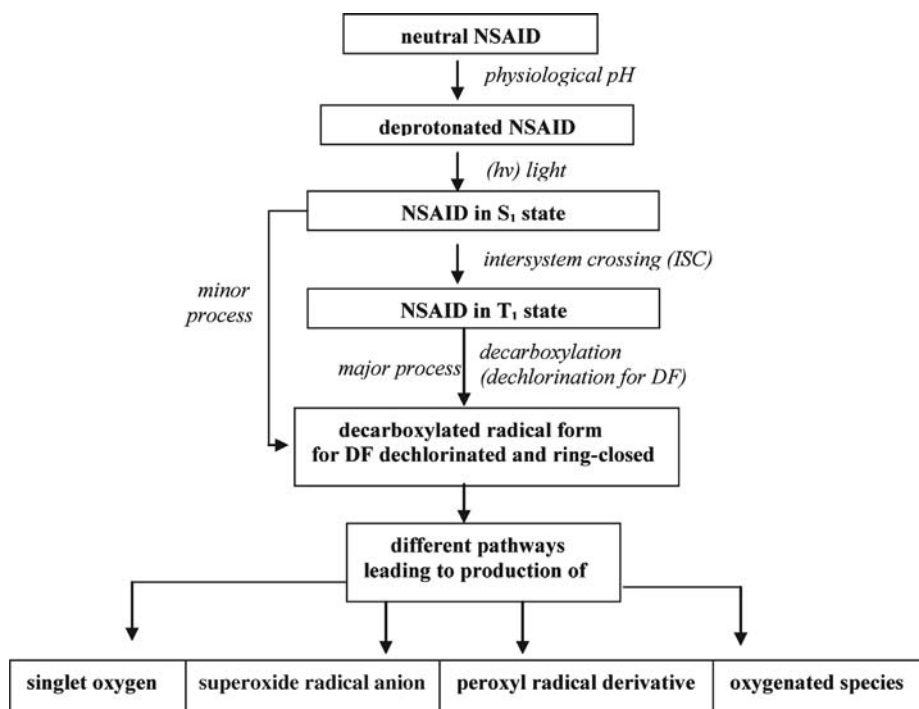


Figure 30.5 Important steps in the NSAID photodegradation mechanisms.

Table 30.6 S_1 (HOMO $\rightarrow$ LUMO) excitations in nm and kcal mol $^{-1}$, and their corresponding probabilities.

NSAID	Vertical S ₁ excitation (HOMO → LUMO)						Reference
	Neutral form			Deprotonated form			
	nm	kcal mol ^{−1}	Probability	nm	kcal mol ^{−1}	Probability	
KP	349	82	0.001	783	36.5	0.001	[103]
IBU	236	121	0.013	347	82.4	0.016	[104]
FBP	262	109	0.436	493	58	0.044	[105]
SUP	349	82	0.001	728	39.2	0.051	[106]
NP	299	95	0.050	505	56.6	0.027	[107]
MNAA	305	94	0.055	508	56.3	0.008	[107]
DF	288	99.3	0.019	364	78.5	0.007	[108]

steps common for the initial and subsequent pathways in the photodegradation reactions.

Table 30.6 displays the vertical S_1 excitation from HOMO to LUMO for both the neutral and the deprotonated species. For the neutral species, the first vertical excitation of KP, IBU, FBP, SUP, NP, MNAA and DF occur at 349, 236, 262, 349, 299, 305 and 288 nm, in the UV-regime of the spectrum, and with oscillator strengths of 0.001, 0.013, 0.436, 0.001, 0.050, 0.055 and 0.019, respectively. These are hence very small oscillator strengths (probabilities) except for FBP, indicating that they will not constitute the main absorption bands.

For the deprotonated species of the above-mentioned drugs, the lowest lying excitations are again from the highest occupied MOs to the LUMO, and occur at 783, 347, 493, 728, 505, 508, and 364 nm, again with very small oscillator strengths, 0.001, 0.016, 0.044, 0.051, 0.027, 0.008 and 0.007, respectively, listed in the same order as the neutral compounds. Owing to the poor overlap between the highest occupied MOs (localized on the carboxylate group) and the lowest unoccupied ones localized on the conjugated rings, the probabilities for these lowest-lying excitations occurring are too small to play any role in the photochemistry of the species.

The singlet excited system, in either the neutral or deprotonated form, upon ISC leads to formation of the triplet state. The optimized lowest-lying triplet of the deprotonated form of KP, IBU, FBP, SUP, NP, MNAA and DF lie 32.9, 77.4, 66.6, 35.1, 60, 55.5 and 39 kcal mol $^{-1}$ above their corresponding optimized singlet ground states (X^-) respectively (Table 30.7); for the protonated species, the corresponding free energy difference is 62.7, 78.4, 64.9, 55.8, 58.3, 55.2 and 57.7 kcal mol $^{-1}$ respectively. The values are very little affected by the inclusion of bulk solvation. The energies of the optimized neutral and deprotonated triplets agree well with the vertical T_1 energies obtained from the TD-DFT calculations of the neutral and deprotonated species (67.6, 84.8, 73, 64, 64, 60.4 and 75.8; and 40.8, 73.8, 52, 31, 52, 50 and 67 kcal mol $^{-1}$, respectively), indicating that the structural relaxation stabilizes the systems by 5–18 kcal mol $^{-1}$ (cf. Table 30.7).

Table 30.7 Singlet–triplet gap of the optimized neutral and deprotonated NSAIDs in gas and solvent phase as well as vertical gaps obtained from TD-DFT calculations.

NSAID	S ₀ -T ₁ gap (kcal mol ⁻¹)						Reference
	Neutral			Deprotonated			
	Gas phase	Solvent phase	TD-DFT (vacuum)	Gas phase	Solvent phase	TD-DFT (vacuum)	
KP	61.30	62.7	67.6	30.52	32.88	40.8	[103]
IBU	79.80	78.4	84.8	68.29	77.35	73.8	[104]
FBP	66.6	64.9	73	54.40	66.64	52	[105]
SUP	58.5	55.8	64	32.85	35.05	31	[106]
NP	59.85	58.31	64	50.61	60.04	52	[107]
MNAA	55.00	55.22	60.4	48.29	55.51	50	[107]
DF	62.02	57.69	75.8	43.43	38.97	67	[108]

All substances explored are, hence, provided the T_1 state is sufficiently stable to not undergo rapid decomposition, capable of generating singlet oxygen under aerobic conditions [$\Delta E (^1O_2) \sim 23$ kcal mol⁻¹].

30.9.1

Photodegradation from the T_1 State

The critical step to initiate the various photodegradation mechanisms for these NSAIDs is decarboxylation (dechlorination in case of DF) from the corresponding first excited triplet states of the deprotonated species. The decarboxylation process from the deprotonated forms of KP and SUP occur spontaneously and without any energy barriers once optimized in their deprotonated triplet states. For IBU, FBP, NP and MNAA, as mentioned in Table 30.1, there is an elongation of the C–C bond lengths responsible for the decarboxylation from these species. On scanning this bond from the corresponding optimized bond length (Table 30.1), in steps of 0.1 Å until that the bond is completely dissociated, the curves displayed in Figure 30.6 were obtained. They indicate that the decarboxylation processes from the deprotonated triplet states of IBU, FBP, NP and MNAA need to pass low energy barriers of 0.3, 0.4, 0.9 and 2.8 kcal mol⁻¹, respectively. These barriers occur at transition state distances of approximately 1.97, 2.08, 2.13 and 2.25 Å, respectively.

In contrast, DF is more likely to dechlorinate instead of decarboxylate. Once this drug is optimized in its deprotonated triplet state, dechlorination takes place spontaneously without any energy barrier. The net result of this step is the formation of HCl, whereas the remaining molecule will form the main photoproduct in the DF photodegradation mechanism, 8-chlorocarbazole acetic acid (CCA). CCA can in turn be decarboxylated from its deprotonated triplet state by passing an energy barrier of 4.3 kcal mol⁻¹ at a transition state distance of 2.37 Å (Figure 30.6).

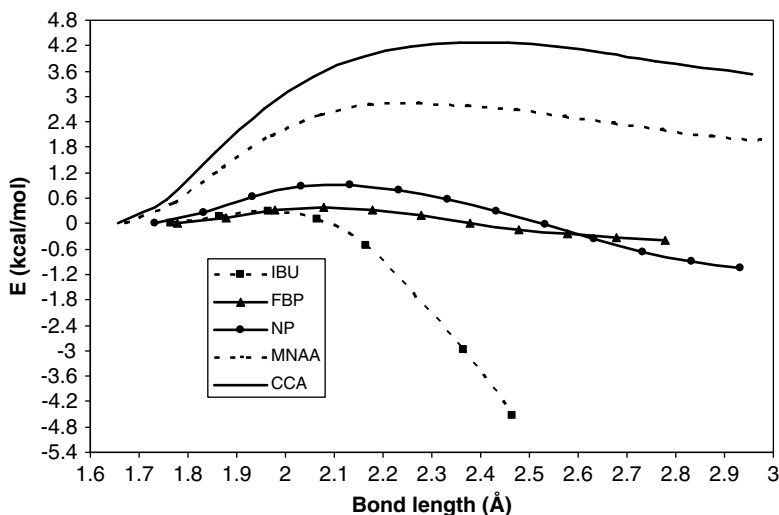


Figure 30.6 Decarboxylation energy curves for IBU, FBP, NP, MNAA and CCA deprotonated triplet states.

Based on these results, the NSAID decarboxylation process from the deprotonated triplet states species can be ranked in order of increasing photostability as $KP = SUP < IBU = FBP < NP < MNAA < CCA$.

30.9.2

Possible Photodegradation from Singlet Excited States

As mentioned above, the photodegradation mechanisms of the NSAIDs occur mainly through decarboxylation from the first excited triplet state of the deprotonated species. This process may, however, also occur from an initial excited singlet state of the same species. To investigate if the excited singlet states of each drug may also spontaneously decarboxylate (or needs to overcome an energy barrier), the C–C bond responsible for the decarboxylation was scanned outward from the optimized value of the S_0 state of the deprotonated form of KP, IBU, FBP, SUP, NP and MNAA, in steps of 0.1 Å. At each new point, the structures were reoptimized and the vertical excitations calculated. The resulting energy curves including the ground state and a set of the lowest singlet excitations of each drug were thus obtained at the TD-B3LYP/6-31 + G(d,p) level. From the resulting energy curves we conclude the following:

- 1) For KP, the S_1 , S_3 and S_5 states appear to be strictly endothermic, while the S_6 – S_9 excitations show exothermic decarboxylation reactions proceeding by way of a small energy barrier, of the order of 2–3 kcal mol^{−1}, at C–C distances of 1.8–2.1 Å. Hence, several of the excited singlet states of the deprotonated species of KP can be expected to result in decarboxylation. Nonetheless, the behavior is

quite different from the first excited triplet state where the decarboxylation occurs spontaneously during geometry optimization [103].

- 2) For IBU, the ground state and most of the lowest excited singlet state surfaces (except S_4 and S_7) are strictly endothermic, and hence show no sign of decarboxylation. The S_4 state displays an apparent transition barrier with a maximum at a C—C distance of ~ 2.4 Å and a barrier to decarboxylation that is rather high (around 20 kcal mol^{-1}). The S_7 state, however, lies far too high in energy to be of relevance: $115 \text{ kcal mol}^{-1}$ above the ground state. Thus, we can conclude that decarboxylation is not likely to occur from the excited singlet states of IBU [104].
- 3) For FBP, similar to IBU, the ground and most of the lowest excited singlet state surfaces are strictly endothermic, and hence show no sign of decarboxylation. The exception is state S_4 (and possibly S_7) which displays a transition barrier with a maximum at a C—C distance of ~ 2.2 Å. Again the barrier to decarboxylation is high, around 21 kcal mol^{-1} . This indicates that decarboxylation is not likely to occur from the excited singlet states of FBP [105], which is in excellent agreement with experimental data that show that the quantum yield for this process is less than 0.01 in PBS and organic solvents [130].
- 4) For SUP, the six lowest singlet excited states are either strictly endothermic or involve energy barriers above 20 kcal mol^{-1} , similar to the case of FBP and IBU. Interestingly, from these calculations, the state that appears most benevolent towards decarboxylation is the S_0 ground state [106].
- 5) As expected, the ground and most of the lowest excited singlet state surfaces of NP and MNAA are strictly endothermic. The S_4 state is also herein an exception, although it lies far too high in energy to be relevant. The transition barrier of that state has a maximum at a C—C distance of ~ 2.4 Å. The barrier is rather high (around 24 kcal mol^{-1}) and hence neither NP nor MNAA will show signs of singlet decarboxylation. This situation is thus again similar to the case IBU [107].

Notably, although the neutral species of these drugs are not present at physiological pH they will show no sign of decarboxylation from their excited singlet states. For instance, our study of KP shows the ground state and three lowest excited singlet states are strictly endothermic throughout the scan of the C—C bond responsible for the decarboxylation [103]. In general, initiation of the photodecarboxylation in various NSAID photodegradation mechanisms is energetically preferred to occur from the first excited triplet state of the deprotonated species but not (or with very low probability) from their corresponding excited singlet states.

30.10

Reactive Oxygen Species (ROS) and Radical Formation

Once the first excited triplet state is reached, and provided this does not spontaneously decarboxylate, energy or electron transfer to molecular oxygen will yield a highly reactive oxygen species (ROS) – singlet oxygen or superoxide radical anion, respectively. In addition, the radicals formed through the NSAID photodegradation

pathways may react with molecular oxygen, forming a peroxy radical derivative, as seen in the general chart of photodegradation mechanism of NSAID (Figure 30.5). These ROS react with a wide range of biologic targets (e.g., lipids or DNA) and are known to be involved in both cellular signaling and cell damage, which cumulates into a situation known as oxidative stress.

In the different photodegradation mechanisms of this set of NSAIDs, the relatively high energy of the first excited triplet states as well as the energies of some derivatives formed throughout the photodegradation mechanisms will, by triplet-triplet interaction, transfer their energy to molecular oxygen and thus lead to singlet oxygen formation. The excitation energy of ground state molecular oxygen leading to the formation of singlet oxygen is only $22.5 \text{ kcal mol}^{-1}$ [131]. Once the singlet oxygen is formed it will react rapidly with various electron-rich substrates, including unsaturated fatty acids in lipid membranes, cholesterol, amino acid residues of proteins (particularly cysteine, histidine and tryptophan), and nucleic acid bases of DNA such as guanine and thymine [132–134].

The second reactive oxygen species formed during the photodegradation mechanisms is the superoxide radical anion. The adiabatic electron affinity of molecular oxygen to generate superoxide in solution is estimated to be $90.2 \text{ kcal mol}^{-1}$ (3.91 eV). The $\text{O}_2^{\bullet-}$ undergoes fast bimolecular decay to yield oxidizing species such as H_2O_2 , O_2 and $\bullet\text{OH}$ in near-neutral and acidic solutions. The superoxide anion in itself may also function as a reducing agent for some triplet derivatives [135]. In general, polar media increase the production of superoxide over singlet oxygen due to the large solvent-induced stabilization (increased electron affinity) of the superoxide radical anion [102].

Peroxy radical derivatives are formed as a result of the reaction of the decarboxylated triplet anion with two molecules of oxygen. The first molecule oxidizes the decarboxylated moiety, leaving a doublet centered at the site of decarboxylation plus superoxide. Once formed, the radical site will react with the second O_2 molecule. The reaction energy forming the peroxy radical product is estimated in different NSAID phototoxic mechanisms (such as KP, IBU, FBP and SUP) to be close to $20 \pm 1 \text{ kcal mol}^{-1}$. The generation of the peroxy radical is strictly exothermic under aerobic conditions, and occurs spontaneously and without any barrier.

The different radical species mentioned above (Type I reaction) were noted in both the experimental and theoretical studies of NSAID photodegradation mechanisms. The decarboxylated derivatives have a very reactive site with significant spin density at the carbon atom that formerly connected to the carboxylic group, making it a suitable site for attack by other molecules. This scenario will result in undesired reactions with biomolecules, as discussed in more detail in the following section.

30.11

Effects of the Formed ROS and Radicals during the Photodegradation Mechanisms

ROS have the capability to react with different biomolecules inside the body, affecting DNA, proteins and lipids. This can lead to severe unwanted side effects such as

photosensitivity reactions, depending on the amount of the compound, the level of activating radiation and the quantity of other chromophores in the skin. Hemolysis and mutagenic effects on DNA are also caused by ROS. In this context, singlet oxygen plays a major role in developing various undesired side effects by reacting with different biomolecules such as unsaturated fatty acids, amino acids in protein structures, and DNA nucleobases:

- 1) ROS can react with the bases of DNA, in particularly reactions with pyrimidine bases have been reported in various studies. Phototoxic NSAIDs may induce DNA damage *in vitro* upon irradiation; for example, KP and NP have been investigated for their ability to generate pyrimidine dimers and single-strand breaks. The results show that UV-irradiation of DNA alone causes pyrimidine dimers while single strand breaks were not detected. However, once DNA was irradiated in the presence of KP, single strand breaks were found, along with increased formation of dimers. The DNA cleavage quantum yield was estimated to be 5×10^{-4} in the case of KP. In deaerated solution, KP-photoinduced dimerization of pyrimidines increased and strand cleavage decreased, while in aerated solution the opposite was the case. Based on these finding, a competition has been postulated, between a less efficient energy transfer between the drug and the pyrimidines at the origin of the dimerization process and a radical process leading to DNA cleavage [86, 87].
- 2) NSAIDs such as SUP are frequently associated with a high incidence of phototoxic or photoallergic reactions; in many studies *in vitro* using proteins or whole cells, it has been demonstrated that the NSAID may provoke modifications in proteins and other cell constituents after irradiation [136]. Exploring these reactions in the presence and absence of light is important in understanding how the light affects the mode, site and mechanism of association with the protein. The photobinding to proteins is proposed to occur by two postulated pathways: (a) through association by weak van der Waals or hydrogen bonding interactions the drug binds to the protein and subsequently photobinds upon UV irradiation or (b) in bulk solution, the drug first decomposes to form less polar photoproduct(s) that become more strongly associated to the protein and bind covalently once irradiation takes place [137].
- 3) The addition of molecular oxygen to the radical derivative formed during the NSAID photodegradation (R) to generate the peroxy radical derivative, $R(OO\bullet)$, has been postulated to not only give the various oxygenated derivatives but also to initiate lipid peroxidation reactions. The peroxy radical derivative $R(OO\bullet)$ will in this scenario abstract a hydrogen atom from a lipid molecule (L) that through addition of molecular oxygen to the new radical site ($L\bullet$) creates the propagating radical damage [(30.1)–(30.4)]:





Once initiated, the chain reactions (30.3) and (30.4) will then repeat until terminated by, for example, radical–radical addition or the action of a lipid-soluble antioxidant such as vitamin E.

30.12

Conclusions

- 1) KP, IBU, FBP, SUP, NP, MNAA and DF studies by means of quantum computational density functional theory (DFT) and time-dependent DFT (TD-DFT) provide a detailed molecular basis on which to understand their photodegradation mechanisms, and hence play an important role in preventing or at least reducing the phototoxic side effects by enabling the development of safer drugs in this area.
- 2) The molecular orbital configurations show that for each drug there is a marked difference between the neutral and deprotonated species. Hence, using the neutral species to rationalize the energetics and photochemistry of each drug may lead to wrong conclusions regarding the actual mechanism involved.
- 3) The main absorption peaks of the neutral form of each drug are located at shorter wavelengths (higher energy), mainly UV-C (200–280 nm) or UV-B (280–320 nm), whereas the absorbance of the deprotonated species is shifted towards longer wavelengths, mainly in the UV-A (320–400 nm) and visible regions, albeit with relatively lower probabilities. The computed spectra are in line with previous experimental findings.
- 4) The critical step in NSAID photodegradation is decarboxylation (or dechlorination in the case of DF). The reactions will primarily take place from the first excited triplet state. In the case of KP and SUP, the T_1 states spontaneously decarboxylate in their deprotonated forms while the other NSAIDs need to pass low energy barriers (in all cases $<3 \text{ kcal mol}^{-1}$). DF is similar to KP and SUP in that it undergoes spontaneous decomposition from the T_1 state, although in this case in the form of dechlorination followed by ring closure. Its main photo-product, 8-chlorocarbazole acetic acid (CCA), is in turn decarboxylated from T_1 by passing an energy barrier of less than $4.5 \text{ kcal mol}^{-1}$. The possibility of decomposition from the corresponding excited singlet states is less and requires the passage over relatively high energy barriers (around 20 kcal mol^{-1}) in most cases. The one exception is KP, where the barrier for the excited singlet states (S_6 – S_9) is in the range 2 – 3 kcal mol^{-1} .
- 5) Reactive oxygen species such as singlet oxygen, superoxide radical anions and peroxy radical derivatives, as well as other radical species, are formed in different steps along the photodegradation mechanisms of each drug from the above-mentioned NSAIDs. These highly reactive species have the capability to affect biomolecules by, for example, initiating a propagating lipid peroxidation or reacting with DNA or proteins.

Acknowledgments

The MENA program (KAKM), the Swedish Science Research Council and the Faculty of Science and Technology (LAE) are gratefully acknowledged for financial support. We also acknowledge generous grants of computing time at the National Supercomputing Center (NSC) in Linköping.

References

- 1 Tilstone, C., Report by the Medicines and Healthcare products Regulatory Agency (MHRA), Commission on Human Medicines (CHM), Drug safety update, August 2008, **2**, 1–10.
- 2 Bednar, B. (2009) *Nephrol New Issues*, **23**, 38–40.
- 3 Yusuff, K. B. and Yusuf, A. (2009) *Journal of the American Pharmacists Association*, **49**, 432–435.
- 4 The Pharmacovigilance Unit; Irish Medicines Board (2000) *Drug Safety Newsletter*, 9th edn., 1.
- 5 Bond, C.A. and Raehl, C.L. (2006) *Pharmacotherapy*, **26**, 601.
- 6 Chan, F.K.L. (2006) *Nat. Clin. Pract. Gastroenterol. Hepatol.*, **3**, 563.
- 7 Aldenkamp, A.P. (2004) *Epilepsia*, **45**, 47.
- 8 Gronroos, M., Chen, M., Jahnukainen, T., Capitanio, A. *et al.* (2006) *Pediatr. Blood & Cancer*, **46**, 624.
- 9 Vargas, F., Rivas, C., Diaz, Y., and Fuentes, A. (2003) *J. Photochem. Photobiol. B*, **72**, 87.
- 10 Kao, Y.H., Chong, C.H., Ng, W.T., and Lim, D. (2007) *Occupational Med. - Oxford*, **57**, 535.
- 11 Musshoff, F., Gerschlauser, A., and Madea, B. (2003) *Forensic Sci. Int.*, **134**, 234.
- 12 Werner, J.J., McNeill, K., and Arnold, W.A. (2005) *Chemosphere*, **58**, 1339.
- 13 Uwai, K., Tani, M., Ohtake, Y., Abe, S. *et al.* (2005) *Life Sci.*, **78**, 357.
- 14 Di Paola, A., Addamo, M., Augugliaro, V., Garcia-Lopez, E., Loddo, V., Marci, G., and Palmisano, L. (2006) *Int. J. Photoenergy*, **2006**, 6.
- 15 Han, K.D., Bark, K.M., Heo, E.P., Lee, J.K. *et al.* (2000) *Photodermatol. Photoimmunol. Photomed.*, **16**, 121.
- 16 Thoma, K. and Holzmann, C. (1998) *Eur. J. Pharm. Biopharm.*, **46**, 201.
- 17 Andrisano, V., Gotti, R., Leoni, A., and Cavrini, V. (1999) *J. Pharm. Biomed. Anal.*, **21**, 851.
- 18 Brisaert, M. and Plaizier-Vercammen, J. (2000) *Int. J. Pharm.*, **199**, 49.
- 19 Andrisano, V., Ballardini, R., Hrelia, P., Cameli, N. *et al.* (2001) *Eur. J. Pharm. Sci.*, **12**, 495.
- 20 Andrisano, V., Hrelia, P., Gotti, R., Leoni, A., and Cavrini, V. (2001) *J. Pharm. Biomed. Anal.*, **25**, 589.
- 21 Stott, C.W., Stasse, J., Bonomo, R., and Campbell, A.H. (1970) *J. Invest. Dermatol.*, **55**, 335.
- 22 Spielmann, H., Lovell, W.W., Holzle, E., Johnson, B.E. *et al.* (1994) *Alternatives Lab. Animals*, **22**, 314.
- 23 DeRosa, M.C., Crutchley, R.J., (2002) *Coordination Chemistry Reviews* **233–234**, 351–71.
- 24 Doll, T.E. and Frimmel, F.H. (2003) *Chemosphere*, **52**, 1757.
- 25 Ray, R.S., Misra, R.B., Farooq, M., and Hans, R.K. (2002) *Toxicol. In Vitro*, **16**, 123.
- 26 Quintero, B. and Miranda, M.A. (2000) *Ars Pharm.*, **41**, 27.
- 27 Mroz, P., Pawlak, A., Satti, M., Lee, H., Wharton, T., Gali, H., Sarna, T., and Hamblin, M.R. (2007) Functionalized fullerenes mediate photodynamic killing of cancer cells: Type I versus Type II photochemical mechanism. *Free Radical Biology and Medicine*, **43**, 711–719.
- 28 Bosca, F. and Miranda, M.A. (1998) *J. Photochem. Photobiol. B*, **43**, 1.
- 29 Miranda, M.A. (2001) *Pure App. Chem.*, **73**, 481.
- 30 Reid, C.D., Chemical Photosensitivity Another Reason to Be Careful in the Sun. Published by the U.S. Food and Drug Administration. *FDA Consumer magazine*

- (May 1996). http://sun1.awardspace.com/Causes_Photosensitivity/chemical_photosensitivity.htm.
- 31 Zambre, A.P., Ganure, A.L., Shinde, D.B., and Kulkarni, V.M. (2007) *J. Chem. Inf. Mod.*, **47**, 635.
 - 32 Penning, T.D., Talley, J.J., Bertenshaw, S.R., Carter, J.S. *et al.* (1997) *J. Med. Chem.*, **40**, 1347.
 - 33 Prasit, P., Wang, Z., Brideau, C., Chan, C.C. *et al.* (1999) *Bioorg. Med. Chem. Lett.*, **9**, 1773.
 - 34 Talley, J.J., Brown, D.L., Carter, J.S., Graneto, M.J. *et al.* (2000) *J. Med. Chem.*, **43**, 775.
 - 35 Talley, J.J., Bertenshaw, S.R., Brown, D.L., Carter, J.S. *et al.* (2000) *J. Med. Chem.*, **43**, 1661.
 - 36 Frolich, J.C. (1997) *Trends Pharmacol. Sci.*, **18**, 30.
 - 37 Frolich, J.C. (1997) *Trends Pharmacol. Sci.*, **18**, 312.
 - 38 Jose, V.M. and Antony, T.T. (2003) *Indian J. Pharmacol.*, **35**, 318.
 - 39 Cardenas, S., Gallego, M., Valcarcel, M., Ventura, R., and Segura, J. (1996) *Anal. Chem.*, **68**, 118.
 - 40 Meloun, M., Bordovska, S., and Galla, L. (2007) *J. Pharm. Biomed. Anal.*, **45**, 552.
 - 41 Smith, W.L., DeWitt, D.L., and Garavito, R.M. (2000) *Annu. Rev. Biochem.*, **69**, 145.
 - 42 Herschman, H.R., Xie, W.L., and Reddy, S. (1999) *Eicosanoids and other Bioactive Lipids in Cancer, Inflammation and Radiation Injury*, **469**, 3.
 - 43 Griswold, D.E. and Adams, J.L. (1996) *Med. Res. Rev.*, **16**, 181.
 - 44 Liu, H., Huang, X.Q., Shen, J.H., Luo, X.M. *et al.* (2002) *J. Med. Chem.*, **45**, 4816.
 - 45 Marnett, L.J. and Kalgutkar, A.S. (1999) *Trends Pharm. Sci.*, **20**, 465.
 - 46 Allison, M.C., Howatson, A.G., Torrance, C.J., Lee, F.D., and Russell, R.I. (1992) *New England J. Med.*, **327**, 749.
 - 47 Lucio, M., Bringezu, F., Reis, S., Lima, J., and Brezesinski, G. (2008) *Langmuir*, **24**, 4132.
 - 48 Price, M.L.P. and Jorgensen, W.L. (2000) *J. Am. Chem. Soc.*, **122**, 9455.
 - 49 Shao, J.Y., Sheng, H.M., Aramandla, R., Pereira, M.A. *et al.* (1999) *Carcinogen.*, **20**, 185.
 - 50 Williams, C.S., Luongo, C., Radhika, A., Zhang, T. *et al.* (1996) *Gastroenterology*, **111**, 1134.
 - 51 Eberhart, C.E., Coffey, R.J., Radhika, A., Giardiello, F.M. *et al.* (1994) *Gastroenterology*, **107**, 1183.
 - 52 Murphy, J.F., (2008) *Clinical Medicine: Cardiology*, **2**, 257–262.
 - 53 Llorens, O., Perez, J.J., Palomer, A., and Mauleon, D. (2002) *J. Mol. Graphics Mod.*, **20**, 359.
 - 54 Llorens, O., Perez, J.J., Palomer, A., and Mauleon, D. (1999) *Bioorg. Med. Chem. Lett.*, **9**, 2779.
 - 55 Xie, W.L., Robertson, D.L., and Simmons, D.L. (1992) *Drug Develop. Res.*, **25**, 249.
 - 56 Breitner, J.C.S. (2003) *Lancet Neurol.*, **2**, 527.
 - 57 Etminan, M. (2003) *Br. Med. J.*, **327**, 751.
 - 58 Price, D. (2003) *Br. Med. J.*, **327**, 752.
 - 59 Robertson, M. (2003) *Br. Med. J.*, **327**, 751.
 - 60 Aisen, P.S., Schafer, K.A., Grundman, M., Pfeiffer, E. *et al.* (2003) *J. Am. Med. Assoc.*, **289**, 2819.
 - 61 Trujillo-Murillo, K., Rincon-Sanchez, A.R., Martinez-Rodriguez, H., Bosques-Padilla F. *et al.* (2008) *Hepatology (Philadelphia)*, **47**, 1462.
 - 62 Marjanovic, M., Zorc, B., Pejnovic, L., Zovko, M., and Kralj, M. (2007) *Chem. Biol. Drug Des.*, **69**, 222.
 - 63 Yamazaki, T., Muramoto, M., Oe, T., Morikawa, N. *et al.* (2006) *Neuropharmacology*, **50**, 558–567.
 - 64 Munroe, D.G. and Lau, C.Y. (1995) *Chem. Biol.*, **2**, 343.
 - 65 Herschman, H.R. (1996) *Biochim. Biophys. Acta - Lipids Lipid Metabolism*, **1299**, 125.
 - 66 Langford, R.M. (2006) *Clin. Rheumatol.*, **25**, S2.
 - 67 Jones, R., Rubin, G., Berenbaum, F., and Scheiman, J. (2008) *Am. J. Med.*, **121**, 464.
 - 68 Hasan, J., Beharry, K.D., Gharraee, Z., Stavitsky, Y. *et al.* (2008) *Prostaglandins Lipid Mediators*, **85**, 81.
 - 69 Chemtob, S., Laudignon, N., Beharry, K., Rex, J. *et al.* (1990) *Develop. Pharmacol. Therapeut.*, **14**, 1.
 - 70 House, A.A., Oliveira, S.S., and Ronco, C. (2007) *Int. J. Artificial Organs*, **30**, 1042.
 - 71 Hinz, B., Renner, B., and Brune, K. (2007) *Nat. Clin. Pract. Rheumatol.*, **3**, 552.

- 72 Marret, E. and Bonnet, F. (2007) *Ann. Fr. d'Anesthésie Réanimation*, **26**, 535.
- 73 Suleyman, H., Demircan, B., and Karagoz, Y. (2007) *Pharmacol. Rep.*, **59**, 247.
- 74 Yorgason, J.G., Fayad, J.N., and Kalinec, F. (2006) *Expert Opin. Drug Safety*, **5**, 383.
- 75 Diaz, R.L., Gardeazabal, J., Manrique, P., Raton, J.A. *et al.* (2006) *Contact Dermatitis*, **54**, 239.
- 76 Bagheri, H., Lhiaubet, V., Montastruc, J.L., and Chouini-Lalanne, N. (2000) *Drug Safety*, **22**, 339.
- 77 Sugiura, M., Hayakawa, R., Xie, Z.L., Sugiura, K. *et al.* (2002) *Photodermatol. Photoimmunol. Photomed.*, **18**, 82.
- 78 Nakajima, A., Tahara, M., Yoshimura, Y., and Nakazawa, H. (2005) *J. Photochem. Photobiol. A*, **174**, 89.
- 79 Nakazawa, T., Shimo, T., Chikamatsu, N., Igarashi, T. *et al.* (2006) *Arch. Toxicol.*, **80**, 442.
- 80 Liu, S., Mizu, H., and Yamauchi, H. (2007) *Biochem. Biophys. Res. Commun.*, **364**, 650.
- 81 Castell, J.V., Gomez, M.J., Miranda, M.A., and Morera, I.M. (1987) *Photochem. Photobiol.*, **46**, 991.
- 82 Miranda, M.A., Morera, I., Vargas, F., Gomezlechón, M.J., and Castell, J.V. (1991) *Toxicol. In Vitro*, **5**, 451.
- 83 Bergner, T. and Przybilla, B. (1992) *J. Am. Acad. Dermatol.*, **26**, 114.
- 84 Miranda, M.A., Castell, J.V., Gomezlechón, M.J., and Martínez, L.A. (1993) *Toxicol. In Vitro*, **7**, 523.
- 85 Packer, J.L., Werner, J.J., Latch, D.E., McNeill, K., and Arnold, W.A. (2003) *Aquatic Sci.*, **65**, 342.
- 86 Chouini-Lalanne, N., Defais, M., and Paillous, N. (1998) *Biochem. Pharmacol.*, **55**, 441.
- 87 Artuso, T., Bernadou, J., Meunier, B., and Paillous, N. (1990) *Biochem. Pharmacol.*, **39**, 407.
- 88 Castell, J.V., Gomezlechón, M.J., Grassa, C., Martínez, L.A. *et al.* (1993) *Photochem. Photobiol.*, **57**, 486.
- 89 Jimenez, M.C., Miranda, M.A., and Tormos, R. (1997) *J. Photochem. Photobiol. A*, **104**, 119.
- 90 Partyka, M., Au, B.H., and Evans, C.H. (2001) *J. Photochem. Photobiol. A*, **140**, 67.
- 91 Martínez, L.J. and Scaiano, J.C. (1998) *Photochem. Photobiol.*, **68**, 646.
- 92 Canudas, N., Moulinier, J., Zamora, D., and Sanchez, A. (2000) *Pharmazie*, **55**, 282.
- 93 Canudas, N., Zamora, D., Villamizar, J.E., Fuentes, J. *et al.* (2005) *Pharmazie*, **60**, 604.
- 94 Bosca, F., Canudas, N., Marin, M.L., and Miranda, M.A. (2000) *Photochem. Photobiol.*, **71**, 173.
- 95 Castell, J.V., Gomezlechón, M.J., Grassa, C., Martínez, L.A. *et al.* (1994) *Photochem. Photobiol.*, **59**, 35.
- 96 Condorelli, G., Costanzo, L.L., Deguidi, G., Giuffrida, S., and Sortino, S. (1995) *Photochem. Photobiol.*, **62**, 155.
- 97 Starrs, S.M. and Davies, R.J.H. (2000) *Photochem. Photobiol.*, **72**, 291.
- 98 Castell, J.V., Gomezlechón, M.J., Miranda, M.A., and Morera, I.M. (1992) *J. Photochem. Photobiol. B*, **13**, 71.
- 99 Cuk, A., Skerlj, M., Palka, E., and Murn, M. (1991) *Z. Vestnik*, **60**, 267.
- 100 Encinas, S., Bosca, F., and Miranda, M.A. (1998) *Chem. Res. Toxicol.*, **11**, 946.
- 101 Encinas, S., Bosca, F., and Miranda, M.A. (1998) *Photochem. Photobiol.*, **68**, 640.
- 102 Moore, D.E. (1998) *Mutat. Res.-Fundamental Mol. Mechanisms Mutagen.*, **422**, 165.
- 103 Musa, K.A.K., Matxain, J.M., and Eriksson, L.A. (2007) *J. Med. Chem.*, **50**, 1735.
- 104 Musa, K.A.K. and Eriksson, L.A. (2007) *J. Phys. Chem. B*, **111**, 13345.
- 105 Musa, K.A.K. and Eriksson, L.A. (2008) *J. Photochem. Photobiol. A*, **1**, 200, 48.
- 106 Musa, K.A.K. and Eriksson, L.A. (2009) *J. Phys. Chem. B*, **113**, 11306.
- 107 Musa, K.A.K. and Eriksson, L.A. (2008) *J. Phys. Chem. A*, **112** (43) 10921.
- 108 Musa, K.A.K. and Eriksson, L.A. (2008) *Phys. Chem. Chem. Phys.*, **11**, 4601.
- 109 Becke, A.D. (1993) *J. Chem. Phys.*, **98**, 5648.
- 110 Lee, C.T., Yang, W.T., and Parr, R.G. (1988) *Phys. Rev. B*, **37**, 785.
- 111 Stephens, P.J., Devlin, F.J., Chabalowski, C.F., and Frisch, M.J. (1994) *J. Phys. Chem.*, **98**, 11623.
- 112 Cossi, M., Scalmani, G., Rega, N., and Barone, V. (2002) *J. Chem. Phys.*, **117**, 43.

- 113 Mennucci, B. and Tomasi, J. (1997) *J. Chem. Phys.*, **106**, 5151.
- 114 Cances, E., Mennucci, B., and Tomasi, J. (1997) *J. Chem. Phys.*, **107**, 3032.
- 115 Bauernschmitt, R. and Ahlrichs, R. (1996) *Chem. Phys. Lett.*, **256**, 454.
- 116 Casida, M.E., Jamorski, C., Casida, K.C., and Salahub, D.R. (1998) *J. Chem. Phys.*, **108**, 4439.
- 117 Stratmann, R.E., Scuseria, G.E., and Frisch, M.J. (1998) *J. Chem. Phys.*, **109**, 8218.
- 118 Frisch, M.J.T., Trucks, G.W., Schlegel, H.B., Scuseria, G.E. *et al.* (2004) Gaussian 03, Gaussian, Inc., Wallingford CT.
- 119 Ristila, M., Matxain, J.M., Strid, A., and Eriksson, L.A. (2006) *J. Phys. Chem. B*, **110**, 16774.
- 120 Lhiaubet, V., Gutierrez, F., Penaud-Berruyer, F., Amouyal, E. *et al.* (2000) *New J. Chem.*, **24**, 403.
- 121 Monti, S., Sortino, S., DeGuidi, G., and Marconi, G. (1997) *J. Chem. Soc., Faraday Trans.*, **93**, 2269.
- 122 Du, L.W., Liu, X.H., Huang, W.M., and Wang, E.K. (2006) *Electrochim. Acta*, **51**, 5754.
- 123 Charoo, N.A., Shamsher, A.A.A., Kohli, K., Pillai, K.K., and Rahman, Z. (2005) *Chromatographia*, **62**, 493.
- 124 Martinez-Pla, J.J., Martin-Biosca, Y., Sagrado, S., Villanueva-Camanas, R.M., and Medina-Hernandez, M.J. (1047) *J. Chromatogr. A*, **2004**, 255.
- 125 Sortino, S., De Guidi, G., Marconi, G., and Monti, S. (1998) *Photochem. Photobiol.*, **67**, 603.
- 126 de Micalizzi, Y.C., Pappano, N.B., and Debattista, N.B. (1998) *Talanta*, **47**, 525.
- 127 Mehta, S.K., Bala, N., and Sharma, S. (2005) *Colloids Surfaces A – Physicochem. Eng. Aspects*, **268**, 90.
- 128 de Cordova, M.L.F., Barrales, P.O., and Diaz, A.M. (1998) *Anal. Chim. Acta*, **369**, 263.
- 129 Kenawi, I.M., Barsoum, B.N., and Youssef, M.A. (2005) *J. Pharmaceut. Biomed. Anal.*, **37**, 655.
- 130 Jimenez, M.C., Miranda, M.A., Tormos, R., and Vaya, I. (2004) *Photochem. Photobiol. Sci.*, **3**, 1038.
- 131 Lissi, E.A., Encinas, M.V., Lemp, E., and Rubio, M.A. (1993) *Chem. Rev.*, **93**, 699.
- 132 Ravanat, J.L., Berger, M., Buchko, G.W., Benard, J.F. *et al.* (1991) *J. Chim. Phys. Phys.-Chim. Biol.*, **88**, 1069.
- 133 Geiger, P.G., Korytowski, W., Lin, F.B., and Girotti, A.W. (1997) *Free Radical Biol. Med.*, **23**, 57.
- 134 Ali, H. and van Lier, J.E. (1999) *Chem. Rev.*, **99**, 2379.
- 135 Llano, J., Raber, J., and Eriksson, L.A. (2003) *J. Photochem. Photobiol. A*, **154**, 235.
- 136 Sarabia, Z., Hernandez, D., Castell, J.V., and van Henegouwen, G. (2000) *J. Photochem. Photobiol. B*, **58**, 32.
- 137 Moser, J., Hye, A., Lovell, W.W., Earl, L.K. *et al.* (2001) *Toxicol. In Vitro*, **15**, 333.
- 138 Paul, B.J. (2004) *Calicut Med. J.*, **2**, e8.

Part Five

Biochemical Signature of Quantum Indeterminism

31

Quantum Indeterminism, Mutation, Natural Selection, and the Meaning of Life

David N. Stamos

“What I do not understand is why most philosophers of science believe the problems of the philosophy of science can be solved by logic.”

Ernst Mayr ([1], p. ix)

31.1

Introduction

Ever since the beginning of modern evolutionary theory, which resides predominantly in the works of Darwin and Mendel, scientists and philosophers have been wondering about the nature of the variations upon which natural selection feeds. Darwin himself thought it safe to say that while the variations are random with respect to the environment, the variations themselves are not the result of random events but instead have determinate causes. As Darwin put it in the *Origin* [2], “chance . . . is a wholly incorrect expression, but it serves to acknowledge plainly our ignorance of the cause of each particular variation” (p. 131). This remained the dominant view for quite some time. Over half a century later, T.H. Morgan [3], for example, on the topic of “accidental variation” in the theory of natural selection, wrote:

“I need not repeat before this body of naturalists that to-day we have dropped entirely the antiquated word chance as something not subject to the laws of mechanics. That conception of chance arose, no doubt, because chance events are those that can not be predicted individually and what he can not predict seems to the confused thinker to disobey the causal law. Out of his ignorance he imagines blind happenings”

([3] p. 203).

This Laplacean view of the universe, a universe permeated with strict determinism – everything has a cause or causes, and the same causes always produce the

same effects – gave way with the advent of the quantum revolution, during which the view of the universe with strict causal laws was replaced at the subatomic level with statistical laws. Since that relatively rapid revolution in science, the Einsteinian interpretation of quantum physics – that “God does not play dice” with the universe – was gradually eclipsed by the realist interpretation of quantum physics (which stems from the instrumentalism and logical positivism of the Copenhagen interpretation), according to which the statistical laws of quantum physics reflect the fundamentally statistical nature of reality (quantum indeterminism). By the time of the experiments conducted by Alain Aspect in the early 1980s on quantum entanglement, and partly because of them and others like them, the consensus became hardened: there are no hidden variables lurking behind the statistical laws of quantum physics, the quantum world is fundamentally statistical, with irreducible probabilities and uncertainties [4, 5].

With the development of quantum physics and the accumulation of evidence and argument leading to the conclusion that the quantum world of the atom is irreducibly statistical, with genuine ontological chance, scientists and philosophers have wondered about and speculated on what this means for the chemistry of life, and more generally for the meaning of biological evolution. Molecular biologists generally took it to mean that evolution is essentially a matter of real chance. Jacques Monod [6], for example, in the context of his discussion on “quantum perturbations” and “mutation,” wrote that

“chance *alone* is at the source of every innovation, of all creation in the biosphere. Pure chance, absolutely free but blind, at the very root of the stupendous edifice of evolution: this central concept of modern biology is no longer one among other possible or even conceivable hypotheses. It is today the *sole* conceivable hypothesis, the only one that squares with observed and tested fact”

([6], pp. 112–113).

And again:

“A mutation is in itself a microscopic event, a quantum event, to which the principle of uncertainty consequently applies. An event which is hence and by its very nature *essentially* unpredictable”

([6], p. 115).

Or as Alan Weiner put it (personal communication 2000), “chemistry constrains but chance rules.”

Philosophers of biology, however, have been remarkably divided on this topic. As we shall see, some have denied that chance at the quantum level has any relevance for evolutionary processes, while others have taken issue with them, and so a real debate has ensued. Philosophers are particularly sensitive not just to the nature of argument but to conceptual issues, so it should be no surprise that philosophers of biology have spent a lot of effort trying to clarify key concepts in biology as much as possible, in this case “chance,” “natural selection,” and “evolution.”

I became a part of this debate shortly after it dawned on me that all were ignoring important evidence concerning the mechanisms of mutation, particularly the different ways in which a quantum event could trigger a point mutation. That seemed to me crucially important, for knowledge is one of the forces in conceptual change. This is especially so in science, the history of which makes abundantly obvious. As knowledge grows in science, new concepts are introduced and old ones are often forced either to change or to disappear altogether. In biology, for example, the concept of “vital force” went extinct, just as did the concept of “phlogiston” in chemistry, while the concepts of “species” and “gene” have undergone and continue to undergo descent with modification. One way or another, *facts matter to concepts*.

What I shall attempt to accomplish in the present chapter is to (i) provide a brief sketch of the recent history of the debate in philosophy of biology over the nature of chance in natural selection and evolution, (ii) provide a summary and update of the evidence for the role of quantum chance (indeterminism) in the origin of point mutations, (iii) provide an argument which concludes that this evidence forces us to modify the concepts of “chance,” “mutation,” “natural selection,” and “evolution,” and (iv) provide a discussion on what all of this means for the big question of life, namely, the meaning of life.

31.2

A Short History of the Debate in Philosophy of Biology

Elliott Sober, in his important book on the nature of natural selection [7], raised the issue of whether quantum chance can “percolate up” to the macro-level of biological evolution. He himself was not convinced that mere large numbers would cancel out the indeterminism at the micro-level, leaving the macro-level deterministic, since “if *enough* elementary particles had behaved differently, the behavior of the macro-object (the organism, the population) that they compose would also have been different.” Sober goes on almost immediately to say that “If chance is real at the micro-level, it must be real at the macro-level as well” (p. 121).

But is that a good reason to think that the processes of biological evolution are fundamentally indeterministic? What the skeptic against genuine chance in evolution wants is a good reason to believe that chance events at the micro-level can occur in ways where they don’t cancel each other out, ways which thereby affect with genuine chance the paths of evolution. The skeptic, moreover, will want to know how a quantum chance event can make a change in an organism, particularly in a way that affects evolutionary outcomes. Without any specifics on this matter, the skeptic will remain unconvinced, and rightly so I would add.

Interestingly, the very point where one would expect the connection – namely, mutations – is overlooked by Sober. He takes it as given that “In evolutionary theory, mutation and selection are treated as *deterministic* forces of evolution” and only random genetic drift “is the stochastic element in evolution” ([7], p. 110). Random genetic drift, of course, is not in itself (if considered apart from mutation) an example

of genuine irreducible chance. The nature of chance in that process, instead, is epistemological, not ontological. Natural selection is another matter, to be dealt with separately. Particularly telling is the implication in Sober's statement that evolutionary theory is some kind of edifice, that it is static not dynamic, that it has not undergone any real change since Darwin, especially with respect to the concepts of mutation and natural selection. We shall see this again in other philosophers of biology. The idea is that evolutionary theory is already given and we just have to sharpen our understanding of it. Nothing, however, could be further from the truth. Ever since Darwin, mutation ("heritable variation" in Darwin's writings) has always been a major part of the theory, and if the theory of mutation has changed since the time of Darwin to the present (and it has), then evolutionary theory has changed as well. Even more, the concept of natural selection may have to be next.

For the moment, it is important to notice that skepticism about genuine chance in biological evolution abounds in philosophy of biology. According to Alex Rosenberg [8], for example, "evolution can proceed in the absence of mutation and would do so even if mutation were a thoroughly deterministic process" (p. 60). The absence of mutation, of course, would allow for only micro-evolution, not macro-evolution, certainly not the evolution of naked genes to present-day biodiversity, while the absence of indeterminism in mutation would profoundly affect our understanding of evolutionary processes. But the mutation-denial does not stop there. For Rosenberg [8], "quantum probabilities involved in biological processes are so small . . . [that] By the time nature gets to this level [the macro-level of biological processes], it has long since asymptotically approached determinism" (p. 61). A very different kind of mutation-denial is exemplified by Barbara Horan [9], who accepts that some mutations may be the result of genuine chance at the quantum level. But she then goes on to say, concerning "the statistical character of evolutionary theory," that

"There has been some discussion in the literature about whether the theory describing genetic processes that create the basis on which evolutionary forces act should be regarded as *part* of the theory of evolution or *ancillary* to it. Standard population genetics textbooks treat mutation as a *source* of variation for, for example, natural selection. The randomness of the *process* of mutation might therefore be regarded as irrelevant"

([9], p. 83n).

So either quantum indeterminism does not significantly affect evolutionary processes or if it does it is irrelevant to evolutionary theory. Something is seriously amiss here. I suspect that a major source of the error, again, is the belief that evolutionary theory is a static rather than dynamic entity. But an additional source of the error, and not unrelated, is the belief that biology and chemistry/physics are sufficiently discrete sciences and that this reflects the phenomena they investigate. A very different view, of course, is exhibited by the authors in this volume, and also by the authors of its nominal predecessor. As Pullman and Pullman [10] put it, one of their goals was "to show biochemists how quantum mechanics can yield answers to

the problems of the structure and mode of action of the constituents of living matter" (p. v). Underlying the very idea of *quantum biochemistry*, then, would appear to be the view that the division of the sciences into physics, chemistry, and biology is largely if not totally a man-made one, born out of the need for professional specialization, but where the reality collectively studied is a seamless whole. Understanding how quantum chance events can affect evolution via mutations contributes to this understanding in no small way. But it is not strictly a scientific issue. Our understanding of the meaning of the terms "mutation," "natural selection," and "evolution" are at stake here, as well as, ultimately, the very meaning of "life" itself. Hence the interest by philosophers of biology.

The debate in philosophy of biology over chance in evolution became especially interesting with the publication of a paper by Brandon and Carson [11]. In that paper they provide a simple theoretical example to illustrate how a single mutation, caused by a quantum chance event, could possibly alter an evolutionary outcome. The example involves a haploid population with two alleles of the same gene, *A* and *a*, such that the population is stable when all the members of the population have either the *A* allele or the *a* allele or when the ratio is 50 : 50. If the frequency of the alleles in the population is actually 50 : 50, then a mutation that turns an *A* into an *a* will be selected for and the population will be driven to fixation for *a*, and likewise to *A* if instead the mutation is one that turns an *a* into an *A*. If the point mutation is caused by a quantum chance event, then "quantum uncertainty would 'percolate up' in a powerful way to the level of populations" (p. 320). Brandon and Carson provide no idea on how a quantum event could actually cause a point mutation, but they think the connection is plausible given the basis of biology in chemistry and physics.

The debate became even more interesting with the reply paper by Graves *et al.* [12]. In that paper they argue that evolutionary theory is probabilistic not because any of the processes of evolution are indeterministic but because of human epistemic limits, in that humans cannot possibly know let alone predict all the nitty-gritty causal events that occur in evolutionary processes. Along the way they make some rather startling claims. For one, they say "Consider the shape and complexity of an adenosine molecule. The changes required to mutate this molecule into a guanine molecule would be quite considerable, clearly involving a substantial aggregation of micro-processes" (p. 144n7). To this they add the consideration that the smallest genes have a high number of nucleotides, that a mutation in one direction in a population can be offset by a mutation in the opposite direction, that the genetic code has a high level of redundancy in it, and that amino acid substitutions in homologous proteins have a small effect. From all of this they conclude that the odds of a quantum chance event ever percolating up to the level of biological processes "asymptotically approach zero" (p. 145). They also claim that the burden of proof in this debate lies on the shoulders of the indeterminist (much as the analogous debate did in physics), that determinism at the level of biological processes should be the "default position" (p. 152), the "working assumption," that "Even quantum mechanics recognizes that at the level of the macromolecule, nature asymptotically approaches determinism," and that "the indeterminist is no biologist" (p. 153).

That last point is odd, much like saying the quantum physicist is no physicist. The second last point is also odd, since DNA certainly qualifies as a macromolecule and yet no modern physicist, aware of the nature of mutations, and certainly no molecular biologist, would claim that DNA replication asymptotically approaches determinism. In my reply paper [13] to Graves *et al.* [12] I pointed out furthermore that point mutations do not proceed by one nucleotide changing into another but by the wrong nucleotide being inserted during DNA replication, that there is a large and constantly growing number of known examples of point mutations (whether in small or large genes) that have a noticeable phenotypic effect, that the relative infrequency of point mutations renders unlikely the chance that a “forward mutation” will be offset by a “back mutation,” that the redundancy in the genetic code is almost completely confined to the third position in codons (thus severely limiting the proportion of silent mutations in coding DNA), and that the study of the high similarity in homologous proteins from different species does nothing to minimize the importance of single mutations since such examples by their very nature overlook the much more numerous *dysfunctional* proteins that must have arisen in evolutionary history and that were selected against (and, I should add here, they also overlook the evolution of the ancestral proteins in the first place).

But the main thrust of my paper was to bring together the various pieces of evidence for how a quantum chance event could trigger a point mutation. (Since most mutations are point mutations, as well as most favorable mutations, that seemed the logical focus.) I was troubled not only by those philosophers who were skeptical of any such connection, but by those philosophers *and molecular biologists* who were sure that there was such a connection but never specified its nature. In other words, I wanted to know the actual mechanism or mechanisms. But no one author seemed to have the answer, not even the molecular biologists (I read and contacted many). Instead, I found only bits and pieces here and there and had to put the picture together as best I could.

In the section following the next, I provide a summary of that picture and attempt to update the evidence. But first it is useful to look at some of the replies to my paper, as important foils for the present chapter.

31.3

Replies to My Paper

Alex Rosenberg [14], for example, concedes that much of what I wrote about “the quantum-mechanical sources of mutation is . . . correct,” so that “biology is indeed indeterministic in at least some of its most important fundamental processes.” But then he immediately claims that I still have failed “to rightly identify the source of the statistical character of the theory of natural selection” (p. 537). On the one hand he admits (what Reference [12] did not) that “in the case of mutations quantum mechanical percolation occurs,” but then on the other hand he states that “Stamos infers directly and illicitly that therefore the probabilities of the theory of natural selection must be the pure probabilistic propensities of quantum mechanics”

(p. 538). That, however, was not what my paper was at all about. It was not about natural selection, but about evolutionary biology, the body of theory that includes natural selection but is certainly not confined to it. (That too, by the way, was the focus of the paper by Graves *et al.* [12].) I concluded that evolutionary theory has in fact substantially changed due to the accumulated evidence supporting the view that point mutations have a basis in quantum indeterminism, that this connection diminishes the autonomy of biology from physics and chemistry, and that philosophers of biology need to take cognizance of all of this. I did not conclude anything about natural selection, the process or the theory. Of course, if one thinks of natural selection in the traditional way, as not including the source of the heritable variation on which it feeds, then Rosenberg's points stands with regard to the nature of probability in the theory of natural selection (even though he missed my point). In the section following the next, however, I will indeed argue against that traditional conception of natural selection, something I did not do in my paper.

In connection with his reply, Rosenberg ([14], pp. 541–543) asserts that natural selection is a statistical law much like the second law of thermodynamics, in that it is a domain-neutral law concerned only with the behavior of large ensembles of individuals, not the individuals that compose them. More recently he has distanced himself from the thermodynamics analogy, because with thermodynamics entropy is not derived from the properties of the individuals that make up the ensemble, whereas in biology the fitness of a population is derived from the fitness levels of the individuals that make up the population ([15], pp. 701–704; [16], pp. 158–159, 170, 185–186). Others, however, have maintained the thermodynamics analogy (e.g., [17], pp. 71–72, 81–83; [18], pp. 285–289). (Presumably, Rosenberg wants to say that natural selection is derived from the properties of the individuals in a population and their environment.) At any rate, the main problem with the ensemble approach (derived or underived) is that it precludes the percolation argument almost as if by definition. The question still remains, and it is an extremely important one, as to whether individual mutations make a difference to evolutionary trajectories, especially within the context of natural selection. That is a question I will attempt to answer in the section following the next one.

In a vein similar to Rosenberg's, Roberta Millstein [19] misses the upshot of my paper. She claims that "the determinism question," even though it is currently a hot topic among philosophers of biology, is a question that does not really matter since "it appears to garner little or no attention from evolutionary biologists" (p. 1318). Following population genetics, according to which selection, drift, and evolution are population-level processes, "it is the population-level factors that will make a difference over the long run, not the facts about individual organisms" (p. 1327). The main problem here is to think of population genetics as the core or foundation of modern evolutionary theory. It is neither, not even for the principle of natural selection [20], nor can it be a static entity if it truly, as Hartl [21] claims, "cuts across so many disciplines, among them molecular biology, . . ." (p. ix).

In another paper, Millstein [22] concedes that my paper provides good reason to believe that point mutations are "fundamentally indeterministic," but she adds that it does not really matter to the concept of chance in evolutionary biology. One can just

as well imagine, she says, that the point mutations that play a role in natural selection and random drift are completely deterministic, in which case “the concepts of chance would remain the same.” This shows, she claims, that “the chance phenomena under discussion are independent of the lower-level causes that produce mutation” (p. 684). But that does not follow at all. If the concept of point mutation in biology changes from being that of a deterministic process to an indeterministic one, then the concept of chance in evolutionary biology necessarily changes as well. In the former scenario, the deterministic biologist would hold that two absolutely identical populations in absolutely identical environments would undergo absolutely identical processes of natural selection and genetic drift, and hence would have absolutely identical evolutionary outcomes. But the indeterministic biologist would not share that view. Instead, he or she would emphasize genuine contingency in the evolutionary outcomes. Moreover, not only the sense but the reference of the word “chance” would be different, since the deterministic biologist would be using the word only in the sense of epistemic finitude and with reference only to macro-level processes, while the indeterministic biologist would be using the word with an expanded sense and an expanded reference, the sense expanded to include genuine contingency rooted in the ultimate source of variation and the reference expanded to include the quantum level.

In a related line of response to my paper, Timothy Shanahan [23] argues that the evidence I brought forward has no bearing on the real issue in philosophy of biology, which is the debate between adherents of the evolutionary determinism thesis and those of the evolutionary indeterminism thesis. The former thesis states that the processes of evolution are fully deterministic and that the probabilities in evolutionary theory are purely epistemic, while the evolutionary indeterminism thesis states that the probabilities in evolutionary reflect genuine indeterminacies in evolutionary processes themselves (p. 164). More precisely, the evolutionary indeterminism thesis, says Shanahan, states that evolutionary processes, independently of any quantum indeterminism that might percolate up to the biological level, has genuine “irreducible” indeterminacies of its own (p. 165). The problem with stating the debate in this way, however, is that it commits a false dichotomy. While Graves *et al.* [12] are clearly defenders of the evolutionary determinism thesis, my own view [13] falls neither into that category or the other. But there is a third alternative, one that fails to get stated in Shanahan’s paper but which needs to be clearly stated, which is that the probabilities in evolutionary theory have to be substantially (not completely) reconceptualized so as to include quantum indeterminism. What makes this alternative a real one is the basic idea underlying quantum biochemistry and the growing evidence supporting it, altogether the idea that biology is a science that is not fully autonomous from physics and chemistry. To deny that possibility *tout court* is to beg the question. My suspicion is that most if not all of the philosophers of biology who reject that third alternative are guilty of this fallacy and the territorialism that goes with it.

With these foils in mind, it is time now to turn to an updated look at the case for the basis of point mutations in quantum indeterminism and to an expanded concept of natural selection that includes that basis.

31.4

The Quantum Indeterministic Basis of Mutations

31.4.1

Tautomeric Shifts

The evidence for the basis of mutations in quantum indeterminism (my focus will be on point mutations entirely) lies in no one experiment or piece of evidence, but rather in the convergence of various lines of evidence. When I worked on this topic for my paper [13], no one person or publication (that I could find) appeared to have the whole picture, so I endeavored to put the pieces together myself. What emerged was very strong evidence for two robust pathways between the quantum world and mutations, with a third pathway that seemed to me quite plausible (and still does) given the physics of DNA and its medium. The following is a summary with some updating.

One of the robust pathways is that of a point mutation caused by a tautomeric shift. Following the suggestion by Erwin Schrödinger [24] that a mutation might be an isomeric transition caused by a “quantum jump” (p. 48), a “chance fluctuation of the vibrational energy” (p. 63), Watson and Crick, in their seminal papers on the nature of DNA published in 1953 and 1954, applied his idea to DNA. Each of the four DNA bases has a major tautomeric form, a relatively stable state of chemical equilibrium. With each of the four bases their hydrogen atoms have their preferred locations. In the double helix, adenine and thymine, connected by two hydrogen bonds, form one of the standard pairs, while guanine and cytosine, connected by three hydrogen bonds, form the other standard pair. (Interestingly, Watson and Crick thought the latter involved two hydrogen bonds as well. The latest picture for hydrogen bonds in both base pairs is more complicated, see Parthasarathi *et al.* [25], but the complications can be ignored here.) The equilibrium of each base is sometimes offset temporarily, resulting in a minor tautomeric form. This can happen to a base either in the template strand or in a free nucleotide. One of the hydrogen atoms will actually jump to a different position on the base, and then jump back. During DNA replication the minor tautomeric form makes it impossible for the base to pair with its complementary base, but it can pair with a different base that has the appropriate structure, during the instant before the hydrogen atom in the minor tautomer jumps back to its original position. Figure 31.1 is based on the diagram provided by Watson and Crick ([26], p. 272), with the middle step added. Adenine (A) undergoes a tautomeric shift when the hydrogen atom that is bonded with the nitrogen atom at the top of the molecule jumps to the nitrogen atom below it, resulting in the minor tautomer (A*). During DNA replication the minor tautomer (A*) will then bond with cytosine (C) rather than with thymine (T).

In the next round of replication, the A* will most likely revert back to an A and bond with a T, while the C will bond with a G. Hence the original base pair A · T results in the base pair G · C, which is a mutation. Figure 31.2 shows the tautomeric shifts for each of the four bases.

The connection between a tautomeric shift and quantum indeterminism resides in the phenomenon known in quantum physics as a “quantum transition” or “quantum

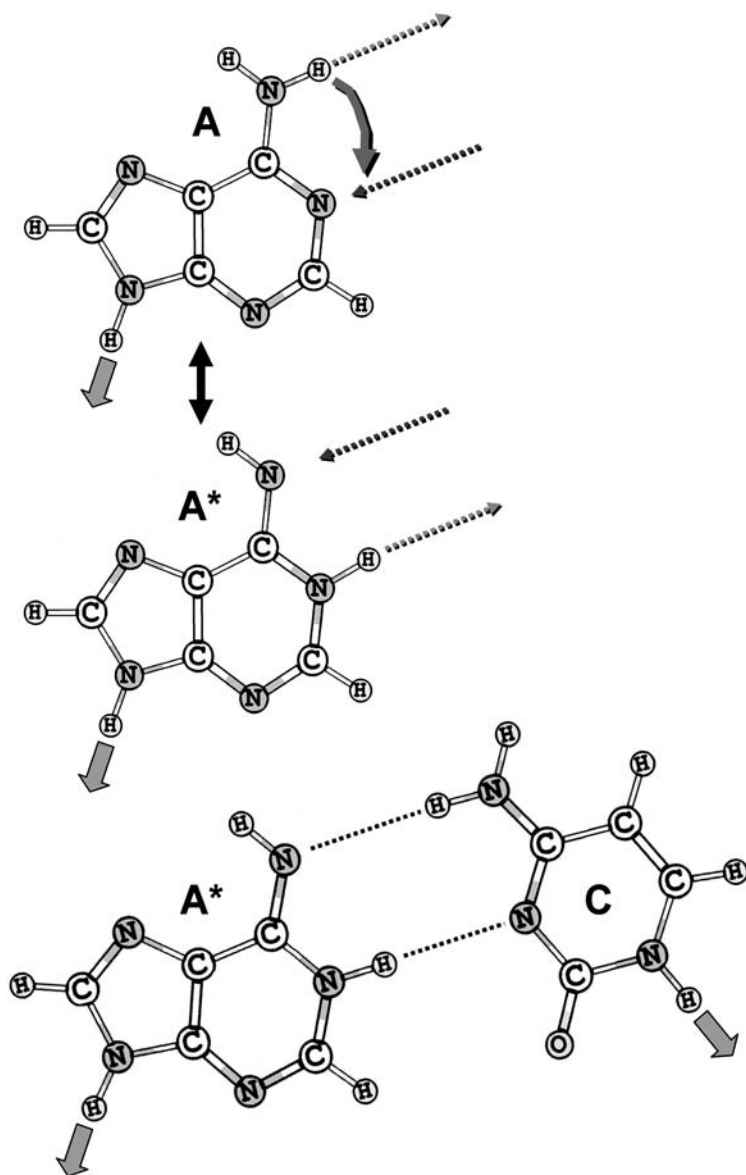


Figure 31.1 Point mutation caused by a tautomeric shift. (All figures by courtesy of Chérif F. Matta © 2009.)

jump” (the latter comes from the days of the old Bohr model of the atom). With regard to electrons in an atom, a quantum transition is a spontaneous transition of an electron from one orbital to another. The first electron ever isolated was accomplished in 1973 by Hans Dehmelt using a Penning trap, who then isolated the first atom in a similar trap six years later. The modern view of electrons in an atom, of course, is not

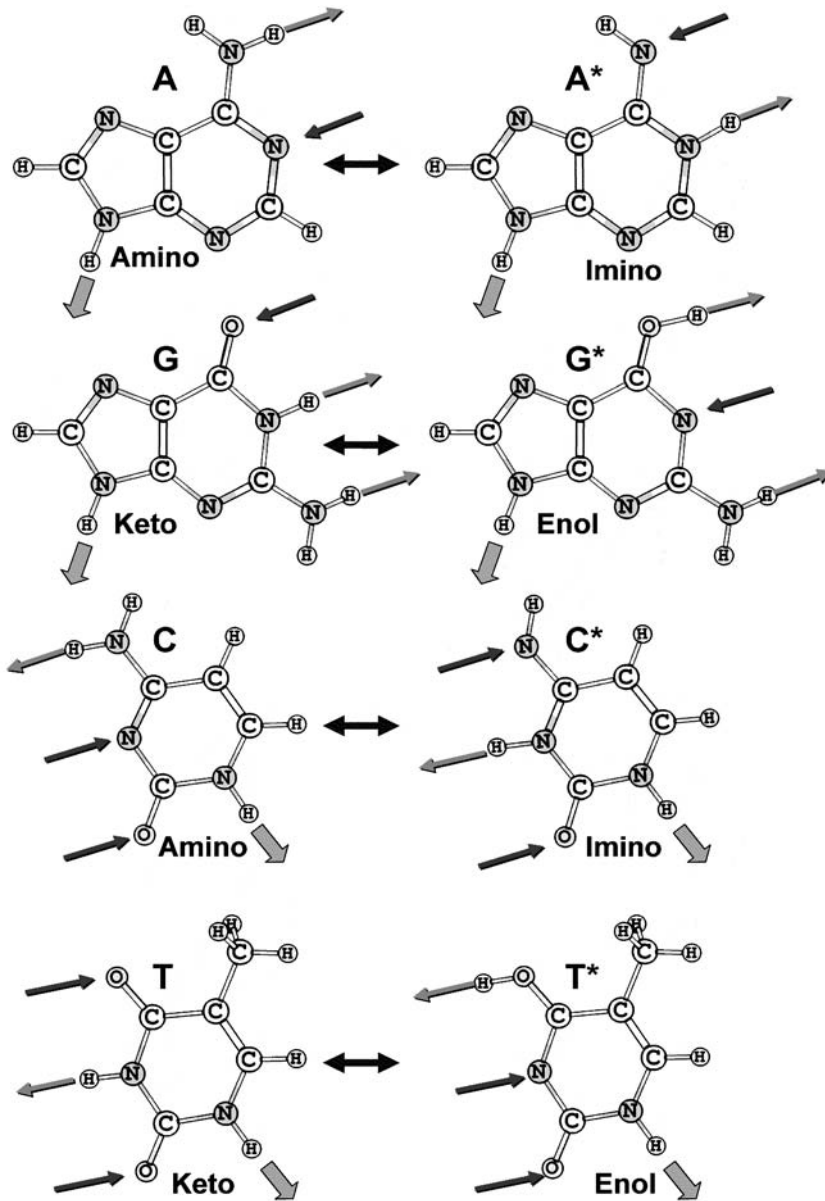


Figure 31.2 The tautomers on the left are the most stable form of these bases present in the (normal) Watson–Crick base pairs while the ones on the right are unstable tautomers that can lead to mismatch mutations. (Figure courtesy of Chérif Matta © 2009.)

that of a particle but of a cloud of weighted probability density surrounding the nucleus. On this view, a quantum transition entails a spontaneous change in the electronic structure of the atom. With a trapped atom illuminated by, for example, ultraviolet light, one can actually see the spontaneous change, albeit indirectly, when the atom suddenly goes into an unreceptive state, temporarily failing to absorb or emit photons ([27], p. 99). In this way one can actually see an event of pure quantum chance.

By the early 1970s, Monod [6], for example, claimed factual status for the Watson–Crick model of mutation by tautomeric shift, conceived, moreover, as a “quantum event” (p. 114), of which mutagens only “augment the probability” (p. 192). His confidence in the Watson–Crick model, however, was premature. Von Borstel [28], in particular, while noting the model’s popularity, especially in textbooks, claimed not only that direct experimental evidence of mutation by tautomeric shift is lacking, but that the length of time for a tautomeric shift is too short to be responsible for a mutation: “the rare tautomer does not *persist* in the double helix and without its persistence, its effect on mutation production is a mirage” (p. 132). Instead, he summarized recent evidence for different causes of mutations, none of which involve tautomeric shifts:

- 1) **Bond angles:** a change in the angle of a chemical bond in a single strand of DNA during replication can sometimes result in a non-standard base pair.
- 2) **Wobble:** guanine and cytosine normally pair using three hydrogen bonds, but each can sometimes become misaligned, allowing the former to pair with thymine using only two hydrogen bonds and similarly cytosine with adenine.
- 3) **Protonation:** The addition of a proton, usually to a ring nitrogen, results in an ionized form of the base (sometimes stable), and hence in a different bonding configuration (similar to a minor tautomer but without tautomerization). This can sometimes be combined with wobble. Hunter and Brown [29] provide over 20 diagrams of non-standard base pairs, many involving altered bases, including a base pair that involves four hydrogen bonds between the bases. In addition, von Borstel’s [28] diagrams include an A · C wobble with only one hydrogen bond between the bases (p. 133).
- 4) **Transient misalignment and dislocation:** Occasionally during DNA synthesis a base in a single-stranded template will temporarily swing backward, thus failing to pair with a base, but then it will spring back to its original position, pairing with the base previously paired with its neighbor, which might not be a complementary base.

In a different line of criticism, some researchers appealed to evidence based on experiments with non-polar DNA base analogs, in which analogs incapable of forming hydrogen bonds were nevertheless paired by DNA polymerase with a standard base. Moran *et al.* [30] and Goodman [31], for example, conclude that conventional hydrogen bonds may be of minor importance for high efficiency and fidelity during DNA replication, and that geometry (size and shape of the molecule) is probably the most important if not the only factor.

The Watson–Crick model of mutation by tautomeric shift, however, was not done. Some defenders raised doubts about the alternative models discussed above. Florián

and Leszczyński ([32], p. 3010), for example, claim that the methods used by researchers to determine ionized and/or wobble base pairing involve conditions different from those that actually obtain during DNA replication – moreover that wobble mismatches are more easily located and excised by the repair mechanisms than mismatches involving minor tautomers. Other researchers used base analogs and synthetic bases that involve much greater frequencies of minor tautomers and hence are more amenable to experiment. Combined with techniques such as X-ray crystallography and nuclear magnetic resonance spectroscopy, their results reaffirmed the roles of hydrogen bonds and tautomeric shifts (e.g., References [33–38]). Computer models have also begun to be used to explore the role of tautomeric shifts in mutation, with positive results (e.g., Reference [39]).

Of considerable importance is the medium of DNA, which is mainly aqueous. Kwiatkowski *et al.* [40] review the experimental evidence for the conclusion that minor tautomers of the four standard bases occur in aqueous solution in a ratio of 1 in 10^4 – 10^5 (p. 124; ditto [41], p. 723). They also review the evidence that shows that the ratio between major and minor tautomeric forms of the bases is roughly equal when isolated in an inert gas. Poltev *et al.* [41] draw the obvious conclusion from these and other lines of research, which is that “hydration should be the main, if not the only, source of the preference for the keto [major] tautomer in aqueous solutions” (p. 723). This conclusion has enormous implications for von Borstel’s line of criticism discussed above. Doublié and Ellenberger [42] point out that at the replication fork, when the DNA polymerase is fitting the nucleotide opposite the template base, “More than 90% of the surface of the base pair is buried from [the aqueous] solvent” (p. 708). Given the conclusion of Poltev *et al.* [41] quoted above, this suggests that a minor tautomer while grasped in the “hand” of the DNA polymerase (Doublié and Ellenberger actually use the words “palm,” “fingers” and “thumb” in their diagram) can remain stable long enough (because it is not in the aqueous solution) for insertion opposite the template base.

What further aids this picture is growing evidence for the role of metal ions, such as cadmium. Not only can they influence the formation of rare tautomeric forms of the bases, but they can also help stabilize already existing ones, hence increasing the role of free rare tautomers in mutagenesis [43, 44].

All of this does not mean, of course, that other processes may not be responsible for point mutations as well, such as wobble, temporary misalignments and proton tunneling (discussed below), but it does strongly suggest that the Watson–Crick model of tautomeric shifts is here to stay and will maintain a major place in the mutation model mix.

31.4.2

Proton Tunneling

Another robust pathway between the quantum world and mutations involves what is known as “proton tunneling,” a well-known statistical phenomenon in which a quantum particle penetrates a potential barrier, such as the escape of an alpha particle from an atomic nucleus. This is possible because a particle in quantum physics is not

conceived classically as a solid object but as a “wave packet” subject to the statistical laws of quantum theory. Per-Olov Löwdin [45] was the first to develop in detail a theory of spontaneous point mutation caused initially by tautomerization *within* the double helix, not involving rare tautomers in the aqueous environment. Explicitly using from quantum theory the concept of the “quantum jump” (p. 222), he argued that occasionally a proton (the nucleus of a hydrogen atom) could tunnel through the potential energy barrier of a hydrogen bond in a DNA base pair. This would be impossible for a proton conceived as a classical particle, but not for a proton conceived as a wave packet “obeying the laws of quantum theory” (p. 219). The energy thresholds required for these proton transfers, he noticed, are high enough to make them relatively rare, close to the estimated rates of point mutations. Löwdin also argued that since each of the four standard bases has a neutral charge, the proton transfer would almost certainly cause a simultaneous anti-parallel proton transfer through another hydrogen bond in the base pair. Visualized as a “double-well potential,” this would keep the charges neutral and restore equilibrium. In this way minor tautomers produced by proton tunneling could sometimes be involved in DNA replication, causing mutations. For example, parallel proton tunneling in the A·T bond would produce the tautomeric bond A*·T*. When split during DNA replication, each minor tautomer would have a different pairing pattern and so would not serve as a template for the standard complementary base but instead for a non-complementary one, resulting in the pairs A*·C and G·T* in the separate strands. Figure 31.3 is based on Löwdin’s main diagram (p. 220), not his one with subsequent tunneling (p. 335).

Löwdin also theorized that a single proton could tunnel between complementary base pairs, likely as the result of exposure to a mutagen such as ionizing radiation. This would lead to base pairs such as A⁻·T⁺ or G⁺·C⁻. If any of these ionized tautomers were to remain ionized at the time of DNA replication, it would be unlikely to pair with a neutral standard base, resulting most likely in a frameshift mutation in the form of a deletion.

Löwdin offered his model of mutation as an addition to the Watson–Crick model of tautomeric shifts, recognizing that it was impossible to say which model accounted for more mutations than the other. Two main objections to his model, however, remained before his theory could hope to gain wide acceptance. One was the uncertainty over the nature of the hydrogen bond, while the other was the lack of experimental evidence of proton tunneling in DNA base pairs.

Beginning with the latter, it would not be an exaggeration to say that during the 1990s to the present a substantial amount of evidence in support of Löwdin’s model has been pouring in. For example, Florián and Leszczyński [32] examined the energy thresholds required for proton tunneling in the standard base pairs. In the G·C pair, they found that there are only two possibilities for double proton transfer wherein the base pair charges remain neutral, and that one of them is so unlikely that it is not worth consideration. The remaining possibility has an energy threshold that results in an estimated ratio of 1 in 10⁶–10⁹ (G*·C*: G·C), which they point out is of obvious significance for the topic of spontaneous mutation. The average rate of spontaneous mutation in DNA, prior to proofreading and mismatch repair by the

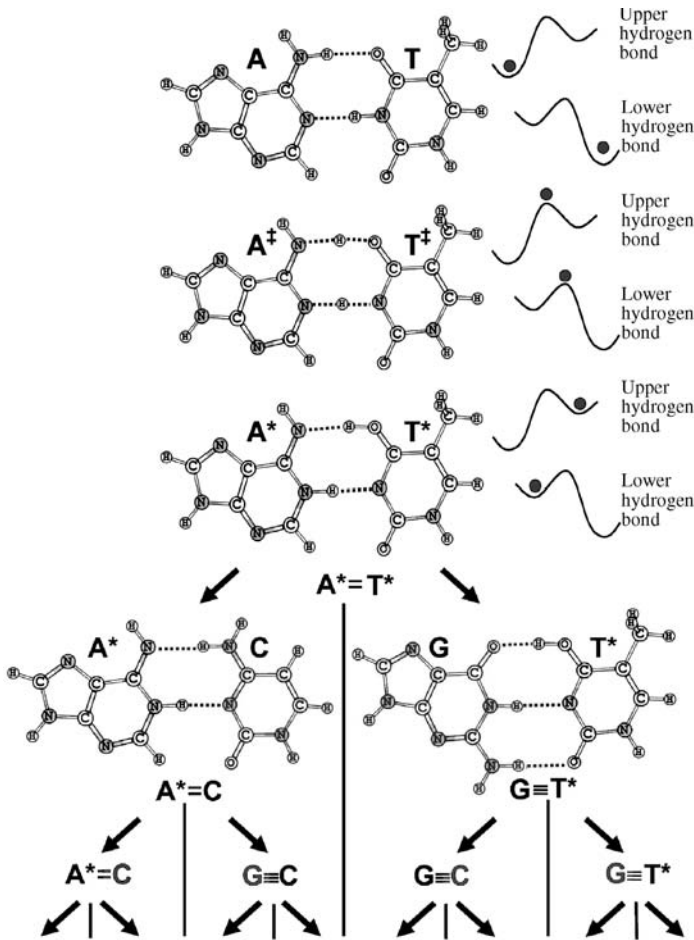


Figure 31.3 Löwdin's mechanism for proton tunneling.

DNA polymerase, ranges in the vicinity between 1 in 10^4 to 1 in 10^8 per base-pair replication ([46], p. 16; [31], p. 640), which is reduced by DNA repair mechanisms by a factor of 10^2 – 10^3 ([39], p. 879).

As for the energy threshold for the transition from A·T to A*·T*, Florián and Leszczyński [32] found it to be much higher, resulting in a ratio of 1 in 10^{12} (A*·T*: A·T), which, they point out, eliminates that part of Löwdin's theory as a probable cause for spontaneous mutation. More recent evidence [47] has similarly indicated that double-proton transfer is much more likely in the G·C pair than in the A·T pair, but also that the hydrogen bonds in the resulting pairs (A*·T*, G*·C*) are stronger than in the original pairs, which may be of significance for their stabilization during DNA replication. Kryachko and Sabin [48], on the other hand, provide strong evidence for a novel mechanism of proton tunneling in the A·T pair, involving what they call "base flipping," which results in an estimated ratio of 1 in 10^6 – 10^9 .

($A^* \cdot T^* : A \cdot T$), which is of obvious significance for mutation. In their concluding section, however, Kryachko and Sabin leave it open to further investigation as to whether proton tunneling leads to more conversions in $A \cdot T$ compared with $G \cdot C$. What is surely significant is that their proposed mechanism for proton tunneling in $A \cdot T$ involves at least three stages (proton tunneling here is not purely tautomeric, but serves instead to break either of the hydrogen bonds, leading then to base flipping) while proton tunneling in $G \cdot C$ involves only one stage, and moreover that mutation rates for the four bases have been established to be higher for G and C, specifically A: 20.3%, T: 20.4%, G: 29.7%, C: 29.5% (from Reference [49], p. 126, although interestingly they offer no explanation for the differences).

The remaining objection to the Löwdin model of mutation concerns the nature of the hydrogen bond. (This objection also relates to what seems to me a very plausible third robust pathway from the quantum world to point mutations, which is to follow below.) Löwdin himself [45, 50], in interpreting proton tunneling as a quantum transition, explicitly aligned his theory with Linus Pauling's model of the hydrogen bond developed in the 1930s, in which the hydrogen bond is partly covalent, involving significantly overlapping wave functions. Although Pauling's model has long remained a popular one, many have argued against it, in favor of an essentially electrostatic model.

In 1998 the evidence finally began to take a clear direction, in favor of the Pauling model. An experiment conducted in Grenoble established that the hydrogen bond in ice is roughly 10% covalent, the rest electrostatic ([51], p. 405). More recent work on the water dimer (two water molecules bound by a single hydrogen bond) yielded similar results [52]. Given that the environment of DNA is mainly water, and therefore also of DNA polymerase during DNA replication, this suggests a third robust pathway from the world of the quantum to the level of mutations.

31.4.3

Aqueous Thermal Motion

DNA polymerase is a large and complex molecule, made of approximately 1000 amino acids. As with the DNA molecule itself, every part of the DNA polymerase is subject to Brownian motion, more accurately thermal motion (wind, noise). Molecular biologists virtually all agree that thermal motion has to be responsible, in part, for DNA polymerase infidelity (Tom Martin, Tom Schneider, personal communications, 2000). In other words, the idea is that thermal motion can occasionally be significant enough to cause base-pair mismatches such as wobble configurations. This is where quantum indeterminism can play a further role. The thermal motion constantly acting on every part of DNA and DNA polymerase is caused mainly by the motion of surrounding water molecules, groups of H_2O molecules (dimers, trimers, etc.) each held together by one or more hydrogen bonds. Proton tunneling within those bonds would be expected to change the motion of the water molecule. This is because although proton tunneling does not change the overall charge of the water molecule, it does change its relative polarity, and hence its interaction with other molecules. Whether singly or in conjunction with other water

molecules, it is conceivable that proton tunneling within water molecules, events involving pure quantum chance, could affect the motion of one or more parts of DNA and DNA polymerase, and consequently the processes of DNA replication and repair.

Pugliano and Saykally [53, 54] were the first to provide definitive measurements of proton tunneling in the water dimer and trimer. A few years later Tuckerman *et al.* [55] found pronounced proton tunneling in H_3O_2^- , while they found that H_5O_2^+ behaves in an essentially classical manner.

The theory of point mutation caused by quantum-induced thermal motion on DNA and DNA polymerase remains a pet speculation, which may or may not receive substantial empirical support or refutation, but at the very least it opens the door to the possibility that the indeterminism of the quantum world can affect the world of mutations in more ways than just tautomeric shifts within bases and proton tunneling between base pairs.

Mutagens and other factors, moreover, do not substantially change the overall picture. Given that chemical mutagens such as nitrous acid and physical mutagens such as X-rays, UV light, and heat can dramatically change mutation rates ([56], pp. 93–97), that flanking base pairs can have an effect on mutation rates ([57], p. 331), that natural selection can drive mutation rates ([58], pp. 140–142), and that there are phenomena such as mutational hot spots ([49], pp. 37–38) and directed mutation [59], all of it does nothing to change the fundamentally indeterministic nature of mutations. As Monod [6] put it long ago, they merely “augment the probability” (p. 192). In the end, just as with radioactive decay, we are still dealing with, to use John Drake’s apt phrase (personal communication 2000), “constrained randomness,” the implications of which are profound.

31.5

Mutation and the Direction of Evolution

Can individual point mutations make a difference to evolutionary trajectories? This question is important for the percolation argument, for if individual point mutations do not make a difference, then their foundation in quantum indeterminism is irrelevant for evolution, defined as the origin of adaptations and of species.

The focus on point mutations is a legitimate one, for several reasons. First, while many point mutations are silent, most of the remainder are only mildly deleterious, resulting in recessive alleles, while a small proportion are, of course, advantageous. Larger mutations, chromosomal mutations, are not only much more rare, but (with the exception of polyploidy) commonly result in abnormalities and are often lethal. Recombination, while it accounts for far more variation in any generation than point mutations, produces results that are neither lasting nor the ultimate source of variation. And finally, as we have seen in the previous section, the role of a single hydrogen bond in point mutations constitutes the connection to the quantum domain, and hence to genuine contingency.

Interestingly, Darwin himself, though of course unaware of point mutations, held the view that a single small change in the hereditary material, what today would refer

mainly to an individual point mutation, could possibly alter an evolutionary trajectory. As he put it in the *Origin* [2],

“A grain in the balance will determine which individual shall live and which shall die, – which variety or species shall increase in number, and which shall decrease, or finally become extinct”

(p. 467).

Alfred Russel Wallace, on the other hand, suggested to Darwin,

“(what I certainly believe to be the fact) that *variations of every kind are always occurring in every part of every species*, – and therefore that favorable variations are *always ready when wanted*”

([60], p. 229; *emphases in the original*).

Darwin’s view can be found in modern authors. According to G.C. Williams [61], for example,

“Minor events such as a particular mutation in a particular individual at a particular time could have important effects on all descendants of the mutant individual and all others that interact with these descendants. An analogy is the butterfly effect in meteorology [chaos theory]”

(p. 6).

Wallace’s view can also be found among modern authors. In Peter Grant’s [62] study of Darwin’s Finches, for example, mutations are explicitly rejected as unnecessary in accounting for adaptive evolutionary trajectories, in favor of natural selection operating on preexisting variation (pp. 278–288, 294, 312).

Is there any evidence that could help settle the matter? Presumably it is safe to say that many adaptive evolutionary trajectories are the result of natural selection operating on preexisting variation. (I am assuming at this point, purely for the sake of argument, that natural selection is a process separate from the origin of heritable variation.) But are all adaptive evolutionary trajectories to be explained in this way? Is there any evidence, instead, to support the view that the occurrence of particular mutations in evolutionary trajectories can be crucial for those trajectories, especially when they are adaptive?

Darwin himself seemed to furnish some strong evidence in his work on orchids (first published in 1862 [63]). John Beatty [64] focuses on one of Darwin’s examples concerning the labellum, the lowest and usually the largest of the three orchid petals and the one that secretes nectar. In most species of orchids, a 180° twist in the ovary allows the labellum to serve its role as a landing pad for pollinators. Natural selection, according to Darwin, operated on the order of variations allowing for this result. In the species *Malaxis paludosa*, however, the ovary had twisted a full 360°, restoring the labellum to its uppermost (presumably ancestral) position. Darwin suggested that in some circumstances the uppermost position of the labellum might

be advantageous, possibly because of the nature of insect pollinators new to that species. But that natural selection in the case of *Malaxis paludosa* directed its evolution to a full 360° twist in the ovarium rather than to the original untwisted form was most likely due, Darwin suggested, to chance variations leading in that direction rather than the other.

Years later, based on the enormous work of geneticists on *Drosophila*, H.J. Muller [65] suggested that “a single decisive differentiating mutation” (p. 201) could result in non-crossability between sub-groups of a population, without any geographic isolation, and subsequently be augmented by “complementary” mutations resulting in “hybrid sterility and inviability” (p. 202).

The role of key mutations in evolutionary divergence would seem to be especially plausible in founder populations and population bottlenecks, where population numbers are close to what David Raup [66] has called the minimum viable population number, around “a few tens or hundreds of individuals” (p. 126). In such small populations drift can swiftly increase a rare allele to fixation, while an advantageous allele in a novel environment can be quickly favored by selection, especially if dominant (thereby bypassing the heterozygosity stage). Of possibly even greater importance would be key mutations in early simple forms of life with less developed DNA repair mechanisms, most notably species existing around the time of the Cambrian explosion. With these species, one or a few key mutations could result in divergence toward a new *bauplan*, and ultimately toward a new phylum.

Is there a way to test any of this? Observations of populations in the wild would be out of the question, given all of the variables involved and the needed time dimension. Theoretically, one could inject a novel advantageous mutant allele into a single individual in an experimental population and observe the results. Ideally, the experimental population would have to be compared with a control population (and the experiment would have to be repeated many times), with both populations and their environments being otherwise absolutely identical. But not only, of course, would the identity conditions be impossible to realize, and usually the time dimension required, but even if possible the experiment would arguably be unethical (much like introducing a new species into an environment).

31.6

Mutational Order

Fortunately, there is evidence that circumvents all of these problems, namely, evidence from computational simulations and from controlled studies of micro-organisms. While the Wallace type of view regarding pre-existent variation has dominated the theoretical landscape for many decades, evidence has been growing in favor of Darwin’s view.

An important contribution to computational simulation studies was provided by Mani and Clarke [67]. In various computational studies, they found that what they call “mutational order” can sometimes be more important in accounting for evolutionary divergence than random genetic drift. What often matters for divergence, they found,

is not simply the amount of existing variation, combined with drift and selection, but the order of appearance of new, advantageous mutations. They compared two initially identical populations, subject to the same forces of drift and selection, the only difference being that in one population the introduction of new mutations was ordered, whereas in the other population it was random. They found that when the initial population is far removed from its optimal state, mutational order proved to be more important for divergence than drift. This importance varied proportionally with population size and selection force; the greater either of these two factors, the more important the role of mutational order.

In an earlier paper, Clarke *et al.* [68] conducted similar simulations, though some without selection, only drift, and found that mutational order does not make a difference to evolutionary divergence when there is no selection. This led them to suggest that mutational order may be the main contributing factor in founder-effect speciation when the founder population is below its adaptive optimum and is exposed to strong selection pressures. It may not be different selection environments that alone account for divergent isolated races from a parental population, but also different genetic environments due to different sequences of mutations.

The importance of mutational order has been corroborated in recent years by studies on microorganisms. For example, in a long-term study on *Escherichia coli*, begun in 1988, Blount *et al.* [69] studied 12 populations from the same clone subjected to identical environmental conditions, involving a citrate-rich medium, which *E. coli* cannot use as an energy source. In one of the populations, after 31 500 generations, a citrate-using variant finally evolved, which had a weak capacity to use citrate. Over the next 2000 generations that capacity was further evolved, leading to dominance of the adaptation in the population and a massive population increase. That the appearance of this adaptation took an enormous amount of generations, and that it occurred in only one of the 12 populations, indicates that the adaptation was not easy to evolve. It further indicates, along with the fact that it took 2000 generations to further evolve, that the genetic basis of the adaptation involved rare historical sequences of mutations rather than an unusually rare mutation. Replays of the experiment, beginning with the weak citrate-positive variant, corroborated the original conclusion.

Blount and colleagues plan on furthering their research using whole-genome sequencing, so as to identify the actual mutations, especially the potentiating mutation or mutations, to determine their precise role (epistasis, hitchhiking, etc.). Given that whole-genome sequencing has already recently been done on experimental populations of *E. coli* in a glycerol-based medium (which is not as challenging to *E. coli* as a citrate-rich medium) so as to monitor the occurrence of beneficial mutations [70], what the research by Blount *et al.* [69] reveals is not only a vindication of what Mani and Clarke [67] claimed about the importance of mutational order, but also that it is still largely neglected as a stochastic process in evolution (since it is missing in similar research such as in Reference [70]). One can reasonably predict exciting new developments in this field.

All of the above research, of course, assumes that the origin of heritable variation, particularly of point mutations, is separate from the process of natural selection itself. Given the evidence for the foundation of point mutations in quantum indetermin-

ism, however, the assumption needs to be challenged, especially as an interesting extension of quantum biochemistry.

31.7

The Nature of Natural Selection

Time and again the claim is made, especially in philosophy of biology, that natural selection is a *deterministic* process. Elliott Sober [7], for example, as we have seen earlier, wrote that “In evolutionary theory, mutation and selection are treated as *deterministic* forces of evolution” (p. 110). Barbara Horan [9] claims that

“Evolutionary theory, as a general theory of evolution, a theory about one or another evolutionary force, or as a fitness theory – is deterministic. In this sense it is like classical statistical mechanics”

(pp. 93–94).

What motivates claims such as these is the belief that the generation of heritable variation is a separate process from natural selection, that natural selection simply feeds on heritable variation. The distinction goes back to Darwin. But as the concept of mutation has changed due to the connection with quantum indeterminism, it may well be that the principle of natural selection will have to change as well. Here I briefly argue that it should, with a fuller treatment of the topic to come in a later publication [71].

Darwin himself [2] defined “natural selection” as follows: “This preservation of favorable variations and the rejection of injurious variations, I call Natural Selection” (p. 81). This, however, was a terse definition, not his full meaning. What he argued throughout his book is that for natural selection to operate it requires heritable, profitable variations, superfecundity, competition or “struggle for existence,” and reproductive success. If all of these features are present, then over generations we get cumulative natural selection in its evolutionary sense.

The role of chance in all of this, for Darwin, was clearly epistemic only. An organism with a profitable variation in comparison with its conspecifics is not certain to out-survive and out-reproduce them, but it only has a “better chance” (p. 5). This is because the “conditions of life,” the relations of an organism with other organisms and with its environment as a whole, are “infinitely complex” (p. 61), such that there are many other factors that determine whether a particular organism reaches adulthood and reproduces.

As for the causes of particular heritable variations, the only meaning of “chance” that Darwin gave to them was in the sense of being random with regard to the environment. Following the evidence from the domestication of animals and plants, it was obvious to Darwin that variation was not directed (p. 30). Nevertheless, he thought that for each heritable variation “a cause for each must exist” (p. 170). As we have seen earlier, “chance,” he said, “is a wholly incorrect expression, but it serves to acknowledge plainly our ignorance of the cause of each particular variation” (p. 131).

In sum, then, heritable variation for Darwin provides “the materials for selection to work on” (p. 40) but it is not part of the process of natural selection itself. Darwin may have conceived of natural selection in this way because of the close analogy he drew between natural selection and artificial selection, since for the latter the origin of variation is clearly not part of the process of selection.

Whatever his reasons, when we turn to modern discussions on the nature of natural selection we find that the vast majority of biologists and philosophers of biology follow Darwin’s lead. Some prefer terse definitions, such as “the differential survival of entities” ([72], p. 33), “the differential rates of reproduction among different genotypes” ([73], p. 335), “the differential reproduction of genetically distinct individuals or genotypes within a population” ([49], p. 41), “a sampling process that discriminates, in particular, on the basis of *fitness* differences” ([74], p. 190), or simply “nonrandom differential reproductive success,” the consensus definition in population genetics. Some (usually philosophers) prefer an actual formalization of “natural selection,” such as Robert Brandon [75], who defines the “principle of natural selection” as follows: “If *a* is better adapted than *b* in environment *E*, then (probably) *a* will have greater reproductive success than *b* in *E*” (p. 11). And then some prefer to list the preconditions for natural selection to occur, such as John Endler [76], who lists “variation,” “fitness differences,” and “inheritance” as “necessary and sufficient for the process of natural selection to occur” (p. 4). Douglas Futuyma [77] basically follows this definition and adds his own terse redefinition: “the difference in rates of increase among biological entities that is *not* due to chance. *Natural selection is the antithesis of chance*” (p. 350).

And then there is Ernst Mayr, who says that “Natural selection is not deterministic, and therefore not absolutely predictive” ([78] p. 490). This is because, for Mayr,

“natural selection is a two-step process. At the first step, the production of genetic variability, accident, indeed, reigns supreme. However, the ordering of genetic variability by selection at the second step is anything but a chance process”
(pp. 519–520).

It is interesting to track Mayr’s thinking on this topic up through the strata of his writings. A few years later, after reading Sober [7] and writing a book review of it, Mayr changed his definition of “natural selection,” such that

“natural selection proper is only the second stage of a two-step process . . . yet selection would not be possible without the continuous restoration of variability”
([79], p. 98).

For roughly the next ten years, Mayr waffled between this definition and the former, between natural selection as a two-step process ([80], 88) and natural selection as the second step in that two-step process ([81], 188). At some point thereafter, however, Mayr’s two-step definition hardened, in the sense that he became adamant that “natural selection” is meaningless except as a two-step process. In an interview

of “the grand old man of evolution” conducted by Michael Shermer and Frank Sulloway [82], Mayr states that

“I consider the production of variation as part of natural selection. They are inseparable. Each is meaningless without the other. Natural selection is a two-step process: (1) variations produced, and (2) variations sorted, with the elimination of the less fit so that you end up with a ‘selection’ of the best”

(p. 81).

This remained his view to the end ([83], pp. 119–120, 281; [1], p. 136).

How is this difference to be settled, between the final view of Mayr on the one hand and the vast majority of biologists (including Darwin himself) and philosophers of biology on the other hand? It is absolutely crucial that this difference be settled if we are to understand natural selection properly.

One possibility is to view the difference simply as a matter of perspective. Non-molecular biologists (if I may use the term) and philosophers tend to look at the process of natural selection from the top down, which might explain why their definitions of “natural selection” are not two-step definitions. They tend to look at natural selection as a process that is “fed” by heritable variation but which does not include the origin of that variation. Peter Grant, for example, in his study of Darwin’s Finches [62], conceives of natural selection in the usual way, in the sense that “novel mutations are material for selection to act on” (p. 294). Geneticists, on the other hand, particularly molecular biologists, tend to look at evolutionary processes from the bottom up, beginning with the origin of the heritable variation itself. Masatoshi Nei [73], for example, argues that “mutation is the driving force of evolution at the molecular level” (p. 431), which he extends to the level of phenotypic evolution, while not denying the importance of natural selection. Some molecular biologists seem to go even further, seeing one unified process instead of two. Barry Glickman [84], for example, states that “Mutation is the driving force of evolution” (p. 47).

The solution, I suggest, is to go back to Darwin, where we find a *reference potential* with regard to the principle of natural selection, a fundamental ambiguity in his thinking that allows the principle of natural selection to take either of two different paths. In the *Origin* [2], Darwin often wrote of natural selection as a “power” (e.g., p. 61). But then Darwin ran into the problem of hypostatization, as many of his critics pointed out. “Natural selection” implies a selecting agent, on the analogy of artificial selection, but Darwin wanted to avoid that implication. Natural selection is a process, nature is not an agent. As much as two years before the publication of the *Origin* Darwin conceded, in a letter to Asa Gray, that

“I had not thought of your objection of my using the term ‘natural Selection’ as an agent; I use it much as a geologist does the word Denudation, for an agent, expressing the result of several combined actions”

([85], p. 492).

Following the publication of the *Origin*, and further criticism of his term “natural selection,” Darwin responded that he no more implied a person than physicists when they talk of the “Attraction of Gravity” ([86], p. 88). He also toyed with other names, most notably “natural preservation” (e.g., p. 243), but that didn’t seem right to him since the comparison with artificial selection would then be lost. The term “selection,” it should be added, was commonly used by animal and plant breeders since 1800 ([87], p. 213). So given the importance of “selection” and that “natural selection” had already been used in the *Origin* and had quickly become famous, Darwin’s readers would just have to get used to the term. But in 1866, in a letter from Alfred Russel Wallace, the term “survival of the fittest” was suggested to him ([60], p. 227). The term had been coined a couple of years earlier by Herbert Spencer and used by him in his writings when referring to Darwin’s principle of natural selection. Darwin thought it a “great objection,” however, that Spencer’s term “cannot be used as a substantive governing a verb” (p. 235). Nevertheless he started using “survival of the fittest” in his writings, beginning with the fifth edition of the *Origin*, in conjunction with “natural selection.”

So Darwin himself, despite many years earlier having rejected the idea that natural selection is an agent, continued to resist the idea that the verb “selection” does not require a noun. In so doing, he was perhaps guilty of falling victim to what the German philosopher Friedrich Nietzsche in 1888 called “the metaphysics of language,” part of which is the idea that every “doing” requires a “doer” ([88], p. 483). For Nietzsche, this is a fact of grammar, but to suppose that it is also fact of reality is simply a metaphysical presumption.

Today, molecular biologists especially seem cognizant of avoiding that error. Robert Haynes [46], for example, specifically on the topic of natural selection, states that

“selection is not a force of nature analogous to gravity or electromagnetism. Rather it arises as a consequence of the pre-existence of hereditary variation and reproductive excess within populations”

(p. 5).

But if we think of natural selection in this way, as a non-teleological process, as a process without an agent, as Darwin put it “the result of several combined actions,” then it would seem that we should, even must, include not only those actions that are normally called “selection pressures,” but also those actions involved in the production of heritable variation. For only both of those actions, combined with the particular histories of differential reproductive success, give us the process of natural selection in biology. The exclusion of the production of heritable variation in favor of the rest is the subtlest and yet still the most pervasive example of taking Darwin’s metaphor too seriously.

This point becomes strengthened once the principle of natural selection is combined with what Darwin intended it for, namely, evolution. It does no good to define “evolution” merely in the sense of the change in the frequencies of alleles or genotypes in a population, as population geneticists normally do (e.g., [21], p. 69; [89], p. xi), even as an adaptive change, since creationists would then clearly believe in evolution and creationism can then be taught in public school science classes.

Instead, “evolution” needs to be defined in terms of the origin of adaptations and of species ([90], p. 10; [78], p. 400). (It helps to know that in Darwin’s writings the distinguishing characteristic between any two species is at least one distinguishing adaptation. Hence for Darwin not only was natural selection the only process that produces adaptations in nature, but it was the only process that produces species; see Reference [91].)

If we think of evolution in this way, then Mayr’s two-step definition of “natural selection” begins to make more sense. It is to distinguish what is tersely defined as “natural selection” and what Dawkins [92] calls “cumulative natural selection,” the latter of which, he says, includes not just a sifting or sorting process but fairly faithful reproduction (where the offspring resemble their parents more than other members of the population), mutation, and repeated sifting or sorting generation after generation. Only cumulative natural selection, says Dawkins, not “single-step” selection, can give us “living organization” (p. 45).

Thinking of natural selection as Mayr’s two-step process repeated over and over again certainly goes against our intuitions, and for various reasons (historical, educational, and so on), but intuitions are not arguments, moreover they can be played against each other. Take the common phrase “evolution by natural selection.” If natural selection is considered as only the second step in Mayr’s two-step process, it follows that natural selection in itself is a deterministic process. The intuition then is to say that evolution by natural selection is a deterministic process. But if we accept that the variations upon which natural selection feeds are at bottom indeterministic, then we have a dissonance in our intuitions. We do not still want to say that evolution by natural selection is a deterministic process. It just doesn’t seem right. We want to say that it is a contingent process, more or less in line with Stephen Jay Gould’s [93] claim that “any replay of the tape [of life] would lead evolution down a pathway radically different from the road actually taken” (p. 51).

Here is a concrete example that helps to bring out the dissonance in intuitions, two statements from the same paragraph in Douglas Futuyma’s highly successful textbook on evolutionary biology [94]. At the beginning of the paragraph (p. 7) he states that (i) “natural selection . . . accounts for the diverse adaptations of organisms to different environments,” and at the end of the paragraph he states that (ii) natural selection is “purely mechanical as gravity.” Now, (ii) is true only if the production of variation is not considered as part of natural selection, given that point mutations have a basis in quantum indeterminism. But if the production of variation is excluded from natural selection, then (i) becomes false, since natural selection excluding the production of variation does not in itself account for adaptations. To have a full, truthful account of adaptations we have to include the nature of heritable variation, which at bottom happens to be indeterministic. In other words, given the basis of point mutations in quantum indeterminism, (i) and (ii) cannot both be true. We have to make a choice. If we want to keep (i) as it is, then we have to reject (ii). If we want to keep (ii) as it is, then we have to reject (i). The former is the more inclusive and less drastic choice.

Naturally, there will be much resistance to this choice. Part of it will simply be the result of historical inertia. Others will combine the inertia with reasons. For example,

Rosenberg ([14], p. 541), as we have seen earlier, and Millstein ([22], p. 684) after him, claim that even if point mutations have their foundation in quantum indeterminism, the probabilities in natural selection do not come from that source. The problem with this view is that it begs the question: it assumes that the principle of natural selection is a done deal, that it is not in need of revision. But it is if we view natural selection, going back to Darwin, as “the result of several combined actions.” The probabilities employed using natural selection as an explanation of adaptive evolution and of biodiversity, then, will have multiple sources, part of which will come from quantum indeterminism, while another part will come from our epistemic limitations.

A further source of resistance is the belief, or desire, that natural selection is domain neutral. Darwin himself thought that natural selection operates not just on heritable variation in organisms but on heritable variation in words in language evolution (e.g., [95], pp. 60–61). The same claim has been made today by several historical linguists (e.g., [96], pp. 376–381). The principle of natural selection has been applied to other domains as well, to natural selection operating on theories [97], on memes in cultural evolution ([72], ch. 11), and on baby universes ([98], ch. 7).

The temptation is to think of each of the above – biological, linguistic, theory, cultural, and cosmological evolution – as instances of what Dawkins [99] has called “Universal Darwinism,” even though Dawkins only meant to say that if there is life elsewhere in the universe he would expect it to be “Darwinian life” (p. 423), life evolved principally by natural selection. But is natural selection itself, not the concept or principle, truly domain neutral?

The distinctively human temptation, combined with the desideratum in science of unification, is to group similar things and processes under the same name, so that processes that *look* like natural selection are taken to *be* examples of natural selection. (The same is even more obviously true for the term “evolution”.) We may just as well, however, think of linguistic, theory, cultural, and cosmological evolution as constituting examples of evolution by *something like* natural selection, but not by natural selection *proper*. (A Tasmanian tiger, after all, is not a real tiger.) What is surely significant is that the recent advocates of domain-neutral natural selection can be found hedging their bets. Popper [100] writes of “a process *closely resembling* what Darwin called ‘natural selection’” (p. 261; italics mine), Dawkins [72] writes of “the *analogue* of natural selection” (p. 194; italics mine), Smolin [98] writes of “some mechanism *analogous* to natural selection” (p. 106; italics mine), while Lass [96] can be caught saying “the ontological bases may be the same or *similar*” (p. 113; italics mine).

One cannot have it both ways, with linguistic, theory, cultural, and cosmological evolution involving natural selection proper *and at the same time* “something like” natural selection in biology. The ambivalence indicates that at some mental level each of them was not quite sure that the process they called “natural selection” is truly identical across domains.

The problem is not merely verbal. Important facts are involved, facts that can make important differences. In biology, for example, the heritable variation in the process of natural selection involves high copying fidelity, but this is arguably not the case in cultural, linguistic, and cosmological evolution. Moreover, the latter domains, along

with the domain of theories, have nothing comparable to the genotype–phenotype distinction in biology. Further still, variation in biology is not directed; not only is it random with regard to the environment but it has a basis in quantum chance, whereas directed variation is normally the case in theory evolution and is often the case in cultural and linguistic evolution. And finally, one has to wonder if it can at all be meaningful to say that a universe has an environment. All of these are differences that make the nature of selection in these domains appear ontologically different from natural selection in the biological domain.

The differences in those domains make it all the more important to get the concept of natural selection as clear and right as possible in the biological domain. For, arguably, the processes of evolution in the linguistic, theory, and cultural domains are unquestionably deterministic, whether we include the generation of the heritable variations as part of the processes or not. With cosmological evolution, on the other hand, the generation of the variant universes is almost certainly indeterministic, as quantum fluctuations. Would we then want to say that cosmic evolution by “natural selection” is a *deterministic* process, if the source of the variation is irrelevant to the process? To say it is a deterministic process sounds extremely odd, to say the least, given its basis in quantum indeterminism. But the same problem arises for biological evolution by natural selection, as already pointed out.

Quantum biochemistry, then, adds a new dimension to the principle of natural selection, in line with Mayr’s two-step process. It also adds a new dimension to the very meaning of life.

31.8

The Meaning of Life

If quantum indeterminism is confined to the micro-level, the subatomic level, and in no way affects macro-level objects and processes, then the patterns and processes of evolution are fundamentally deterministic. And if they are fundamentally deterministic, then at any time the future of evolution is already set by antecedent causes. Evolution, then, would in a sense be an “unfolding,” with the future as set as the past. Humans would be a destiny, as well as every other species past, present, and future.

What the chapters in the present volume attest to is the fact that quantum indeterminism is not confined to the micro-level, so much so that quantum indeterminism must be taken into account to fully understand biological processes, including fields such as medicine. That point mutations, especially, have a basis in quantum indeterminism, combined with the role of mutational order, gives substance to the claim that biological evolution is contingent – really, truly, genuinely contingent. Whatever we take the meaning of life to be, whether of all life or of particular lives, it must now be viewed as fundamentally contingent.

Indeed it is the quantum world, and the quantum world alone, that justifies Gould’s [93] claim that “replaying life’s tape” (p. 48) would result in different outcomes every time. In particular, it justifies his claim that “*Homo sapiens* is an entity, not a tendency” (p. 320), that if we could replay life’s tape on Earth “the chance

becomes vanishingly small that anything like human intelligence would grace the replay" (p. 14). Gould's focus was not on mutations, however, but mainly on "chance" events such as asteroid impacts and introduced species, what he considered to be merely "improbable" events (p. 51) (improbable though in effect deterministic). What a focus on mutations adds to the picture, particularly an up-to-date one, is not merely chance or indeterminism in an epistemic sense (such as the lack of predictability in chaos theory), but *genuine* contingency. In fact, it adds a whole new meaning to the phrase "Darwinian life," far beyond what Dawkins [99] gave to it. It means not only that if there is life elsewhere in the universe it must have evolved principally by natural selection, but that the paths both of individual lives and of evolution are ontologically indeterministic.

Opposed to the latter view is the one that takes general trends in evolution seriously, argued perhaps most forcefully by Simon Conway Morris [101]. For Conway Morris, "convergence is ubiquitous" (p. 283), not just the repeated evolution of wings and eyes in Earth's history, but all sorts of traits. The significance is that "different routes will not prevent a convergence to similar ends" (p. 13), that if we could rerun the tape of life over and over again "the end result would be much the same" (p. 282), that "something like ourselves is an evolutionary inevitability" (p. xv) or "a near-inevitability" (p. 328), and that life has "metaphysical implications" (p. xv).

All of this is ripe pickings for a philosopher. For a start, it seems that convergence is at least partly in the eye of the beholder, akin to lumpers and splitters in classification. Where some might see convergence, others might see mere similarity. It all depends, of course, on how one defines "convergence."

But even with the same definition, some might still see convergence where others do not. This is because the theory-dependency of observation is not confined merely to definitions, and so it is free to override them. Conway Morris, for example, in his "Convergences Index," includes song among humans and birds. Defined as an elaborate vocal signal, clearly both humans and birds have song. But it is by no means obvious or established, nor even plausible, that song in birds and humans is a common evolutionary solution to a common adaptive problem. Rather it would seem simply an accidental, superficial resemblance.

More telling is Conway Morris's inclusion of intelligence and communication. For Conway Morris, "the emergence of human intelligence is a near-inevitability" (p. xii). He arrives at this conclusion by looking mainly at dolphins, as a different path that led to much the same end as in humans (pp. 256–257). The problem is that part of human intelligence is intimately connected with what Noam Chomsky called the "language organ," the essence of which is a complicated set of syntactic rules (or switches), a "universal grammar." Conway Morris accepts much too easily various claims that bonobos and dolphins and perhaps some other species have something similar as well. Contrary to what he wants to believe, the evidence strongly suggests that the linguistic abilities in these species is rudimentary at best, deserving to be called "protolanguage" rather than "language" ([102]; [103], ch. 14). Conway Morris, nevertheless, believes that "what we call language is an evolutionary inevitability" (p. 253). Were he not so strongly inclined toward the ubiquity of convergence, his argument would be better served by seeing in these other species examples of

convergence in protolanguage only, where the path to true language occurred either gradually by natural selection in a single small population of ancestral humans [104] or as the result of a fortuitous mutation (possibly chromosomal) in a single hominid line, an extremely rare event linking into one system previously nonlinguistic and protolinguistic functions ([105], pp. 165, 190; [106], p. 83). Either way, whether gradual or catastrophic, the conclusion remains that no other species we know of has anything remotely close to the complexity of human language, and, because of it, anything like the kind or kinds of intelligence that go with it. As such, Gould's rejoinder ([93], pp. 319–320) still stands, that a species like ours is extremely unlikely: the dinosaurs ruled the Earth for over 100 million years and evolved nothing like us, there was not even a trend toward bigger brains, and at any point in time during hominid evolution in Africa our species could easily have been preempted by the contingencies of extinction.

This leads us to the “metaphysical implications” of convergence. Conway Morris's book is ultimately a plea for theistic evolution. He is greatly impressed not only by the ubiquity of convergence, especially the repeated emergence of sentience, but by the extraordinary fine-tuning (“eerie perfection”) of both the genetic code and the cosmic constants of the universe. He quotes, highly favorably, theistic evolutionists such as Arthur Peacocke, Michael Polanyi, and John Greene, while he unashamedly bashes Richard Dawkins. He admits that his view of the universe neither proves nor requires a God, but that it is “congruent” with it and to think otherwise is to wear “dark glasses” (p. 330). In short, for Conway Morris “Either we are a cosmic accident, without either meaning or purpose, or alternatively . . .” (p. xiii, ellipsis his).

Recent work on the genetic code should seriously dampen any eerie feelings [107], while the argument based on the supposed fine-tuning of the cosmic constants fares no better ([98], ch. 7 and 15; [108], ch. 20). But what is especially odd about Conway Morris's view is that it takes genuine chance to preclude meaning: if we are a cosmic accident, then our lives have no meaning or purpose. This is odd because if traits such as sentience and intelligence can emerge out of non-sentience and non-intelligence respectively, and convergently so, then why not meaning also? In this way meaning does not have to come from without, from a maker or Maker, but can come from within. Conway Morris has taken the traditional view, that meaning can only come from without. But in so doing he has set up a false dichotomy. A further possibility is that meaning has emerged from non-meaning, and repeatedly so. Atheistic existentialists, with their claim that we are free to choose the meaning of our lives, would be representatives of this third alternative, but not the only ones.

There is yet a fourth alternative to Conway Morris's two, which combines a creator God with genuine ontological chance, a God that creates species through chance in evolution. The philosophers of religion Robert Russell [109] and Thomas Tracy [110], for example, have advocated for an interpretation of evolution wherein God guides evolution within the statistical laws of quantum physics, generating the desired mutations through the quantum physics of the hydrogen bond. This view is not altogether new. The Harvard botanist Asa Gray [111], for example, a friend of Darwin's and his main advocate in America, insisted that “variation has been led [by God] along certain beneficial lines” (p. 414).

Darwin's answer to Gray was swift and to the point, and it applies just as much to Gray's version of theistic evolution as to its modern counterparts. Gray's view, said Darwin, "seems to me to make natural selection entirely superfluous, and indeed takes whole case of appearance of new species out of the range of science" ([112], p. 226).

It should not go unnoticed that Darwin's concept of natural selection here (unlike elsewhere in his writings) would seem to include the *generation* of variation. If so, this is an important precedent for Mayr's view. At any rate, a further possibility for theistic evolution is to admit genuine chance in evolution, with God not guiding individual mutations, however, but simply setting up the whole system of the universe so as to allow, by genuine chance, the evolution of something like us or even higher. This is the view advocated by the Catholic cell biologist Kenneth Miller, in his book mistakenly titled *Finding Darwin's God* [113]. For Miller, evolution by genuine chance variation and natural selection was the "only way" (pp. 279, 291) that God could get what he wanted, namely, creatures with genuine free will, creatures that would make genuine moral choices, creatures that would come to understand and share in the glory of his creation, creatures that would worship and love him above all else (pp. 238–239).

Never mind that free will, contrary to Miller, is not entailed by quantum indeterminism, and that the natural sciences weigh heavily against the very possibility of free will ([114], pp. 86–87, 98). The main objection to Miller's view – and to the other views examined here and in fact to all theistic evolutionary views, including process theology – is a *moral* one, based on the evolution of sentience. This brings us back to Conway Morris. One can share his view (as I do) that sentience is not confined to one or a few species on Earth, that it is shared in various ways with a great many species, possibly even sprinkled throughout the universe on other planets. One can also share his view (as I do) that this has great moral implications. But one cannot then automatically think (as Conway Morris does but I do not) that this bodes well for theology. The problem is traditionally known as the problem of evil, and it is massively amplified if we take evolution seriously. In short, if there is a God whose telos – Conway Morris explicitly uses the term (p. 313) – behind creation via evolution is the creation of something like humans or beyond, then the mind-boggling enormity of the pain and suffering of untold trillions of creatures spread over hundreds of millions of years on this planet alone (and possibly others) – the fear and horror of predation, the agony of starvation, the torture of parasitism and disease, the brutality of temperature extremes – would require a God that is seemingly infinitely callous, and hence not worthy of belief. As Darwin put it much too mildly in his reply to Gray,

"There seems to me too much misery in the world. I cannot persuade myself that a beneficent and omnipotent God would have designedly created the *Ichneumonidæ* with the express intention of feeding within the living bodies of caterpillars, or that a cat should play with mice"

([86], p. 224).

One might reply by employing the concept of *kenosis*, that God empties himself into his creation partly by experiencing the suffering of his creatures along with them,

that is, from the inside. The problem of evil for theistic evolution remains, however, by rendering highly questionable, to say the least, whether the telos – morally good humanoids, God-worshipping souls, or whatever else – is really good, given the enormous cost.

Thomas Tracy [110] has argued that when it comes to animal suffering spread throughout evolutionary history – the “collateral damage,” as he puts it – we need to exercise what he calls “epistemic humility,” given our “epistemic finitude,” since we cannot possibly look at any particular example of suffering and determine that it was or is truly gratuitous, given the big picture of things. This is, of course, to miss the proverbial forest for the trees. Moreover, I should think it far more humble (epistemically or otherwise) to view ourselves, along with all living things, as the products, at bottom, of genuine raw chance, not mind, as parts of a system, a universe, with no goal or purpose. To Darwinize Shakespeare in *Macbeth*, “It [the four-billion-year pageantry of evolution on Earth] is a tale told by an idiot [natural selection including chance variation], full of sound and fury [biodiversity], signifying nothing [death and eternal extinction].”

For most people, this is a view much too bleak to ponder, let alone accept. For others, however, it is the only one that makes full sense of the world and all that science has uncovered, including especially the Uncaused Cause of the quantum.

Acknowledgment

I am deeply indebted to Cherif Matta, not only for the invitation to contribute this chapter, but also for his molecular graphs and his very helpful feedback on the manuscript. The mistakes that remain are entirely my own.

References

- 1 Mayr, E. (2004) *What Makes Biology Unique?: Considerations on the Autonomy of a Scientific Discipline*, Cambridge University Press, Cambridge.
- 2 Darwin, C. (1859) *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, John Murray, London.
- 3 Morgan, T.H. (1910) Chance or purpose in the origin and evolution of adaptation. *Science*, **31**, 201–210.
- 4 Rohrlich, F. (1983) Facing quantum mechanical reality. *Science*, **221**, 1251–1255.
- 5 Aczel, A.D. (2002) *Entanglement: The Greatest Mystery in Physics*, Raincoast Books, Vancouver.
- 6 Monod, J. (1971) *Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology*, (A. Wainhouse, trans.), Knopf, New York.
- 7 Sober, E. (1984) *The Nature of Selection: Evolutionary Theory in Philosophical Focus*, University of Chicago Press, Chicago.
- 8 Rosenberg, A. (1994) *Instrumental Biology, or The Disunity of Science*, University of Chicago Press, Chicago.
- 9 Horan, B.L. (1994) The statistical character of evolutionary theory. *Philos. Sci.*, **61**, 76–95.
- 10 Pullman, B. and Pullman, A. (1963) *Quantum Biochemistry*, John Wiley & Sons, New York.
- 11 Brandon, R.N. and Carson, S. (1996) The indeterministic character of evolutionary

- theory: no 'no hidden variables proof' but no room for determinism either. *Philos. Sci.*, **63**, 315–337.
- 12 Graves, L., Horan, B.L., and Rosenberg, A. (1999) Is indeterminism the source of the statistical character of evolutionary theory? *Philos. Sci.*, **66**, 140–157.
 - 13 Stamos, D.N. (2001) Quantum indeterminism and evolutionary biology. *Philos. Sci.*, **68**, 164–184.
 - 14 Rosenberg, A. (2001) Discussion note: indeterminism, probability and randomness in evolutionary theory. *Philos. Sci.*, **68**, 536–544.
 - 15 Bouchard, F. and Rosenberg, A. (2004) Fitness, probability and the principles of natural selection. *Br. J. Philos. Sci.*, **55**, 693–712.
 - 16 Rosenberg, A. (2006) *Darwinian Reductionism, Or, How to Stop Worrying and Love Molecular Biology*, University of Chicago Press, Chicago.
 - 17 Matthen, M. and Ariew, A. (2002) Two ways of thinking about fitness and natural selection. *J. Philos.*, **99**, 55–83.
 - 18 Walsh, D.M. (2007) The pomp of superfluous causes: the interpretation of evolutionary theory. *Philos. Sci.*, **74**, 281–303.
 - 19 Millstein, R.L. (2003) Interpretations of probability in evolutionary theory. *Philos. Sci.*, **70**, 1317–1328.
 - 20 Glymour, B. (2006) Wayward modeling: population genetics and natural selection. *Philos. Sci.*, **73**, 369–389.
 - 21 Hartl, D.L. (1988) *A Primer of Population Genetics*, 2nd edn, Sinauer, Sunderland, MA.
 - 22 Millstein, R.L. (2006) Discussion of 'four case studies on chance in evolution': philosophical themes and questions. *Philos. Sci.*, **73**, 678–687.
 - 23 Shanahan, T. (2003) The evolutionary indeterminism thesis. *BioScience*, **53**, 163–169.
 - 24 Schrödinger, E. (1944) *What is Life?*, Cambridge University Press, Cambridge (Reprinted 1992).
 - 25 Parthasarathi, R. *et al.* (2004) Bader's and reactivity descriptors' analysis of DNA base pairs. *J. Phys. Chem. A*, **108**, 3817–3828.
 - 26 Watson, J.D. and Crick, F.H. (1953) The Structure of DNA, Cold Spring Harbor Symposia on Quantitative Biology, **18**, pp. 123–131. Reprinted in G.S. Stent, (ed) (1980), *The Double Helix* W.W. Norton, New York, pp. 257–274.
 - 27 von Baeyer, H.C. (1992) *Taming the Atom: The Emergence of the Visible Microworld*, Random House, New York.
 - 28 von Borstel, R.C. (1994) Origins of spontaneous base substitutions. *Mutat. Res.*, **307**, 131–140.
 - 29 Hunter, W.N. and Brown, T. (1999) Single-crystal X-ray diffraction studies on the non-Watson–Crick base associations of mismatches, modified bases, and non-duplex oligonucleotide structures, in *Oxford Handbook of Nucleic Acid Structure* (ed. S. Neidle), Oxford University Press, Oxford, pp. 313–330.
 - 30 Moran, S. *et al.* (1997) A thymidine triphosphate shape analog lacking Watson–Crick pairing ability is replicated with high sequence selectivity. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 10506–10511.
 - 31 Goodman, M.F. (1999) On the wagon – DNA polymerase joins 'H-bonds anonymous'. *Nat. Biotechnol.*, **17**, 640–641.
 - 32 Florián, J. and Leszczyński, J. (1996) Spontaneous DNA mutations induced by proton transfer in the guanine · cytosine base pairs: an energetic perspective. *J. Am. Chem. Soc.*, **118**, 3010–3017.
 - 33 Fazakerley, G.V. *et al.* (1993) Base-pair induced shifts in the tautomeric equilibrium of a modified DNA base. *J. Mol. Biol.*, **230**, 6–10.
 - 34 Evans, T.A. and Seddon, K.R. (1997) Hydrogen bonding in DNA – a return to the status quo. *Chem. Commun.*, **21**, 2023–2024.
 - 35 Robinson, H. *et al.* (1998) 2'-Deoxyisoguanosine adopts more than one tautomer to form base pairs with thymidine observed by high-resolution crystal structure analysis. *Biochemistry*, **37**, 10897–10905.
 - 36 Suen, W. *et al.* (1999) Identification by UV resonance Raman spectroscopy of an imino tautomer of 5-hydroxy-2'-deoxycytidine, a powerful base analog transition mutagen with a much higher

- unfavored tautomer frequency than that of the natural residue 2'-deoxycytidine. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 4500–4505.
- 37 Harris, V.H. *et al.* (2003) Recognition of base-pairing by DNA polymerases during nucleotide incorporation: the properties of the mutagenic nucleotide dPTPaS. *J. Mol. Biol.*, **326** (5), 1389–1401.
 - 38 Guerra, C.F. *et al.* (2006) Adenine tautomers: relative stabilities, ionization energies, and mismatch with cytosine. *J. Phys. Chem. A*, **110**, 4012–4020.
 - 39 Triampo, D. *et al.* (2007) The stochastic model of non-equilibrium mutagen-induced alterations of DNA: implication to genetic instability in cancer. *BioSystems*, **90**, 870–880.
 - 40 Kwiatkowski, J.S. *et al.* (1986) Quantum-mechanical prediction of tautomeric equilibria. *Adv. Quantum. Chem.*, **18**, 85–130.
 - 41 Poltev, V.I. *et al.* (1996) Modeling DNA hydration: comparison of calculated and experimental hydration properties of nucleic acid bases. *J. Biomol. Struct. Dyn.*, **13** (4), 717–725.
 - 42 Doublié, S. and Ellenberger, T. (1998) The mechanism of action of T7 DNA polymerase. *Curr. Opin. Struct. Biol.*, **8**, 704–712.
 - 43 Monajjemi, M. *et al.* (2003) Metal-stabilized rare tautomers: N4 metalated cytosine ($M = Li^+, Na^+, K^+, Rb^+$, and Cs^+), theoretical views. *Appl. Organomet. Chem.*, **17**, 635–640.
 - 44 Wu, Y. *et al.* (2009) Theoretical studies on the bonding of Cd^{2+} to adenine and thymine: tautomeric equilibrium and metalation in base pairing. *Chem. Phys. Lett.*, **467**, 387–392.
 - 45 Löwdin, P.-O. (1965) Quantum genetics and the aperiodic solid: some aspects of the biological problems of heredity, mutations, aging, and tumors in view of the quantum theory of the DNA molecule. *Adv. Quantum. Chem.*, **2**, 213–360.
 - 46 Haynes, R.H. (1987) The 'purpose' of chance in light of the physical basis of evolution, in *Origin and Evolution of the Universe: Evidence for Design?* (ed. M. Robson), McGill-Queen's University Press, Kingston and Montreal, pp. 1–32.
 - 47 Gorb, L. *et al.* (2004) Double-proton transfer in adenine–thymine and guanine–cytosine base pairs. A post-Hartree–Fock ab initio study. *J. Am. Chem. Soc.*, **126**, 10119–10129.
 - 48 Kryachko, E. and Sabin, J.R. (2003) Quantum chemical study of the hydrogen-bonded patterns in A·T base pair of DNA: origins of tautomeric mispairs, base flipping, and Watson–Crick $\Rightarrow$ Hoogsteen conversion. *Int. J. Quantum. Chem.*, **91**, 695–710.
 - 49 Graur, D. and Li, W.-H. (2000) *Fundamentals of Molecular Evolution*, 2nd edn, Sinauer, Sunderland, MA.
 - 50 Löwdin, P.-O. (1969) Some aspects of the hydrogen bond in molecular biology. *Ann. NY Acad. Sci.*, **158**, 87–95.
 - 51 Martin, T.W. and Derewenda, Z.S. (1999) The name is bond – H bond. *Nat. Struct. Biol.*, **6** (5), 403–406.
 - 52 Sterpone, F. *et al.* (2008) Dissecting the hydrogen bond: a quantum Monte Carlo approach. *J. Chem. Theory Comput.*, **4**, 1428–1434.
 - 53 Pugliano, N. and Saykally, R.J. (1992) Measurement of the ν_8 intermolecular vibration of (D_2O_2) by tunable far infrared laser spectroscopy. *J. Chem. Phys.*, **96** (3), 1832–1839.
 - 54 Pugliano, N. and Saykally, R.J. (1992) Measurement of quantum tunneling between chiral isomers of the cyclic water trimer. *Science*, **257**, 1937–1940.
 - 55 Tuckerman, M.E. *et al.* (1997) On the quantum nature of the shared proton in hydrogen bonds. *Science*, **275**, 817–820.
 - 56 Winter, P.C. *et al.* (1998) *Instant Notes in Genetics*, Springer, New York.
 - 57 Chou, S.-H. and Reid, B.R. (1999) DNA mismatches in solution, in *Oxford Handbook of Nucleic Acid Structure* (ed. S. Neidle), Oxford University Press, Oxford, pp. 331–353.
 - 58 Drake, J.W. (1991) Spontaneous mutation. *Annu. Rev. Genet.*, **25**, 125–146.
 - 59 McFadden, J. and Al-Khalili, J. (1999) A quantum mechanical model of adaptive mutation. *BioSystems*, **50**, 203–211.
 - 60 Burkhardt, F. (ed.) (2004) *The Correspondence of Charles Darwin*, Volume

- 14, 1866, Cambridge University Press, Cambridge.
- 61 Williams, G.C. (1992) *Natural Selection: Domains, Levels, and Challenges*, Oxford University Press, Oxford.
- 62 Grant, P.R. (1986) *Ecology and Evolution of Darwin's Finches*, Princeton University Press, Princeton.
- 63 Darwin, C. (1862) *On the Various Contrivances by which British and Foreign Orchids are Fertilised by Insects, and the Good Effects of Intercrossing*, John Murray, London.
- 64 Beatty, J. (2006) Chance variation: Darwin on orchids. *Philos. Sci.*, **73**, 629–641.
- 65 Muller, H.J. (1940) Bearings of the 'Drosophila' work on systematics, in *The New Systematics* (ed. J. Huxley), Oxford University Press, Oxford, pp. 185–268.
- 66 Raup, D.M. (1991) *Extinction: Bad Genes or Bad Luck?*, W.W. Norton, New York.
- 67 Mani, G.S. and Clarke, B.C. (1990) Mutational order: a major stochastic process in evolution. *Proc. R. Soc. London B Biol. Sci.*, **240**, 29–37.
- 68 Clarke, B.C. *et al.* (1988) Frequency-dependent selection, metrical characters and molecular evolution. *Philos. Trans. R. Soc. London B Biol. Sci.*, **319** (1196), 631–640.
- 69 Blount, Z.D. *et al.* (2008) Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.*, **105** (23), 7899–7906.
- 70 Herring, C.D. *et al.* (2006) Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nat. Genet.*, **38**, 1406–1412.
- 71 Stamos, D.N. The New Higgledy-Piggledy: Natural Selection as Mayr's Two-Step Process (in preparation).
- 72 Dawkins, R. (1976) *The Selfish Gene*, Oxford University Press, Oxford.
- 73 Nei, M. (1987) *Molecular Evolutionary Genetics*, Columbia University Press, New York.
- 74 Beatty, J. (1984) Chance and natural selection. *Philos. Sci.*, **51**, 183–211.
- 75 Brandon, R.N. (1990) *Adaptation and Environment*, Princeton University Press, Princeton.
- 76 Endler, J.A. (1986) *Natural Selection in the Wild*, Princeton University Press, Princeton.
- 77 Futuyma, D. (1998) *Evolutionary Biology*, 3rd edn, Sinauer, Sunderland, MA.
- 78 Mayr, E. (1982) *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*, Harvard University Press, Cambridge.
- 79 Mayr, E. (1988) *Toward a New Philosophy of Biology: Observations of an Evolutionist*, Harvard University Press, Cambridge.
- 80 Mayr, E. (1991) *One Long Argument: Charles Darwin and the Genesis of Modern Evolutionary Theory*, Harvard University Press, Cambridge.
- 81 Mayr, E. (1997) *This Is Biology: The Science of the Living World*, Harvard University Press, Cambridge.
- 82 Shermer, M. and Sulloway, F.J. (2000) The grand old man of evolution: an interview with evolutionary biologist Ernst Mayr. *Skeptic*, **8** (1), 76–82.
- 83 Mayr, E. (2001) *What Evolution Is*, Basic Books, New York.
- 84 Glickman, B.W. (1987) The gene seemed as inaccessible as the materials of the galaxies, in *Origin and Evolution of the Universe: Evidence for Design?* (ed. J.M. Robson), McGill-Queen's University Press, Kingston and Montreal, pp. 33–57.
- 85 Burkhardt, F. and Smith, S. (eds) (1990) *The Correspondence of Charles Darwin, Volume 6, 1856–1857*, Cambridge University Press, Cambridge.
- 86 Burkhardt, F. (ed.) (1993) *The Correspondence of Charles Darwin, Volume 8, 1860*, Cambridge University Press, Cambridge.
- 87 Hodge, M.J.S. (1992) Natural selection: historical perspectives, in *Keywords in Evolutionary Biology* (eds E.F. Keller and E.A. Lloyd), Harvard University Press, Cambridge, pp. 212–219.
- 88 Kaufmann, W. (ed.) (1954) *The Portable Nietzsche*, Viking Penguin, New York.
- 89 Gillespie, J.H. (2004) *Population Genetics: A Concise Guide*, 2nd edn, Johns Hopkins University Press, Baltimore.
- 90 Thompson, P. (1989) *The Structure of Biological Theories*, State University of New York Press, Albany, NY.

- 91 Stamos, D.N. (2007) *Darwin and the Nature of Species*, State University of New York Press, Albany, NY.
- 92 Dawkins, R. (1986) *The Blind Watchmaker*, Longman, London.
- 93 Gould, S.J. (1989) *Wonderful Life: The Burgess Shale and the Nature of History*, Hutchinson Radius, London.
- 94 Futuyma, D. (1986) *Evolutionary Biology*, 2nd edn, Sinauer, Sunderland, MA.
- 95 Darwin, C. (1871) *The Descent of Man, and Selection in Relation to Sex*, vol. I, John Murray, London.
- 96 Lass, R. (1997) *Historical Linguistics and Language Change*, Cambridge University Press, Cambridge.
- 97 Popper, K.R. (1959) *The Logic of Scientific Discovery*, Hutchinson, London.
- 98 Smolin, L. (1997) *The Life of the Cosmos*, Oxford University Press, Oxford.
- 99 Dawkins, R. (1983) Universal darwinism, in *Evolution from Molecules to Men* (ed. D.S. Bendall), Cambridge University Press, Cambridge, pp. 403–425.
- 100 Popper, K.R. (1979) *Objective Knowledge: An Evolutionary Approach*, Oxford University Press, Oxford.
- 101 Conway Morris, S. (2003) *Life's Solution: Inevitable Humans in a Lonely Universe*, Cambridge University Press.
- 102 Kako, E. (1999) Elements of syntax in the systems of three language-trained animals. *Anim. Learn. Behav.*, 27 (1), 1–14.
- 103 Smith, N. (2002) *Language, Bananas & Bonobos: Linguistic Problems, Puzzles and Polemics*, Blackwell, Malden, MA.
- 104 Pinker, S. (1994) *The Language Instinct*, William Morrow, New York.
- 105 Bickerton, D. (1990) *Language & Species*, University of Chicago Press, Chicago.
- 106 Bickerton, D. (1995) *Language and Human Behavior*, University of Washington Press, Seattle.
- 107 Novozhilov, A.S. et al. (2007) Evolution of the genetic code: partial optimization of a random code for robustness to translation error in a rugged fitness landscape. *Biol. Direct*, 2, 24.
- 108 Weinberg, S. (2001) *Facing Up: Science and Its Cultural Adversaries*, Harvard University Press, Cambridge.
- 109 Russell, R.J. (1998) Special providence and genetic mutation: a new defense of theistic evolution, in *Evolutionary and Molecular Biology: Scientific Perspectives on Divine Action* (eds R.J. Russel et al.), Center for Theology and the Natural Sciences, Berkeley, pp. 191–223.
- 110 Tracy, T.F. (1998) Evolution, divine action, and the problem of evil, in *Evolutionary and Molecular Biology: Scientific Perspectives on Divine Action* (eds R.J. Russel et al.), Center for Theology and the Natural Sciences, Berkeley, pp. 511–530.
- 111 Gray, A. (1860) Darwin and his reviewers. *Atlantic Monthly*, 6, 406–425.
- 112 Burkhardt, F. (ed.) (1994) *The Correspondence of Charles Darwin, Volume 9, 1861*, Cambridge University Press, Cambridge.
- 113 Miller, K.R. (1999) *Finding Darwin's God: A Scientist's Search for Common Ground Between God and Evolution*, HarperCollins, New York.
- 114 Searle, J. (1984) *Minds, Brains and Science*, Harvard University Press, Cambridge.

32

Molecular Orbitals: Dispositions or Predictive Structures?*Jean-Pierre Llored and Michel Bitbol*

The study of structure and reactivity is of the utmost importance for chemists and biochemists. Biological properties of DNA, for instance, are a reflection of its space structure, which itself is the result of interactions between molecules. Photosynthesis and breathing depend on biological cycles that involve a very large variety of chemical species (oxidants, reducers, acids, bases, etc.) whose relative reactivity is determined by enzymes that optimize their spatial orientation and, in doing so, lift the reacting systems over the required energy thresholds.

Since the end of the 1950s, the explanation and the forecast of these biochemical transformations into terms of structure and reactivity have gradually turned to the methods of quantum chemistry. We explore the origins of this shift by identifying the assumptions and contexts used by biochemists. The choice of the prevailing approximation of molecular orbitals depends on techno-scientific contexts and financial constraints but also raises the problem of our representations of living matter, representations that structure our science of the life and the history of chemistry.

On the basis of these considerations, we focus on the importance of the concept of “reactivity” in our understanding of the molecular events. The different reactivity indices, be they static or dynamic, often depend on the underlying electronic structure method, often either molecular orbital self-consistent field (MO-SCF) approach or density functional theory (DFT). This choice of conceptual framework is not fortuitous – it expresses a holistic description of the relations between a whole and its parts that influences our representation of the biochemical transformations. A constructive philosophical dialog between a realistic approach relating to the *dispositions* that conceal the matter and a pragmatic transcendental approach that inquires about the possibility of this scientific knowledge, rather than focusing exclusively on its object, appears crucial to us to question our practices of research.

32.1

Origins of Quantum Models in Chemistry: The Composite and the Aggregate

At the end of the 1920s, many researchers were focused on creating a molecular model that accounts for the various experimental facts of spectroscopy and chemistry. A gap in knowledge existed in connecting the spectral terms of an *isosteric* series of molecules to those of the corresponding atoms. The spectral studies inquire about the possibility of indexing the molecular electronic states by quantum numbers describing the angular momentum, like the one used for atoms. R. T. Birge assumed that the energy levels associated with valence electrons of molecules agree in all essential aspects with those associated with the valence electrons of atoms [1].

Comparison of the spectra [2] suggests the possibility of distributing the first eight electrons of BO and CN around the two nuclei using two molecular orbits, while the ninth electron, more lightly bound, belongs to an orbit similar to that of the valence electron of sodium. Robert S. Mulliken compared his results with those of Sommerfeld [3] and Langmuir [4] for the N₂ and CO molecules. His assumption is that similar electronic distributions was corresponded to similar energy levels and vice versa. This assumption enabled him to foresee in 1925 the existence of bands whose reality would soon thereafter be confirmed by experiment [5]. However, for a given electron, the orbits in the atom necessarily do not correspond to the orbits described in the molecule insofar as several inversions and other divergences such as splitting appear in spectra. On the basis of work by Erwin Schrödinger, Pascual Jordan, Oyvind Burrau, and Max Born, Friedrich Hund [6] was able to describe, at the same time, the exact stationary states of two subsystems from those of the system that includes them. Thus the quantum state of a system can be obtained by superimposing the quantum states of its parts, a result made suitable for the study of the molecules by Hund. He suggested an interpolation between the quantum states of the separated atoms, those of the united atom (a fictitious atom obtained by the coalescence of the two atoms), and the molecule. This continuous and adiabatic quantum description of a system helped Mulliken to propose a systematic classification in which the molecular spectroscopic states and the relevant quantum numbers were rigorously identified [7, 8].

With regards to the concept of valence, considered as an intrinsic property of the atom, Mulliken has opposed the notion of “energy state” deduced from spectra on the basis of an *electronic configuration*, that is, of a distribution of the molecular electrons in different orbits. In this description, each orbit is delocalized over all the nuclei and can contribute, depending on each specific case, a stabilizing or destabilizing energy contribution to the total energy of the molecule. The sum of the energy contributions of each electron in its orbit determines whether the electronic configuration allows for the existence of a stable molecule, that is, whether its energy is stabilizing overall. For Mulliken, the atom no longer exists in a molecule. His concept of molecular state suggests molecular variability of energy and geometry that cannot be considered within the approaches of Gilbert N. Lewis and Langmuir. Mulliken proved that the spectral states of the molecules can be obtained from that of their molecular ions by the mere addition of an electron without changing the quantum numbers and,

thus, worked out his molecular *Aufbauprinzip*. This close connection between quantum theory and spectral studies gave birth to correlation diagrams in 1932 [9]. Such diagrams make it possible to consider the degree of likeness between a molecule and its separated atoms or its united atom thanks, in particular, to the empirical knowledge of the internuclear distances and of the charges of the nuclei. The molecule from then on has been considered as a composite [10], that is, a new entity rather than a mere aggregate of individualized atoms. Its state is described as an interpolation between known spectroscopic states of the constituent atoms.

The change from the concept of molecular *orbit* to that of molecular *orbital* occurred in 1932. The concept of orbital takes all its sense in Max Born's probabilistic interpretation that the square of a molecular orbital corresponds to the probability density of finding this electron in space. According to Mulliken, an atomic orbital is an orbital corresponding to the motion of an electron in the field of a single nucleus plus other electrons, while a molecular orbital corresponds to the motion of an electron in the field of two or more nuclei plus other electrons. Furthermore, both atomic and molecular orbital may be thought of as defined in accordance with the Hartree method of the self-consistent field, to allow so far as possible for the effects of other electrons than the one whose orbital is under consideration [11].

This holistic model of molecular orbital collides with a quantum model suggested by Walter Heitler and Fritz London in 1927 on the basis of the work of Werner Heisenberg and Paul Dirac. In their study of the H_2 molecule, these researchers associated the nature of the chemical bond with the energy of exchange. In this work, the molecule is a juxtaposition of atoms. From this standpoint, the role of the spin in chemical bonding quickly acquired a central importance. John Van Vleck went as far as considering that the spin lies at the heart of chemistry [12]. Anna Simoes and Kostas Gavroglu opined that this first paper related to the H_2 molecule opened a new era in chemistry [13]. Heisenberg was of the opinion that the theory of valence of Heitler and London has the advantage of leading naturally to the concept of valence used by chemists.

At the end of the 1920s and at the beginning of the 1930s, work of British mathematicians allowed the appearance of methods applied to the determination of the analytical form of the wave function. John Lennard-Jones, Douglas Rayner Hartree, Vladimir Fock lined up behind the holistic approach of the molecule. In Germany, Enrich Hückel popularized the approach of molecular orbital using energy assumptions relating to the electrons of the type " σ " and " π ." Linus Pauling and John Slater generalized and developed the work of Heitler and London and introduced, at the beginning of the 1930s, the *ad hoc* concept of hybridization. Pauling answered more directly the concerns of chemists by stressing the three-dimensional structure of molecules, the electrons being the *bonding officers* of the atoms. The valence bond approach that he developed with Slater was more quickly acknowledged by chemists because resonance corresponds to their usual representations.

The two rival theories evolved along partly different paths due to the mental representations of the chemists, representations that are central in the formation of new paradigms as Kuhn has shown [14]. The values and common examples shared by chemists explain their disapproval of the work of Mulliken and Hund which leads

to a dilution of the chemical bond on the whole of the molecule. This context is not unrelated to the refusal of *atomicity* by Pierre Duhem in his energetist work. Opposed to the equivalentists, Duhem preferred to rehabilitate Aristotle by admitting that the composite is really different from its parts, rather than to design the compound as a juxtaposition or an aggregate of atoms. The French historian of chemistry Bernadette Bensaude-Vincent points out that this dual interpretation of phenomena, that is, the *composite* versus the *aggregate*, is recurrent in the history of chemistry. For her, chemists were always confronted by this choice and, depending on the epoch, they chose one or the other interpretation or attempted to reconcile both. But the pluralism of possible interpretations does not cease living in chemistry [15].

These two visions of the molecule (that of Mulliken, and that of Heitler, London, Pauling and Slater) bring to the fore the historical opposition between the composite and the aggregate, that is, the classical question of the comparison of a whole with its parts in order to explain the emergent structures and properties. Before tackling this philosophical question, we suggest an attempt to understand the development of these models in the context of biochemistry where the question of the emergence of life is fundamental.

32.2

Evolution of the Quantum Approaches and Biology

During the 1930s, rival approaches quarreled over representations. Other factors relating to the personalities of the main protagonists influenced the path of the incipient approaches. Mulliken published in journals primarily read by physicists and, as a result, his work may long have remained little known by chemists, whereas Pauling appealed directly to the chemist's common sense. Trained as a crystallographer, Pauling also aroused biologists' interest by his studies of the structure of proteins. The extension of these approaches to the polyatomic molecules has been decisive in explaining the development in biology.

Mulliken studied polyatomic molecules from the standpoint of molecular spectroscopy and group theory. His irreducible representations for the 32 molecular symmetry groups were well-received in 1933 by the all theorists invited to Cambridge. He directed his work towards the explanation of molecular reactivity on the basis of the method of fragments and new correlation diagrams. He also developed the linear combination of atomic orbitals (LCAO) approximation in 1935 on the basis of Lennard-Jones' research, whereas Slater put forward his model of the intermediate molecular structure using the various possible electronic configurations. A conference organized in Paris in 1948 gave rise to up-to-the-minute debates about the nature of the chemical bond. In 1951, Mulliken organized the conference of Shelter Island, which paved the way to a new use of quantum mechanical methods in chemistry. This conference promoted an international agreement on the goals to be reached and the universal sharing of knowledge, with Mulliken's laboratory (The Laboratory of Molecular Structure and Spectroscopy, or LMSS) in Chicago playing a pivotal role. Developments of the approximation of molecular orbitals

picked up pace because of the generalization of the LCAO method to the study of charge-transfer complexes, developments in the calculation of exchange integrals by the method of the self-consistent field of Hartree–Fock, and concepts such as atomic electronic population, hyperconjugation, and so on. These applications of molecular orbitals approximation facilitated its progressive integration in various fields of chemistry in particular in organic chemistry.

Semiempirical methods became extensive, Clemens Roothaan and Bernard Ransil developed data-processing programs within the framework of projects of LMSS that improved the practical implementation of calculations while working in particular on the extent and the choice of the bases of wave function. Data-processing progress made possible complete Hartree–Fock type calculations without recourse to the experimental data whereas calculations related to the approach of Pauling and Slater essentially remained unrealizable. At the end of the 1950s, Mulliken used the *ab initio* approach to study the molecular reactivity of aromatic compounds, which certainly contributed to the rise of this method until the beginning of the 1980s. He developed an approach and a *style* of research that had a considerable influence on other contemporary scientists. Starting from the same line of reasoning of Mulliken, Kenichi Fukui was able to formulate the method of frontier orbitals, first proposed in 1952. Robert Woodward and Roald Hoffmann followed the path of Mulliken's work as well in their recourse to symmetry and correlation diagrams in the analysis of selection rules governing pericyclic reactions. One may conclude that a quantum paradigm is at work in fundamental chemical research, a paradigm brought to fruition primarily by Robert Mulliken.

How we can explain the acceptance in biology of a holistic approach similar to that of Mulliken with the global concept of molecular state after having specified the difficulties that it encountered in chemistry? How could an approach excluding valence spark the interest of biologists in the 1950s and 1960s?

Their interest can be justified partly by the difficulty in applying the approach of Pauling and Slater to biological molecules. The most important difficulty is that the one-electron functions describing the mesomeric forms are not orthogonal. So the matrix elements associated with the Hamiltonian operator in this particular basis appear as a very large number of terms of all the possible permutations of the electrons in the two antisymmetrized products. If the results of the two rival approaches (molecular orbitals and valence bond) converge for diatomic molecules, the valence bond method quickly proved to be unusable for larger molecules and remained impractical until recently as the power of computers has increased sufficiently to provoke renewed interest in the method. A techno-scientific limitation counterbalanced for a long time the intellectual embarrassment caused by dilution of the chemical bond over the whole molecule and supported the prevalence of the molecular orbitals approach. Recourse to a criterion of simplicity of calculations seems unavoidable in explaining the success of the molecular orbitals approach over its rival which is considered to be equivalent. Philosophers of science and scientists often call upon this criterion of simplicity to account for historical evolutions of science. Is it sufficient in this case? To answer this question, let us try to include the motivations of biologists of the time.

Alberte and Bernard Pullman showed that the work of many biochemists between 1949 and 1965 concerned the relation between the molecular orbital of π -systems and life (Reference [16], pp. 547–573). This orientation of research is hardly astonishing given the almost nonexistence of methods of studying σ -systems of sufficient size to interest chemists and biologists. These studies try to clarify the role of electronic delocalization on the structural and functional levels of the molecules of life (proteins, enzymes, coenzymes, steroids, vitamins, etc.) [17]. The role of intermolecular and intramolecular interactions was explored using the molecular orbitals approach that made it possible to explain the stability of pairs of complementary bases of DNA [18]. The delocalization is connected to the stability of these molecules; energies are calculated by semiempirical calculations within the framework of the self-consistent field. The stability of porphyrins has been connected with its selection and its persistence through evolution. Calculations [19] showed that adenine is the purine with the highest delocalization energy and were used to understand its prebiotic synthesis. The remote control of molecular information and the possibility of genetic changes are connected to the interactions between molecules and to the polarizability of the molecules with highly delocalized electrons. Semiempirical calculations allowed the determination of ionization potentials and electron affinities, which made it possible to account for the effectiveness of catalytic cycles. The electronic populations introduced by Mulliken and the recourse to frontier orbitals provided an explanation for the observed reactivity of molecules in the processes of life. In the same paper, A. Pullman and B. Pullman concluded that the catalytic activity of similar coenzymes is essentially related to the transformation of the initial coenzyme–substrate complex into a resonance stabilized and, at the same time, electronically activated intermediate, which is the essential form of the reaction. For them, only conjugated molecules, with their clouds of mobile and deformable electrons, that containing hetero-atoms capable of oscillating between different valence states can bring about such results. This statement expresses the holistic outlook of biochemical research. Far from being only justified by the data-processing limitation related to the use of the valence bond approach, one cannot but conclude that the molecular quantum holism meets the biological holism. Beyond the criterion of simplicity a holistic methodology expresses a change of perception of the link between a whole and its parts, a cell and its components, a body and its organs. A. Pullman and B. Pullman considered the choice and utilization of conjugated compounds as structural components of life as one of the most important quantum effects in biochemical evolution. Are we witnessing the emergence of a new quantum paradigm of research? This question remains to be studied more closely. Clearly, however, the choice of the energetist and holistic approach to molecular orbitals implies a conception of the process of life not only as a flow of information inside local or global networks but also as a process of generating form insofar as the relation between matter and its form depends on a system of interactions.

This new representation of the body can constitute in itself a historical rupture in our approach and our understanding of living matter. Why? Probably because it accelerates a major change, that is, the transition from the schematization of the

body to its in-depth cartography that anatomy has been achieving since the birth of medicine. The smooth body described superficially by mechanics and classical geometry becomes the singular body described in its intimate dimensions by an algorithmic and operational mathematics. The digitalization of the body is made possible by the use of quantum chemistry to express the details of structure and physical properties. The transition from an inanimate-material to a biomaterial that changes form while transmitting a genetic heritage is made possible by this change of mathematical models. The French philosopher Michel Serres points out that the moment mathematics change, a turn in World Vision and, consequently, in its images, starts [20]. Quantum methods allow the study, modeling, understanding and forecast of reactivity and molecular transformation. A quantum footbridge makes it possible to move from the shores of the geometry to those of topology and holistic algorithms. Why? Because the structure of quantum mechanics is first and foremost holistic and contextual. From a cartography resting on a separate analysis of the body's, muscles and bones to a consideration of their mechanical, chemical, or electric links, one passes a mindset where the focus is on an *interacting* dense matter-network. The purpose is not only to seek a classification by means of adequate invariants, that is, of objects connected by continuous deformations (homeomorphisms), such as Leonhard Euler's or Henri Poincaré's, but also and above all to think of the molecular event within a whole of interactions where matter and form maintain a dynamic relation. In his book "*The Interference*," Michel Serres shows that information flows in general and that its junction in pluralized nodes is a condition of choices, innovations, and inventions. In a world from now on foreseen like a superposition of networks, each one effective, constitutive and regulator in its own kind, the consideration of such lattices is the massive epistemological and gnoseological fact. He also highlights that the previous model of independent chains is substituted by a multiplicity of chains, connected each one to each one by the multiplicity of ways of a network in all the orders of life, of its praxis and knowledge [21].

This substitution and this type of research are already expressed in the work of the Pullman couple when they studied the quantum bases of the relation between biological function, natural selection, intermolecular structure and properties. The digitalization of their semiempirical calculations exceeded the purely geometrical vision to reach an understanding of dynamical reactivity and biological specificity. The meeting of these two holisms, quantum on the one hand and biological on the other hand, must now be analyzed more openly and placed in a context of interdisciplinary exchanges. The following are some routes for this "contextual handing-over":

- The algorithms Roothaan and Ransil developed for homologous series of diatomic molecules recall the step that Claude Shannon and Warren Weaver developed, as far back as the end of the 1940s, within the framework of information theory and which supplemented already, among others, those of Norbert Wiener and John von Neumann. In a nutshell, matter and information were described thanks to the same mathematical methods.

- In the 1930s, the study of the human genome leads to a holistic approach. The “adaptive landscape” model of Sewall Wright completely gave up the assumption of a pure and mere additivity of the contributions of genes and defends the total adaptability on the basis of statistical mechanics. According to Wright, the genes form an interacting system where some can retroact on the others.
- At the beginning of the 1940s, the mathematical modeling of the total behavior of the neuron is developed by William McCulloch and William Pitts, to go beyond electrochemical analyses.
- The concept of biological system in interaction took a new form related to quantum physics at the end of the 1940s when Max Delbrück, a collaborator of Niels Bohr, tried to explain genotypical and phenotypical transformations into terms of the molecular transitions from a stationary state towards another.
- The beginning of the 1950s is characterized by the growth of holistic work in biology. Alan Turing sought a mathematical model of morphogenesis while taking as a starting point the work of D’Arcy Thompson on the formation of the biological structures. Cybernetics implies that living and social systems are considered as data processing systems and humankind as a particular machine kind. The concept of landscape is developed especially thanks to Cecile DeWitt-Morette and her famous School of Houches, which brought together physicists and young biologists, but also with Conrad Hal Waddington who introduced his concept of epigenetic landscape in 1952.

The terms systems, networks and landscapes express a holistic approach to living matter and its transformations already active at the time when the choice between rival quantum approaches arose. In his series of books *The Cycle of Hermes*, Michel Serres shows at which point all sciences move in synergy since their Greek origin. Methods and concepts pass from one field of research to another and influence the way in which the world is represented in thought. This leads us to maintain the insufficiency of recourse not only to a criterion of simplicity but also to a purely techno-scientific reason to explain the reception by biologists of the models of quantum chemistry. A wider paradigm of research is at work and underlies an interpretation of the matter considered as a problem to be solved in the form of algorithms. We are able then to grasp at which point a holistic and energetist approach amounts to a possible bridge between quantum chemistry and biology.

A vigorous exchange of concepts has been stimulating research in biochemistry since the end of the 1940s. The approach of complexity and the concepts of percolation and plasticity [22] currently cross all sciences and seek a way of bridging algorithms with empirical fields. If philosophy consists, as underscored by Jacques Bouveresse, not only of a clear conceptualization of the nature of the means that we use to think of reality but also as a better understanding of the manner in which the World is represented in the thought [23], a thorough study must be continued to question the bonds between holism, energetism, information and matter as from the 1940s, while defying hasty analogies [24]. In a collective work edited by Isabelle Stengers, entitled “*From one Science to Another: Wandering Concepts*,” she suggested distinguishing between two modes of propagation of concepts:

- The first is achieved through diffusion: here, the disciplinary origin of the concept is recognized, and we are in the case of an openly metaphorical use.
- The second evolves as an epidemic: the source of the concept is forgotten, it is presented as “pure,” cut off from the natural language, being defined from the formalism of the science that it helps organize [25].

The transition from the quantum concepts of chemistry to biology is undoubtedly done on the second mode. It is based on a mathematical expression of the questions on the matter where the symbols of an algorithm replace the image of an aggregation of mass particles [26]. The capacity of extension and organization of quantum concepts comes consequently from the primarily holistic nature of their formalism. At the end of the 1940s, a composite knowledge of the matter, implying several levels of analysis and several disciplines, appeared and sought to be stabilized despite prevailing reductionist approaches. Gilles Deleuze and Felix Guattari gave an account of this composite knowledge by putting forward the thesis of a mobility inherent to the concept that joins together the pieces or the components that come from other concepts, which answered other problems and supposed other plans. According to them, a concept does not require only one problem under which it alters or replaces preceding concepts, but a crossroads of problems where it is combined with other coexistent concepts [27]. The crossroads of problems connecting quantum chemistry and biology at the end of the 1940s connected temporarily the topics of energy, holism, matter, finality, life and sciences of evolution. This point of convergence allowed a new expression of questions that have crossed our culture since the Greeks and which focus on the definition of life as well as the relation between matter and form. We move from an *ontology of entities* to an *ontology of capacities to change* that the quantum formalism, discussed in the second part of this chapter, enables us to think of in terms of a new translation of probabilities.

The methods evolved (CNDO, NDDO, INDO, new versions of SCF, etc.), the algorithms were refined, energy became a functional of the electronic density and not solely of molecular orbitals, the theoretical framework to go beyond the Born–Oppenheimer approximation became available but these methods are finer and contextualized forms of the same holistic and energetist description of the matter. The choice of method whether *ab initio*, semiempirical, DFT, or post Hartree–Fock depends on the aims and the means invested in a given research project within a particular laboratory. The semiempirical methods have a substantial number of advantages that make them the tool of choice when the evolution of a physical or chemical property in a family of compounds is being questioned. In addition, the many calculations carried out since their appearance made it possible to better determine the weaknesses of each method and, therefore, to delimit their fields of application [28].

However, as mentioned above, the valence bond method has also become the subject of renewed interest, due perhaps to advances in the power of computers and data processing. This method can have important advantages over the molecular orbital approach since it can zoom in on an interesting molecular area (e.g., an active site of a protein) or replace configurations by relevant mesomeric forms. It is

necessary to be careful with the use of this aggregative approach in chemistry and biochemistry because its conceptual treatment can compromise molecular holism. We now suggest going back over the concepts of molecular topology in the light of quantum mechanics and opening a philosophical dialogue that inquires about our practices of research aimed at catching molecular events and their specificities computationally.

32.3

Philosophical Implications of Molecular Quantum Holism: Dispositions and Predictive Structures

32.3.1

Molecular Landscapes and Process

How can we recognize reality? In the unique proliferation of its details. The transition from the geometrical diagram and the old physics to a topology founded on computational algorithms permits a grasp of the uniqueness of events and that of molecular processes. These calculations are approximations addressing problems formulated and described by quantum methods giving access to the details of a structure. The numerical outcomes allow researchers to foresee biochemical reactivity. The landscapes that appear on computer screens of biologists, chemists, and biochemists are full of details that express interactions in selected isolated networks belonging to the total complexity of the biological molecular environment. Hydrogen bonds that frame the structures of proteins, the quantum study of a state of transition explaining the creation of chemical bonds inside the ribosome, and the relation between ligands and central metal are examples of energy landscapes based mainly on the molecular orbital approximation. But are these networks projected on the things of the world, or simply extracted from them? Are these networks constitutive of the things of the world or only enable us to know them? What does the general information of this interlaced net mean?

Michel Serres widens this question by taking a necessary distance from the whole of the outcomes of contemporary science. According to him, this entangled net is almost invariant not only from the local to the global or from the gigantic universe to the smallest particles of matter or place of life, but also from reality to our manners of knowing, from exchange of signs to our collective practices. He wonders if we can consider it as the skeleton of any system referring to Leibniz's philosophy [29]. In the same way, Pierre Levy wonders whether the descriptions of phenomena in terms of calculation or data processing raise a mere analogy, an heuristic metaphor, a model for prediction prevented from any ontological claim, or on the contrary if the identification of the objects highlights their true colors by means of our scientific machines [30].

We will try to bring concise replies to these questions by analyzing the role of the molecular orbital approximation in contemporary biochemistry. A possible starting point is a reflection on the daily practice of the quantum methods in the laboratory.

A close relation between theory, experiment and algorithms makes it possible for researchers to study the matter using fine techniques such as fluorescence or femtosecond lasers, the Aharonov–Bohm effect, nanotechnology, or new approximations allowing calculations of some integrals of energy essential to the study of the molecular reactivity. The crossroads of the concepts corresponds to a crossroads of the techniques. Ilya Prigogine and Isabelle Stengers consider that we have reached the point that constitutes the singularity of modern science: the meeting between the technique and the theory, the systematic alliance between the goal to model the world and that of including it [31]. In the same way, we can write that time is no more the almost exclusive result of deduction (being subjected less and less to experimental testing) but of the elaboration of models by practices that remain techno-scientific from beginning to end, and leading to the reconstruction of the details of sectors of reality as a function of contexts.

How do researchers use the quantum approximations? What are their purposes when they refer to such methods? What do these approximations teach us and what are the limits of this knowledge?

The model of the transition state has stimulated several fine experiments on small molecules in the form of generally homo-kinetic molecular beams whose molecules are brought either to collide with those of another molecular beam or excited vibrationally or electronically by irradiation at a selected frequency. These studies showed the need to take into account the coupling of electronic and nuclear movements, in other words, to go beyond the Born–Oppenheimer approximation. It is possible to specify the nature and the electronic state of the products according to the experimental conditions [32]. Biochemical reactions, however, involve macromolecules at a temperature where, generally, electronic transitions are not induced. It is thus generally admitted, on empirical and theoretical bases, that the point representing the system formed by one or more molecules that are poised to react moves on a surface of potential energy corresponding to a single electronic state (adiabatic surface) and obtained for all possible geometries of this system within the Born–Oppenheimer approximation. A statistical treatment taking into account all the possible initial arrangements makes it possible to predict the nature and the relative proportion of the various products. Within this experimental framework, only a limited number of possibilities are weighted by a probability that is significant enough so that the chemical act can take place on the time scale of the measurement.

How to foresee and understand a chemical reaction in this context of practices of research? A “tracking” of the most probable possibilities proceeds by identifying the reaction pathway that leads to the transition state postulated by the macroscopic kinetic theory. A continuous exchange between quantum models of reactivity and the increasingly refined instrumental techniques, experimental culture, and the macroscopic approach is the beating heart of research. Hence, a footbridge is required between two levels of description: the microscopic level and the kinetic laws on our scale. A prevalent geometrical language, that of the “reaction coordinate,” is used to describe possible reaction pathways. One can then study regions of the potential energy surface that contain the transition state and, at the same time, regions that

contain the least energetic path(s) leading to this transition state. These regions can be explored by means of energy calculations on the supermolecule formed by the reactive species. This holistic energy description enables the study of chemical or biochemical reaction dynamics. However, to understand the reactive properties of a system, a chemist generally seeks criteria that can be used to predict the existence and the geometry of transition states and those of the products given those of the reactants. This is the goal of molecular theories of reactivity.

The static approach to reactivity uses static indexes [33] that are electronic descriptors of an insulated molecule in an equilibrium state. Static indices include, for instance, the atomic charges for the prediction of ionic or dipolar reactions or the free valence index for radical reactions. Several researchers have analyzed the creation of the transition state in stages. They first use as a starting point the wavefunctions of the two separated reagents before tracking their interaction by means of quantum approximations. The method of intermolecular orbital developed by Salem [34, 35] analyzing the reactivity of π -systems is one of most accurate by which to study the various effects that occur in short-range interactions. Successive approximations justified by their context of use make it possible to include the behavior of the reagents according to various types of control (steric, electric, frontier orbital) expressed by the Klopman–Salem equation to determine the energy of interaction.

Perturbational approaches and the variation principle are at the basis of these approximations. Intermolecular energy is considered to be a perturbation of the initial energy of the reagents. How can we explain this recourse of chemists to the perturbation methods often used in other fields of quantum physics as in the quantum theory of the fields? Generally chemists are interested in small energy differences (e.g., energy of activation, enthalpy of reaction, etc.), which is why the perturbation approach is relevant. Within this framework, Fukui's frontier orbital approximation makes it possible to tackle five questions of reactivity: absolute reactivity, relative reactivity, regioselectivity, stereochemistry of the reaction, and stability of the primary product of reaction; and also three questions of structure: structural stability, structural reactivity, and the forecast of the structural anomalies [36]. The Woodward–Hoffmann rules of the conservation of orbital symmetry then make it possible to explain and predict pericyclic reactions [37, 38].

In parallel, the dynamic approach to reactivity uses indexes that characterize the response of a molecule to the approach of a reagent. An example of dynamic index of reactivity is the π -electrons localization energy on an atom in a molecule. These π -electrons can form addition intermediate complexes with an electrophilic reagent, complexes that subsequently break-up to produce substituted compounds. Clearly, just as one cannot describe the creation of a chemical bond between two atoms correctly by considering each reagent independently of the other, energy and electronic characteristics of transition states cannot be completely understood without considering the chemical structure in its totality. In particular, a transition state can result from the transfer of charge from one reagent to another so that any separate treatment of the two reagents would lead to erroneous conclusions.

Necessary to contextualize the study within that of the reaction pathway. (i) the principle of Bell–Evans–Polanyi emphasizes the effects of polarizability of the reagents in concerted reactions and (ii) the Hammond postulate that the structure of a transition state is similar to that of the species closest in free energy (or, to think in terms of potential energy surfaces, of the topologically closest species). Topology takes over geometry and makes possible a holistic understanding of a chemical reaction. The Hammond postulate links the rate of a reaction with the structural characteristics of the reacting species by specifying that the molecular reorganizations must be small at the time of the transformations, implying minimal energy changes and/or differences. This allows structural comparisons between reagents, products, and possible “intermediate states,” explaining why the thermodynamically most stable product is not always favored during the reaction process.

The methods of quantum chemistry make it possible to describe a molecule in energy terms and to analyze its transformations. Reagents and chemical events are studied in a precise context of unceasing exchanges between empirical techniques, data, algorithms, and theories. Isabelle Stengers points out that there are compromises between the way in which the “laws” raise the problem and the manner in which the experimental data are required to solve it, between the theory and “phenomenology” [39].

We are witnessing a transformation from a mechanistic and reductionist view of chemistry that gives way to an explosive expansion of structural information, reactive pathways, and active sites, that is, with landscapes that unite the local and global scales at once. The mathematics carrying this description of reality resists the univocity of traditional reasoning and weaves relations, region by region, by studying the *molecular whole* without making any assumption on its nature. We are leaving an ontological approach of scientific facts. Michel Serres studies this new mathematics that underlies scientific researches. For him, this operational mathematics can be defined as pure, abstracted and practical at the same time. Describing the landscapes, it produces the charts and that is why the image and the diagram are no longer contradictory [40]. Pure reason and great principles are often relegated backstage in favor of a more pragmatic approach. What do these practices of chemists and biochemists teach us? Certainly that molecular reagents that they use and events that they detect are a compromise solution resting on the unceasingly negotiated nature, prolific but conditional, of the alliance between the Hamiltonian formalism and the experimental practice [41].

A question seems essential at this point of our work, it relates to the meaning of molecular orbitals. Does a molecular orbital belong to the molecule or does it only reflect the knowledge that we have about the molecule? These mono-electronic wavefunctions potentially contain information not only on the structural details but also on the molecular capacities to react and to change in experiment. Can we assert that molecular orbitals are *things* of the world? A dialog is necessary between ontology and epistemology to inquire about this form of knowledge, its origins, its status, its range, and its role in contemporary practices of research and the inter-subjective agreements that they suppose and cause.

32.3.2

Realism of Disposition and Predictive Structures

In his *Philosophy of History*, Hegel wrote: “What is in itself, it is a reality, a capacity, but which, from inside, did not arrive yet at existence” [42]. While speaking about the concept, he added:

“The movement of the concept is development, by which development one considers only what is already present in itself. In Nature, it is the organic life which corresponds to the stage of the concept. Thus, for example, the plant develops from its seed. The latter already contains in itself the whole plant, but in an ideal way, and this is why one should not consider its development in such a way that the various parts of the plant, root, trunks, leaves and so on, exist already in the seed but all very small. It is there the assumption that things fit into each other, whose absence consequently implies that what is not present except in an ideal manner is considered as already existing. What remains correct, to the contrary, in this hypothesis is the fact that the concept, during its process, remains within itself and that through this process nothing new is added with regard to the contents, but that only a change of the form is produced.”

Does the seed contain the preformed tree? We could not assert it. Does it contain the power to produce the tree? A whole tradition of thought would undoubtedly affirm it. The latinists and the philologists are best at distinguishing the fundamental difference between a mere possibility and the power to do. The question of dispositions and causal powers has always influenced the work of philosophers since Aristotle’s *potentia*. This question also crosses the history of chemistry in particular through the debate, always open, between the composite and the aggregate evoked in the first part of this chapter: the aggregate locks-up the parts of a compound in action, while the composite incorporates them in power (insofar as only an analysis would reveal them). The main concern of this chapter is our understanding of the novelty in the world, that is, in more contemporary terms, our explanation of the emergent phenomena. Do hydrogen and oxygen contain water “in power?” In his book *De l’explication en sciences* (Explanations in Sciences), Emile Meyerson explained the recourse to the concept of state of power as a condition of “hypostasis” (or entities of thought not directly perceptible) allowing several types of scientific explanations:

“no hypostasis is conceivable, if we do not admit the possibility of an existence in power”

[43]. He added:

“The thinker is prey to two opposite tendencies as soon as he has recourse to this concept of state of power: It is necessary that the potential is distinguished from the actual and it is necessary however that in spite of this distinction it must give rise

to him, which is only possible if potential and actual are identical, if potential is tantamount to actual. It is necessary thus that the thought considers simultaneously these states as similar and as different and that it reconciles or at least seems to reconcile this contradiction. It is to this end that the ingeniousness of the philosophers applied untiringly" [44]. For Meyerson, the recourse to the state of power is an

"unavoidable condition of the process of our reason,"

in his previous book *Identité et réalité* (Identity and Reality) he had already asserted:

"Does science proceed differently when it considers the kinetic energy and the potential energy and even fundamentally different energies, such as heat, electricity, the mechanical energy, as 'forms' of only one and sole fundamental essence (we will say while using of the term of Mnésarque, as modalities being able to transform one into the other without ceasing preserving itself)? How consequently would the philosophy, whose main task consists in the effort tending to compose a coherent image of the Great Whole, escape this inescapable need?"

[45]. This analysis is integrated in a geometrical and essentialist vision of the world with which quantum mechanics is at odds. The state of power is nevertheless a cornerstone of different established interpretations of quantum mechanics, and seems to be even more natural within the framework of this theory. How did this *wandering concept* spread and developed in such a quantum background?

In chemistry and quantum biochemistry, it is not only dispositions or potentialities that get involved, but also their fitting into each other. We can thus ask whether the dispositions or the potentialities of a molecule are translatable in terms of wavefunction (expressing itself potentialities) or in terms of reaction landscapes that result from it. Indeed, indexes of both static and dynamic reactivity indicate to us the possible molecular regions of interaction between two molecules and, thus, the regioselectivity and stereoselectivity of their reaction. We can consider the products of a reaction, their proportions, the physicochemical properties of a molecular complex, or the dynamics of an electrocyclic reaction by jointly using the molecular wavefunctions that would contain them "in power" and some elements of molecular symmetry. We must think about this situation independently of questions about the bases and relevance of this reading of the molecular wavefunction. It is necessary for us to highlight not only the Hilbertian quantum approach but also the practices of compromise that we have described if we want to address the difficult question of quantum realism.

The well-known *measurement problem* plays an important part in the expansion of the concept of power in the fertile ground of quantum mechanics. In the same way as David Albert did with respect to the property represented by the observable Q [46], we can inquire about the meaning for a quantum system to be in the state Ψ that is not an eigenstate of the Hermitian operator corresponding to Q . Several answers

consider that the system has a capacity, a disposition, a propensity to express a particular value of Q when measurement intervenes on its state Ψ . What is the meaning of these assertions? Mauro Dorato aims to study how quantum mechanics are connected with philosophical issues about dispositions [47]. His study stresses the key role of the concept of dispositions whatever the interpretation of quantum mechanics. He refers to the work of Robert Clifton and Constantine Pagonis bearing on an interpretation in terms of dispositions of the properties introduced by Bohm's theory, but he is astonished by the scarcity of philosophical work on this subject while at the same time Werner Heisenberg, one of the "pilgrim fathers" of quantum physics, emphasized that the statistical algorithm of quantum theory contains statements on possibilities, or better tendencies – *potentia* in Aristotle's philosophy – and that such statements are completely objective insofar as they do not depend on any observer. For Heisenberg, the discontinuous change in the probability function, however, takes place with the act of registration, because it is the discontinuous change of our knowledge in the instant of registration that has its image in the discontinuous change of the probability function [48]. Mauricio Suárez notices that the work of Heisenberg consists in connecting the quantum probabilities to this concept of power [49]. How is this connection tied?

Henry Margenau, who is opposed to Bohr's interpretation of quantum mechanics as well as to that of Bohm, evokes a *latency* lying in quantum descriptions that is not expressed in traditional physics [50]. Nicholas Maxwell connects probability and propensity thanks to the quantum formalism. This connection is based on a reflection of the nondeterministic nature of quantum mechanics. Along the same lines, and referring to the Ionian cosmologists, Karl Popper pointed out that the metaphysical program that propensity interpretation suggests could be summarized by the formula "all is propensity." Using Aristotle's terminology, Popper concluded that "to be" is at the same time to be the actualization of a former propensity to become, and to be a propensity to become [51]. Thereafter, Popper went into more details, he explained that in quantum physics, propensities are not those of isolated objects, but of the whole context of experimentation in which they are inserted. He insisted on this situational aspect of the concept of propensity because he considered it decisive for a realistic interpretation of quantum mechanics. For Popper, propensities can act as well as the forces of gravitation, *they are effective, they are real* (italicized in the text of Popper). In doing so, he allotted a kind of reality to pure possibilities, to indexed possibilities [52]. This close connection between potentiality and contextuality is typical of quantum physics. But how can one pass from this potentiality to reality? Michael Readhead analyzed three prevalent interpretations of quantum mechanics and showed that, in an interpretation presupposing propensities and potentialities as real, it is the measurement that has the function of converting "latent" values into actual values [53]. From there, many apparent paradoxes of quantum mechanics vanish. Thus, according to Dorato, it is enough to combine the concepts of contextuality and complementarity with that of potentiality to lift the apparent paradox of the wave–particle duality without necessarily giving up a realistic reading of quantum physics. Indeed, Dorato shows that while antirealist philosophers consider that there is no way to find out which of the two slits

the particles has gone through, it can be asserted in a realistic way that the disposition to manifest the particle-like nature of the system got lost [54]. A topical question is to associate probability with the state of power whereas apart from quantum physics power determines its actualization if all the necessary conditions are met. How is this connection between probability and potentiality achieved? To some extent, this connection relies on two main approaches independently of the underlying interpretation of quantum mechanics.

The first approach seeks an increasingly tighter association between contextuality and indeterminism, crystallized, for example, in the concept of “selective propensities” suggested by Suárez, who states that the crucial difference between a propensity and a sure-fire disposition is that, under the appropriate circumstances, the manifestation of a sure-fire disposition is necessary – deterministic propensity –, while the manifestation of a propensity might only be probable [55]. Suárez agrees with Popper when he concludes that potentialities are inherent of an experimental situation rather than of an object.

Others, such as Christian de Ronde, understand the wavefunction in terms of its faculty or capacity to be interacting. They start from an analysis of the operating modes of the traditional scientific thought and identify three of its fundamental pillars: (i) the principle of existence (ontology), (ii) the identity principle (separability and unicity) that we have already underlined through the analysis of Meyerson, and (iii) the logical principle of non-contradiction. The Hilbertian structure of quantum mechanics is then carefully studied and the key role of the choice of the bases of projection is underlined. Using the theorem of Kochen and Specker, de Ronde concludes that a sole state vector cannot support the existence of its different representations simultaneously, and that is why it is not possible to consider Ψ in terms of something that refers to an entity [56].

Indeterminism, complementarity, and superposition supersede the principles of existence, noncontradiction, and identity, respectively. This new logical and probabilistic (non-Kolmogorovian) context paves the way for a reflection on the modalities. De Ronde proves that, since Aristotle, the combined use of the three principles of traditional physics tends to induce an identification of “actual” with “reality.” He suggests introducing the concept of *ontological potentiality* (p. 206 of Reference [57]). For him, the mode of being of a faculty is potentiality, not thought in terms of possibility – which relies on actuality – but rather in terms of ontological potentiality, as a *mode of existence*. He specifies (p. 207 of Reference [57]) the meaning of the faculty of raising his hand. It does not mean that he will raise his hand or that he will not raise his hand; what it does mean is that here and now he has a faculty that exists in the mode of being of potentiality, whatever will happen in actuality. In this context, entities in classical mechanics, as well as faculties in quantum mechanics, are the basic presuppositions needed for the determination of the classical and quantum experiences; they allow the bringing together of observation and measurement. De Ronde emphasizes then that entities are non-contextual existents whereas faculties are relational contextual existents. The notion of complementarity is crucial here, understood in this case not as bringing together different incompatible representations but, rather, as providing the constraints under which faculties exist.

For this purpose, Ψ is considered to be an expression of the condition of possibility to perform a certain experiment.

This representation of the world undoubtedly has the advantage of being thoroughly isomorphous with the operation of the quantum formalism. It should not, however, be forgotten that it also rests on “hypostasis” within the meaning of Meyerson (on beings of thought escaping from the direct experiment). Because, if “faculties” of De Ronde are given only in relation with what we can do, if they are “*relational contextual existents*,” what can allow us to cut them off from what we do indeed? What can make them independent compared to the whole of the contexts of the experiments that are actually carried out? Only the fact that we need a unifying thread between all the experimental acts carried out, to raise us on the scale of general knowledge, and to be able to predict the results of experiment in all circumstances. Building a “hypostasis” is tantamount here to project on nature a formal unification enabling prediction in all experimental contexts.

How can these reflections be applied to the potentialities in the case of molecular orbitals? In *static reactivity*, the molecular landscape indicates the probable places of the molecular target at the time of an electrophilic, nucleophilic, or free-radical attack by another molecule or species. Quantum approximations supported by symmetry rules provide a framework for understanding of intramolecular cyclization, as in the case of the prediction of regioselectivity in Diels–Alder cycloaddition reactions. In a given context (choices of the reagents, solvent, physical conditions, size of the reactor, procedure, etc.), both forecast and rationalization of the experimental results are possible. In *dynamic reactivity*, a knowledge of the process of transformation is possible and allows, for instance, a better understanding of the enzymatic catalytic phenomena. Faculty and propensity, considered as new forms of dispositions within the quantum framework, seem compatible with the invisible molecular agents on which we act. Does scientific realism find a new breath in this ontological reading of the power by the contemporary practices of research? What are the conditions for the possibility of a realistic approach in this context? These difficult questions are of the greatest interest because they determine our relation with the World and our understanding of this World.

To answer these questions, it is necessary to consider the practice of research from which they originate, that is, the techno-scientific seam that combines quantum mechanics, approximations, instrumental, and algorithmic techniques. The predictive capacity of these approaches does not only depend on the molecular wavefunction but also on a host of approximations and compromises that make it possible for numerical properties and molecular landscapes to be calculated. Faculties or propensities express this pragmatic course of research and not only an ontological indeterminism or an ontological potentiality. Isabelle Stengers points out that these approximations clandestinely bestow a “really causal” role on the coupling perturbation. This causal role is irreducible with the transformation, which, incorporating the perturbation in the Hamiltonian, would eliminate the coupling. In other words, the quantum object defined in the Hilbert space needs to refer to an “exogenic” intervention, that is to an interaction that cannot be described in Hilbertian term [58].

The molecular wavefunction should be understood only within this practice of scientific inventiveness, its faculty is the fruit of a whole of practices and not the simple expression of new “discovered” methods. In *Cosmopolitiques IV*, Isabelle Stengers highlights that the question raised by physics does not return initially to the “great ideas,” realism, determinism or others, but especially with the technique, readable in terms of requirements and obligations. It does not question the existence of a reality in oneself, but perhaps raises the problem of the relevance of the requirements that the physical theories translated into terms of assertions that make their objects exist. However, in any event, it raises it on a singular mode, suitable for the physical and mathematical spaces of inventiveness but also with its historicity. In addition, it should not be forgotten that these calculations and these algorithms depend on the choice of the selected bases, which generally refers to an established culture acquired through the authority of practice. The bases built by Ransil and Roothaan in the LMSS under the direction of Mulliken testify to this compromise related to the approach of the self-consistent field. In the same way, it should not be forgotten that this molecular wavefunction is not the only possible functional of energy but that the electronic density is equally usable. Moreover, Schrödinger’s wave-mechanical formalism is not a unique approach to quantum mechanics; the equivalent matrix approach of Heisenberg has, without any recourse to wavefunctions, expressed the bases of quantum mechanics indeed. If a realism remains possible, it must be a question of a different realism by which we will have to learn to resist the lure that the forms of expression exert on us, to paraphrase Ludwig Wittgenstein in his *Blue Book* [59].

We have shown that the ontological question knowing if the natural laws *are* or *are not* deterministic cannot be solved. Indeed, deterministic appearances can result from a statistical regularity whereas indeterminist appearances can reflect an underlying deterministic phenomenon such as chaos [60]. How to understand then these new forms of ontological realism that intrinsically take for granted the existence of indeterminist propensity in nature?

In his book *Projective Probability*, James Logue has shown that any coherent system of probabilistic evaluations can be interpreted in a realistic manner, that is, it can be understood as expressing proposals the truth value of which is independent of our means of testing them [61]. In other words, the proposals belong to reality and cannot be reduced to our faculty of knowledge, and, in this view, Logue disagrees with Popper. This interpretation in its turn can lead the authors of a probabilistic evaluation to project it on the world. Unsurprisingly, under these conditions, the coherent system of probabilistic evaluations of quantum physics, not counterbalanced by a fertile deterministic program, could be conceived by researchers as eminent as Popper (and even Heisenberg in his way) as translating partly or entirely a “real” or an “existing” characteristic of the world. Incontestably, the defenders of an ontological indeterminism devote themselves here, very much as the defenders of the hidden variables theory, to what Kant would have denounced like an attempt to extend the application of our concepts beyond the limits of the experiment [60]. The question of the philosopher then no consists in asking whether reality is or is not

made of potentialities having the structure of the probabilistic algorithm of the quantum theory, but only if we lose (or do not lose) knowledge by interpreting this algorithm in a realistic way. To answer this question, we may have recourse to an analysis of the status of probabilities in quantum mechanics. The usual functions of probabilities rest on Boolean algebra whereas Hilbert space uses ortho-algebras that contain the previous one as substructure. As long as the contexts can be joined or provided that measurements are independent of the order of intervention of these contexts, nothing prevents us from merging the ranges of the possible in only one range with respect to one global context. It is then possible to bypass this global context and to treat the elements of the range as if they translate as many intrinsic measurements.

Notably, this presupposition and this operating mode of the language are related to a Boolean classical logic and to a Kolmogorovian theory of probability. But as soon as the conjunction of the contexts is impossible or the order of use of the contexts is essential, as it is the case in microscopic physics when one tries to measure canonically combined variables, these methods become unusable. The strategy of not taking into account experimental contexts fails, and the explanation of the contextuality of the determinations becomes imperative. Contextuality is of the utmost importance and another probabilistic background is, thus, required. It is this failure in the conjunction, this imperative character of the clarification of the contextuality, that makes the realistic interpretation of the formalism so arguable. In quantum physics, the relation between the probabilistic substructures and the under-logics related to a context is organized within the framework of a meta-logic and a meta-contextual probabilistic formalism. The quantum theory of probability is a calculation of conditional probabilities depending on the interposition of a given experimental structure fixing its own range of outcomes, and not a calculation of categorical probabilities relating to the elements of a pre-data range of possible events [62].

Predictive information that one must be able to derive from a given setup (preparation) cannot be identified, thus, with a simple list of probabilities applied biunivocally to a list of events determined in advance, but to a general-purpose mathematical tool that can be used to generate as many derived lists of probabilities as the lists of facts corresponding to the types of experiments one could carry out at the end of the setup. Therefore, quantum mechanics is none other than a meta-contextual form of theory of probability with regards to its theoretical function; in other words, this functional approach expresses contextuality without being bound to a precise context. It collects the conditions of possibility of a unified system of probabilistic prediction relating to phenomena inseparable from, what may sometimes be incompatible, contexts. It is then enough to supplement this theory by various symmetries to create as many particular variants of quantum theory. The association of the two elements allows quantum mechanics to become a probabilistic system of evaluation adapted to a class of experimental situations whose extension depends on the implemented symmetries. It is completely possible to understand, from this pragmatic point of view, the results of the experiments on molecular reactivity and the quantum predictions related to structure and stability of molecules without referring to an ontology. This approach dissolves the problem

of measurement and the reduction of the wavefunction or that of molecular orbitals and suggests to us a reassessment of the concept of experiment. What we must grasp is that there is no more “pure reason” and that the reason is primarily pragmatic while giving a priori a functional meaning within the practical framework of the activity of research. An “a priori form” is thus regarded as a structure provided in advance by some kinds of activity of research, which is altered as soon as those activities are given up or redefined. This clearly differs from Kant’s anhistorical and universal a priori.

32.4

Closing Remarks

The techno-scientific seam of which we stressed the importance of understanding the meaning of the molecular landscapes and the compromise evoked by Isabelle Stengers also demands a redefinition of the scientific method. It is no longer a question of defining and characterizing objects from a deduction using universal principles systematically but rather of extending the scientific method to the direct coordination of the experiments by biases of the practices and of the contexts of inventiveness without having recourse to a World in itself. Thus, the ontological scientific paradigm (deeply rooted in ontological metaphysics) cannot allow us to understand modern sciences – a study that takes into account practices of research and contextuality paves the way for a new cognitive relation with the world.

Acknowledgments

We thank Isabelle Stengers and Chérif F. Matta for their suggestions, help, and high humanism.

References

- 1 Birge, R.T. (1926) *Nature*, **117**, 300–301.
- 2 Mulliken, R.S. (1925) *Phys. Rev.*, **26**, 561–572.
- 3 Sommerfeld, A. (1934) *Atomic Structure and Spectral Lines* (translated by H.L. Brose), Methuen, London.
- 4 Langmuir, I. (1919) *J. Am. Chem. Soc.*, **41**, 868–934.
- 5 Mulliken, R.S. (1926) *Proc. Nat. Acad. Sci. U.S.A.*, **12**, 158–162.
- 6 Hund, F. (1927) *Z. Physik*, **40**, 742–764. Reproduced in: Hettema, H. (2000) *Quantum Chemistry: Classic Scientific Papers*, World Scientific Publishing, p. 226.
- 7 Mulliken, R.S. (1930) *Rev. Mod. Phys.*, **2**, 60–115.
- 8 Mulliken, R.S. (1931) *Rev. Mod. Phys.*, **3**, 89–155.
- 9 Mulliken, R.S. (1932) *Rev. Mod. Phys.*, **4**, 1–86.
- 10 Duhem, P. (1985) *Le Mixte et la Combinaison Chimique*, Fayard, Paris.
- 11 Mulliken, R.S. (1932) *Phys. Rev.*, **4**, 49–71.
- 12 Van Vleck, J.H. (1970) *Pure Appl. Chem.*, **24**, 235–256.
- 13 Gavroglu, K. and Simoes, A. (1994) *Historical Stud. Phys. Biol.*, **25**, 47–110.

- 14 Kuhn, T. (1983) *La Structure des Révolutions Scientifiques (The Structure of Scientific Revolutions)*, Flammarion.
- 15 Bensaude-Vincent, B. (2005) *Faut-il Avoir Peur De la Chimie?*, Les empêcheurs de penser en rond éditeurs, Paris, p. 146.
- 16 Pullman, A. and Pullman, B. (1964) π Molecular orbitals and the processes of life, in *Molecular Orbitals in Chemistry, Physics and Biology, A Tribute to R.S. Mulliken* (eds P.O. Löwdin, and B. Pullman), Academic Press, New-York and London.
- 17 (a) Evans, M.G. and Gergely, J. (1949) *Biochim. Biophys. Acta*, **3**, 188–189; (b) Paoloni, L. (1959) *J. Chem. Phys.*, **30**, 1045–1046; (c) Pullman, A. and Pullman, B. (1963) *Nature*, **199**, 467.
- 18 Marmur, J. and Doty, P. (1959) *Nature*, **183**, 1427–1428.
- 19 Pullman, A. and Pullman, B. (1961) *Nature*, **189**, 725–728.
- 20 Serres, M. (1999) *Paysages des Sciences*, Le Pommier, Paris, p. 23.
- 21 Serres, M. (1972) *Hermès II: L'interférence*, Les Editions de minuit, Paris, p. 129.
- 22 Lambert, D. and Rezsöházy, R. (2004) *Comment les Pattes Viennent au Serpent? Essai sur l'étonnante Plasticité du Vivant*, Nouvelle Bibliothèque Scientifique, Flammarion, Paris.
- 23 Bouveresse, J. (1996) *La Demande Philosophique: Que Peut la Philosophie et Que Peut-on Vouloir d'elle?*, L'éclat (Ed), Paris, p. 34.
- 24 To better understand the dangers of analogies, see: Bouveresse, J. (1999) *Prodiges et vertiges de l'analogie: de l'abus des belles-lettres dans la pensée*, Raisons d'agir, Paris, pp. 13–14.
- 25 Stengers, I. (1987) *D'une Science à l'autre: Des Concepts Nomades*, Editions du seuil, Paris, p. 19.
- 26 Serres, M. (1999) *Variations sur le Corps*, Le Pommier (Ed), Paris, pp. 179–187.
- 27 Deleuze, G. and Guattari, F. (1991) *Qu'est-ce que la philosophie?*, Editions du minuit, Paris, pp. 23–24.
- 28 Rivail, J.L. (1999) *Éléments de Chimie Quantique à l'usage des Chimistes*, 2nd éd., Savoirs actuels, EDP Sciences/CNRS éditions, pp. 346–347.
- 29 Serres, M. (1999) *Paysages des Sciences*, Le Pommier (Ed), p. 18.
- 30 Levy, P. (1987) *Le Paradigme de Calcul in D'une Science à l'autre: des Concepts Nomades*, (ed I. Stengers) Editions du seuil, Paris, p. 96.
- 31 Stengers, I. and Prigogine, I. (1979) *La Nouvelle Alliance. Métamorphose de la Science*, Folio Essais, Paris, p. 76.
- 32 Rivail, J.L. (1999) *Éléments de Chimie Quantique à l'usage des Chimistes*, 2nd éd., Savoirs actuels, EDP Sciences/CNRS éditions, pp. 383–384.
- 33 Fukui, K. (1964) A simple quantum-theoretical interpretation of the chemical reactivity of organic compounds, in *Molecular Orbitals in Chemistry, Physics and Biology, A Tribute to R.S. Mulliken* (eds P.O. Löwdin, and B. Pullman), Academic Press, New-York and London, pp. 513–537.
- 34 Salem, L. (1968) *J. Am. Chem. Soc.*, **90**, 543–553.
- 35 Salem, L. (1982) *Electrons in Chemical Reactions: First Principles*, Wiley-Interscience, New York.
- 36 Anh, N.T. (1995) *Orbitales Frontières*, InterEditions/CNRS Editions, Paris.
- 37 Anh, N.T. (1970) *Les règles de Woodward et Hoffmann*, Ediscience, Paris.
- 38 Woodward, R.B. and Hoffmann, R. (1970) *The Conservation of Orbital Symmetry*, Verlag Chemie, Weinheim.
- 39 Stengers, I. (1997) *Cosmopolitiques IV*, Editions de la Découverte, Paris, p. 75.
- 40 Serres, M. (1999) *Paysages des Sciences*, Le Pommier (Ed), Paris, p. 34.
- 41 Stengers, I. (1997) *Cosmopolitiques IV*, Editions de la Découverte, Paris, p. 94.
- 42 Translated from: Meyerson, E. (1995) *De l'Explication dans les Sciences*, Fayard (Ed), Paris, p. 402. "Ce qui est en soi, c'est une réalité, un pouvoir, mais qui, de son intérieur, n'est pas encore parvenu à l'existence".
- 43 Translated from: Meyerson, E. (1995) *De l'Explication dans les Sciences*, Fayard (Ed), Paris, p. 412. "Aucune hypostase n'est concevable, si nous n'admettons la possibilité d'une existence en puissance".
- 44 Translated from: Meyerson, E. (1995) *De l'Explication dans les Sciences*, Fayard (Ed), Paris, p. 417. "C'est que le penseur,

- aussitôt qu'il a recours à ce concept d'état de puissance, se trouve en proie à deux tendances opposées: il faut que le potentiel se distingue de l'actuel et il faut cependant qu'en dépit de cette distinction il puisse lui donner naissance, ce qui n'est possible que s'il est identique, s'il peut être confondu avec lui. Il faut donc que la pensée pose simultanément ces états comme semblables et comme différents et qu'elle concilie ou du moins ait l'air de concilier cette contradiction. C'est là à quoi l'ingéniosité des philosophes s'est inlassablement appliquée".
- 45 Translated from: Meyerson, E., (2000) in *Identité et Réalité*, Broché, pp. 420–421. "La science procède-t-elle autrement quand elle traite l'énergie cinétique et l'énergie potentielle et même des énergies fondamentalement différentes, telles la chaleur, l'électricité, l'énergie mécanique, comme des 'formes' d'une seule et même essence fondamentale (nous dirons en usant du terme de Mnésarque, comme des modalités pouvant se transformer l'une en l'autre sans cesser de se conserver)? Comment dès lors la philosophie, dont la tâche propre consiste dans l'effort tendant à composer une image cohérente du grand Tout, échapperait-elle à cette nécessité inéluctable ?".
 - 46 Albert, D. (1992) Chapter 1, in *Quantum Mechanics and Experience*, Harvard University Press, Cambridge.
 - 47 Dorato, M. (2002) Dispositions, relational properties and the quantum world. *Proceedings of the Conference on Dispositions and Causal Powers, Paris X and Ecole Polytechnique*, 19–21 September 2002 (ed. M. Kistler), Ashgate, 249–271.
 - 48 Heisenberg, W. (1958) *Physics and Philosophy: The Revolution in Modern Science*, Harper and Row, New York, pp. 67–69.
 - 49 Suárez, M. (2004) On quantum propensities: two arguments revisited. *Erkenntnis*, **61**, 1–16.
 - 50 Margenau, H. (1954) *Phys. Today*, **7**, 6–13.
 - 51 Popper, K.R. and Bartley, W. (1989) *Quantum Theory and the Schism in Physics: From the Postscript to the Logic of Scientific Discovery*, Routledge (Ed), pp. 205–206.
 - 52 Popper, K.R. (1997) *A World of Propensities*, Thoemmes Press, pp. 35–40.
 - 53 Readhead, M. (1987) *Incompleteness, Non-locality and Realism*, Clarendon Press, Oxford, pp. 45–50.
 - 54 Dorato, M. (2002) Dispositions, relational properties and the quantum world. *Proceedings of the Conference on Dispositions and Causal Powers, Paris X and Ecole Polytechnique*, 19–21 September 2002 (ed. M. Kistler), p. 13.
 - 55 Suárez, M. (2004) *Erkenntnis*, **61**, 13–21.
 - 56 De Ronde, C. (2005) *Potencialidad Ontologica y Teoria Cuantica* (eds H. Faas, A. Saal, and M. Velazco), *Epistemologia e Historia de la Ciencia*, vol. 11, Universidad Nacional de Cordoba, Cordoba, pp. 204–211.
 - 57 De Ronde, C. (2007) A topological study of contextuality and modality in quantum mechanics. *International Journal of Theoretical Physics*, **47**, 168–174.
 - 58 Stengers, I. (1997) *Cosmopolitiques IV*, Editions de la Découverte, Paris, pp. 92–93.
 - 59 Wittgenstein, L. (1980) *The Blue and Brown Book (1933–1934)*, Harper & Row, New York, pp. 1–74.
 - 60 Bitbol, M. (1998) La mécanique quantique comme théorie des probabilités généralisée, in *Prévision et Probabilités dans les Sciences* (eds E. Klein and Y. Sacquin), Editions Frontières, Paris.
 - 61 Logue, J. (1995) *Projective Probability*, Oxford University Press.
 - 62 Bitbol, M. (1996) *Mécanique Quantique: une Introduction Philosophique*, Nouvelle Bibliothèque Scientifique, Flammarion, p. 146.

Index

a

- ab initio* calculations 681
 - levels of theory 681
- ab initio* continuum solvent study 155
- ab initio* methods 483, 771, 877
- ab initio* molecular orbitals 478
- ab initio* MP2 energies 482
- ab initio* polarizabilities, comparison 395
- ab initio* quantum mechanical theory 147, 151
 - calculations 29
 - interaction energy 43
- ab initio* SCF theory XIV
- ab initio* statistical mechanical theory 425
- ab initio* wavefunctions 675
- acceptor stem mimic (ASM) sugar moiety 505
- acetylcholine 757–761
 - hydrolysis by cholinesterases 757
- acetylcholinesterase (AChE) 757–761, 769, 772, 773, 776, 777
 - active site 777
 - gorge regions 758
 - catalysis 772
 - catalytic serine residue, S203 759
 - deactivation 769
 - effect on 773
 - irreversible inhibitors 777
 - by 3-*N,N*-dimethylphenyl carbamate 769
 - second-order rate constant 769
- acetyl coenzyme A 624
- acetylphenothiazine 770
 - butterfly angles 770
 - molecular theory, levels 770
- acid-base equilibrium 669, 683
 - (de)protonation 683
- active-site model 88, 90, 103, 104, 106, 643, 650, 661
- adaptive landscape model 880
- adaptive Poisson–Boltzmann solver (APBS) 150
- adenine 202, 205, 277
 - formation 202, 208, 210, 211
 - reaction profiles 208, 210, 211
 - interactions, examples 331
 - prebiotic experiments 202
 - precursors, DAMN vs. AICN 205
- adenine–gold interaction 254
 - binding energies 254–263
 - bonding patterns 254–257
 - bond lengths 258
 - complex A-Au₃(N₆) 262
 - features 256, 580
 - gold, propensity 257–263
 - hydrogen bonding 259–262
 - H-bonds, WC intermolecular 289, 290
 - interaction 262
 - potential energy surface (PES) 254, 255
- adenine–histidine system 321
 - stacking, representative examples 319
- adenosine 5'-triphosphate (ATP) 473, 475, 477, 484, 485, 487, 488, 490, 491, 493, 495
 - functioning 475
 - group charges 491, 492
 - distribution 492
 - presence/absence of Mg²⁺ 491
 - hydrolysis 484, 485, 487–489
 - atomic contributions 488
 - bond properties 487
 - energy, atomic contributions 487
 - energy delocalization 485–487
 - exergonic hydrolysis 473
 - global energies 485
 - Lipmann's group transfer potentials 485
 - molecular graphs 487
 - presence/absence of Mg²⁺ 487
 - molecular electrostatic potential 492

- in the absence/presence of Mg^{2+} 492
- reactants and products geometries 477
- ball-and-stick representation 477
- molecular graphs 489
- regions 473
- triphosphate chain bond lengths 490
- adenosylcobalamin (AdoCbl) coenzyme 105
- ADPGV7b molecule 32–34
 - functions, calculations 32–34
 - KEM calculation 33
 - X-ray crystal structure 33
- adsorption, distribution, metabolism, excretion, and toxicity (ADMET) 46
- Aharomov-Bohm effect 883
- albumin 783, 784
 - copper binding, type 2 783, 784
- AlkB family, enzymes 652, 654
 - active site, structural models 655
 - function 654
 - 1-methyladenine 655, 657
 - repair mechanism 657
 - spin-quintet potential energy surfaces 657
 - oxidative dealkylation mechanism 654–657
 - α -ketoglutarate (α -KG) co-substrate 655
 - stages, PESs 655
- Alzheimer's disease (AD) 743, 745, 789
 - amyloid beta peptide ($\text{A}\beta$) 789–791
 - copper 789–791
 - peptide-bound copper (II), reduction potentials 781
 - targeting butyrylcholinesterase 757
 - therapy 757
- AMBER approach 78, 731
 - charges 78
 - force field 608
 - library 531
- ambrosianolides 626
 - stereogenic center, stereochemistry 626
- amino acids 227, 229, 309, 393, 396, 398, 407, 408, 412, 423, 425, 438, 442, 443, 448, 451, 456, 461, 462, 507, 512
 - α -amino/ α -carboxylic groups 439
 - aromatic, stacked dimmers, interaction energies 324, 328
 - atomic numbering 309
 - computational approach 438
 - geometries 438
 - ionization 227–234
 - intramolecular proton-transfer processes 231–234
 - radical cation amino acids, structural features 227–230
 - level of theory 438
 - linear regression equation 442
 - lowest-energy conformers 229, 230
 - optimized geometries 229, 230
 - model 410
 - non-catalyzed reaction 511
 - partial molar volume 440
 - partition coefficients 448
 - genetic mutations, simulation 448
 - polarizabilities 393–398
 - properties 407, 425, 456
 - QTAIM side chain polarizations 408–414
 - ray/neutron diffraction geometries 462
 - reactive surface maps 461
 - residues 451, 758, 771
 - arrangement 758
 - Stick model 406
 - side chains 411, 439, 441, 448
 - Andrews plots 411
 - charge separation index 441
 - neutral, local polarity 441
 - QTAIM atomic properties 439
 - QTAIM variables 411
 - structural propensities 408
 - theoretical variables 408
 - structure 309
 - theoretical classification 408–414
 - zwitter-ionic forms 443
- 4-aminoimidazole-5-carbonitrile (AICN) 203, 207
 - tautomeric forms 207–213
- 3-aminopropane-1-sulfonic acid (3-APS) 751
- Amsterdam density functional (ADF) program 579
 - SCF procedure 558
- amyloid precursor protein (APP) 748, 749
 - processing 749
- Anacystis nidulans*, DNA photolyase 524
- Andrews plots (APs) 410
- angle of edge rotation, definition 314
- anti-cancer drug 723
 - absorption, distribution, metabolism, excretion (ADME) aspects 739
 - problem 739
 - properties 739
- carboplatin 723
- cisplatin 723–726
 - calculation 726–731
 - structure 726
- DNA interactions 726–731
- non-platinum alternatives 735–738
- platinum-based alternatives 732–735
- platinum complexes 723

- (anti)ferromagnetic spin coupling 538
 - antiferromagnetic diiron complexes 539
 - spin density functional theory 539–542
 - apparent surface charge (ASC) method 145
 - apurinic and apyrimidinic (AP) site 525
 - endonucleases 525
 - aromatic DNA-protein components 318
 - interactions 318–326
 - stacking interactions 319–323
 - T-shaped interactions 323–326
 - aromatic hydrocarbon (Ah) receptor 685
 - artificial neural network (ANNs) model 704, 708
 - algorithms 710
 - structure 709
 - aryl selenol framework 599
 - AT–AT complex, HB BCPs 375
 - atomic densities, nonspherical modeling 429
 - atomic dipole interaction model (ADIM) 390
 - atomic dipole moment 437
 - atomic electronic dipole moments 374
 - atomic energy 348, 353, 473, 477, 480, 494
 - *ab initio* methods, QTAIM atomic energies 478
 - ATP, energy richness 473
 - changes, bar graphs 348, 353
 - computational details 484
 - direct evaluation 480
 - energy of reaction, atomic contributions to 484
 - from Kohn–Sham density functional theory method 482
 - theoretical level, choice 477–484
 - BDE, atomic contributions empirical correlation 478
 - problem 477
 - atomic reaction energy profile, construction 595
 - atomic Shannon entropy 367
 - atomic shell approximation (ASA) 688
 - atomic virial theorem 356, 435
 - attractive/repulsive contributions 356–360
 - short-range nature 356–360
 - atom's average energy, definition 495
 - atoms in molecules (AIM) method 137, 493
 - quantum theory 493
 - b**
 - Bacillus cereus*, BcII 612, 614, 615
 - β -lactam antibiotics docking 615
 - two-step mechanism 614
 - Bacillus fragilis* 612, 613
 - CcrA 612
 - di-Zn MbL CcrA, QM/MM structure 613
 - base excision repair (BER) 517
 - basis set superposition error (BSSE) 311, 368
 - Bayesian regularization (BR) approach 710
 - benchmark systems 562
 - benzene derivatives, QSAR models 706
 - benzoquinone-hydroquinone complex, *see* quinhydrone
 - beta-amyloid 748, 750, 752, 753, 790, 792
 - nature 752
 - protein misfolding 748, 750
 - betaine dyes 162
 - example 163
 - solvatochromism, use 162
 - Bettens–Collins isotropic formula 178
 - binuclear antiferromagnets 540
 - biogenetic processes, chemical simulation 625
 - *in vitro* transformations, principles 625
 - bioinorganic catalysis 552
 - biological system 880
 - cycles 873
 - evolution 876–882
 - matrix 154
 - biphenyl-4-yl(10*H*-phenothiazin-10-yl)
 - methanone 774
 - conformation 774
 - β -lactamase 605, 615
 - β -lactam 610, 614
 - antibiotics 605, 617
 - carbonyl carbon, hydroxyl nucleophile attack 614
 - carboxylate moiety 617
 - different types 610
 - drugs 605, 606
 - families 606
 - hydrolysis, single-step mechanism 614
 - proton donor 614
 - ring 613, 614
 - C–N cleavage 613
 - substrates 606
 - metal ions 606
 - metallo β -lactamase hydrolysis 606
- blood-brain barrier (BBB) 712
- B3LYP method 86, 552
 - atomic energies, definition 483
 - calculations 98, 102
 - DFT method 729
 - functional 740
 - 6-31G(d) method 589, 596, 597
 - Hartree–Fock exchange 552
 - hybrid density functional method 796
 - LB//B3LYP/SB level 797
- BLYP DFT functionals 526, 734

- B1 metallo β -lactamase (M β L) 610, 612, 615
 - catalytic mechanism 612
 - cefotaxime enzymatic hydrolysis in BcII 614
 - cefotaxime enzymatic hydrolysis in CcrA 613
 - β -lactam antibiotics, reactivity 615
 - di-zinc MbL 613
 - CcrA 611
 - di-Zn/3H mono-Zn reaction mechanisms 612
 - binding mode 611
 - Michaelis complex 610
 - nucleophile structural determinants 611
 - substrate binding determinants 610
 - reactivity 615
 - subclass, hydrolytic mechanism 617
 - zinc content 615
- B2 metallo β -lactamase (M β L) 616
 - CphA 616
 - ImiS 616
 - Sfh-I 616
- Bohm's theory 888
- bond-breaking process 523
 - computational study 523
- bond critical point (BCP) 366, 697, 431, 489, 509, 510
 - density 491
 - space 672, 674, 681
 - dimension 677
 - ellipticity vs. electron density 673
 - full set 673
 - properties 674, 677, 678, 682, 686
 - representation of phenol 673
 - variables 682, 683
 - VIP plots 684
 - stacking properties analysis 377
- bond dissociation energies (BDE) 474, 475, 478, 479, 495
 - atomic contributions 479
 - atomic partitioning 475
- bond dissociation, enthalpy 474
- Boolean algebra 892
- Born model 149
- Born–Oppenheimer (BO)
 - approximation XIII, 746, 881, 883
- Born–Oppenheimer molecular dynamics approach 520
- Born–Oppenheimer wavefunction 426
- boundary element method (BEM), formalism 146
- B3PW91/6-311G(2df,p) method 587, 592
- bridging algorithms 880
- brute force approach 135
- butterfly angle, definition 770
- butyrylcholinesterase (BuChE) 757–760, 771, 772, 774–778
 - active site gorge regions 758, 771
 - catalysis 772
 - catalytic serine residue S198 759
 - cholinergic neurotransmission, regulation 757–760
 - cholinesterase function 760
 - human wild-type 775
 - inhibition 760, 772, 773
 - mutants 775
 - mutations 778
 - optimizing specific inhibitors 761
 - phenothiazine scaffold 761
 - phenothiazine carbamates 775
 - π - π stacking 776
 - reversible inhibition 774
 - schematic representation 759
 - time-dependant inhibition 776
 - wild-type 776–778
 - pseudo-irreversible inhibitors 774, 777
- butyrylthiocholine, rate of hydrolysis 776
- C**
- Cambridge crystallographic data center (CCDC) 379
- carbamate 773, 777
 - action 773
 - carbonyl group 777
 - cholinesterase inhibitors 767
 - physostigmine/rivastigmine 767
 - derivative 777
 - phenothiazine, butterfly angles 777
 - phenothiazine ring system 777
- carbocation, hydrogen shifts 627
- Carbó index 674
- α -carbon atom 439, 442
 - transferable group properties 442
- carbon monoxide oxidation 247
- carboplatin, Raman spectrum 732
- Car–Parrinello scheme 520, 608
 - molecular dynamics techniques (CPMD) 649, 729
- complete active space self-consistent field (CASSCF), 67, 87, 112, 114, 115, 118, 119
 - wavefunction 92
- catalytic cycles 564, 576, 578
 - cytochrome P450cam 564–567
 - high-spin Ni(II) 578
 - His-Porphyrin models 567–574
 - Ghosh reference data 570
 - Harvey reference data 568–570
 - model systems 571–574

- NiFe hydrogenase 574–578
- spin-states involvement 564–578
- catechol 378, 379
 - face to face (FF) dimer 379
- cation- π interactions 326
 - ligand-antibody binding interactions 326
 - receptor-ligand interactions 326
- C–C bond 673, 674
 - π character 673
- central nervous system (CNS) 711
 - active drugs 713
- ceruloplasmin 781, 786
- C–H/ π complexes 381–384
 - adducts 382
 - bonds 383
 - BCP properties 383
 - geometry features 383
 - model systems, QTAIM analysis 381
 - molecular graphs 382
- charge dependence of the effective atomic polarizability (CDEAP) model 391
- charge separation index (CSI) 441, 442, 450
- charge-transfer complexes 367–371
- CHARMM force field 751, 752
- chemical bond 435
 - interaction 491
 - ionicity 491
 - structure 434
- chemical equilibria 152
 - in molecular aggregation 153
 - pK_a of acids 153
 - tautomeric equilibria 153
- chemically initiated electron-exchange luminescence (CIEEL) 118
- chemical transferability 339–343
- chicory roots, (+)-germacrene A synthetase isolation 625
- 8-chlorocarbazole acetic acid (CCA) 825, 830
 - decarboxylation energy curves 826
 - deprotonated triplet states 826
- cholinesterases 757, 768, 773
 - catalysis, active site serine, acylation/deacylation 768
 - inhibitors 761, 767
 - biological evaluation 761–769
 - phenothiazine derivatives 761
 - pseudo-irreversible inhibitors 773
- choline thioesters 766
 - cholinesterase catalyzed hydrolysis, Ellman's method 766
- chromosomally-encoded enzymes 605
 - *Bacillus cereus* (BcII) 605
 - *Bacteroides fragilis* (CcrA) 605
 - *Elizabethkingia meningoseptica* (BlaB) 605
- circular dichroism (CD) 788
- cisplatin 723, 725, 733, 740
 - adducts, hydrogen bonds 730
 - ammine loss 740
 - hydrolysis 727
 - toxic side effects 733
- cisplatin-water complexes 727
- Clinton's equations 7–10
 - applications 7–10
 - beryllium 7
 - maleic anhydride 9
 - experimental density matrix 10
- closed isolated system 436
- CODESSA program 714
- coenzyme-substrate complex, transformation 878
- collagen 47
 - KEM 47–50
 - interaction energy calculations 49
 - triple helix 49
 - 1A89 47–49
 - calculations, comments 50
 - energy calculations 49
- Collins complex 571
- comparative molecular field analysis (CoMFA) method 706
- complete active space multiconfigurational second-order perturbation theory (CASPT2) 87, 91
- complete basis set (CBS) 311
- complex biomolecules, hydrolysis 757
- complex systems 171
 - fast marching method 171
- composite systems 131
 - vs. aggregate system 876
 - response properties 131
- computational chemistry 61, 308
- computational investigation method 202–213
 - pentamerization, thermochemistry 204
 - step by step mechanism 205–213
- computational modeling 786
- computational quantum mechanical methods 607
- computer-aided drug design (CADD) 403
- conductor-like polarizable continuum model (C-PCM) 394
- conductor-like screening model (COSMO) 147, 702
- conformers 404–408
- continuum solvation methods 148–150, 155, 158, 164, 314
 - apparent surface charge (ASC) methods 148
 - finite difference method (FDM) 150

- finite element method (FEM) 150
- generalized Born model 149
- multipole expansion methods (MPE), limitations 149
- polarized continuum model (PCM) 314
- convergences index 864
- copper 794
 - binding 783
 - albumin-type 2 783
 - ceruloplasmin-type 1 785–787
 - peptides/proteins 787
 - site models 786
 - Cu/A β , Cu(II)/Cu(I) reduction potentials 791–794
 - Cu(3-Clip-Phen), schematic view 735
 - Cu(II) 791
 - His residues 793
 - reduction 791
 - non-reversible redox behavior 794
 - peptide complexes 795
 - computational methodology 796
 - reduction potentials calculation 795
 - proteins, reduction potentials 781, 794
- correlation coefficient 670, 679, 681
- corticosteroid-binding globulin (CBG) binding activity 685
- Coulomb interaction 133, 139
 - energy 134
 - linear scaling 133
 - potential 520
- coupled cluster theory (CC) 310
- Creutzfeldt–Jakob disease (CJD) 745, 787
- Criegee-type rearrangement pathway 99, 100
- crystallographic information files, deposition XX
- crystallographic unit cell 429
- crystal structure analysis 19
- crystal X-ray diffraction 738
- cyclic hexapeptide trihydrate 17
- isodensity surface 17
- cyclobutane pyrimidine dimer (CPD)
 - mechanism 521, 522
- cyclodecane, inter-conversion energy 630
- cyclooxygenase (COX) 808
 - COX-1/2 isozyme 809
- cytochrome P450cam enzymes 551, 565
 - catalytic cycle 552, 565
 - hydroxylation reaction 551
- cytosine-gold (C–Au) interactions 272
 - basic features 274
 - binding energy 272
 - bond lengths 275
 - conformers 272
 - PES 272
- d**
 - Dalton quantum chemistry program 149
 - decarboxylation-elimination reaction 624
 - dechlorination reaction 191
 - catalyzed by *trans*-3-chloroacrylic acid dehalogenase (CAAD) 191
 - density functional theory (DFT) method 86, 220, 308, 310, 478, 502, 505, 520, 552, 568, 573, 587, 596, 643, 649, 650, 694, 729, 738, 873
 - ability 587
 - AMBER approach 731
 - approximation 32
 - BLYP framework, QM/MM calculations 612
 - calculations 91, 93, 137, 339, 444, 568, 576, 608, 698
 - functionals 570, 578, 579
 - exchange-correlation functionals 727
 - systematic failures 579
 - Kohn–Sham equations 482, 502
 - atomic energy 482
 - wavefunction 483
 - optimization 739
 - quantum computations 505, 511
 - SCRF solvation 729
 - density matrices 4–11, 22, 56, 57, 696
 - wavefunction representability, problem 4
 - deoxyHr 543, 545
 - antiferromagnetic binuclear center 543
 - deoxyHrNO 546
 - EPR data simulations 546
 - spin Hamiltonian parameters 546
 - Mössbauer data simulations 546
 - spin Hamiltonian parameters 546
 - deoxyribonucleic acid (DNA) 39, 245, 873
 - adduct, formation 725, 736
 - base dimers 347
 - atomic numbering schemes 347
 - molecular graphs 347
 - base pairing, atomic energies 343–350
 - AA1, energy changes 349
 - CC, energy changes 349
 - dimerization 346–349
 - GG4, energy changes 350
 - TT2, energy changes 350
 - hydrogen bonding 347
 - binding ligands 725
 - biological properties 873
 - calculations, comments 41
 - chain, mid-segmentation 253
 - cleaving agent 734
 - damage 725, 729
 - double-stranded (dsDNA) 253

- features 251
- films 252
- fragment 738
- functionalized gold NPs 250
- glycosylases 525, 529
- gold nanotubes 252
- groove binders 725
- helix 730, 734
- hydrolysis 525–527, 529
- interactions 738
 - Au NPs, interaction 250, 251
 - base-gold interaction 252, 253, 273–286
 - charge state 278–286
 - complexes, anchoring bond 276–278
 - gold-bonding patterns 253, 254
 - modification 738
- KEM 39–41
- molecular systems 41
- nanodimensions 248–253
- nucleobases 309
 - atomic numbering 309
 - structure 309
- oligomers 730, 731
 - photolyase 518
 - thymine dimmers, repair 518
- photolyase catalyzed splitting reaction 524
- polymerase 852
- repair enzymes 517, 518, 533
 - endonuclease IV 518
- structures, X-ray crystallography 39
- Watson–Crick base pairs 457
- without solvent, energy calculation 40
- deprotonation energy (DPE) 261
- diaminocyclohexane (DACH) 723
- diaminomaleonitrile (DAMN) 203, 205
 - photoisomerization 205
- Diels–Alder cycloaddition reactions, regioselectivity 890
- differential scanning calorimetry (DSC) 453
- 1,3-dihydro-1-methyl-2H-imidazole-2-selenol (MSeI) 598
 - GPx-like activity 598
 - structure 598
- dihydrofolate reductase (DHFR) 647
 - pH-independent catalytic rate constant 647
- diiron-oxo proteins 537, 538, 542
 - enzymes 538
 - metallic cores, structure 538
 - Mössbauer spectra, phenomenological simulation 542–546
 - hemerythrin, antiferromagnetic diiron center 542
 - hemerythrin, nitric oxide derivative 542, 543
 - reduced uteroferrin, antiferromagnetic diiron center 545, 546
- dimeric interactions, dispersion energy 146
- dioxetanone intermediate (DO) 117
- direct methods, X-ray crystallography 428
- dissociation constants 609, 644, 686
- ditopic cisplatin-copper complex 735
 - DNA binding 735
- DNA base pairs 221, 345
 - ionization 221–227
 - dimerization energies 222
 - double proton transfer reactions 223–227
 - equilibrium geometries 222
 - single proton transfer reactions 223–227
- molecular graphs 345
- π - π interactions 374–378
- DNA-protein complexes 307, 313, 316
 - potential energy surface scans 313
 - variables 313
- DNA-protein components 326
 - cation- π interactions 326–333
 - charged aromatic amino acids 330
 - charged non-aromatic amino acids 330
 - charged nucleobases and aromatic amino acids 326–330
 - hydrogen-bonding interactions 332
- DNA-protein interactions 307, 308
- DNA-protein interfaces, stair motifs 332, 333
- dose-response approach 769
 - curve, disadvantage 764
- Dronpa 111, 112
 - mutants 111
 - reaction mechanisms 112
 - reversible photoswitching, mechanism 111
- drugs 43, 45, 816
 - ameliorate symptoms 790
 - design 43–47, 669
 - KEM 43–47
 - ligand, pharmacological activity 699
 - paradigm 669
 - quantum chemical methods, applications 669
 - target interaction calculations, comments 46
 - target interaction energies 45
- DNA interactions 734
 - electron affinity (EA), calculation 816
 - ionization potential (IP), calculation 816
 - photosensitivity 806–808
 - photoallergies 807
 - photophobia 807
 - phototoxicity 807
 - receptor interaction effects 697
- RNA complex, crystal structure 45
- safety 805

e

- ebselen 589
- catalytic cycle 591, 598
- derivatives 591
- GPx-like activity 589, 596, 597
- effect of the molecular environment 598–600
- substituent effects 596–598
- hydrogen peroxide, reduction 591
- atomic electronic energy 595
- catalytic cycles, summary 590
- Gibbs energy barriers, summary 592
- schematic reaction mechanisms 593, 594
- thiol exchange reactions 596, 600
- interconversion pathway 590
- preparation 590
- structure, electronic/steric modifications 597
- effective core potentials (ECPs) 553, 586, 587
- Ehrenfest and virial theorems 340
- Ehrenfest force 341
- eigenvector following (EF) method 173
- eikonal equation 175, 176, 181
- electric field-induced second-harmonic generation (EFISH) process 159
- electron density(ies) 21, 424, 426–430, 432, 435, 458, 671, 672, 675, 677, 680, 881
- *ab initio* 671, 672
- comparison 21
- conditions 427
- Dirac's question 428
- distributions, transferability 342
- function 132
- Laplacian 458
- three-dimensional topography 430
- topology 594, 597
- electronegativity equalization method (EEM) 708
- electron-electron repulsion energy 356
- electronic embedding 62
- advantages 62
- formalism 89
- electronic indices methodology (EIM) 698
- electronic multipole moments, *see* polarizations
- electron localization function (ELF) 670
- electron-nuclear potential energy 356
- electron paramagnetic resonance (EPR) 788
- silent electronic configuration 785
- electron-rich ligands 788
- electron transfer mechanisms 567, 647
- electrostatic models 343
- atom-atom interaction energy 355
- effects 276, 746
- interactions 346, 744
- stabilization 528
- Ellman reagent 761
- 5,5'-dithio-bis(2-nitrobenzoic acid) (DTNB) 761
- energy decomposition analysis 139
- hydrogen bonding 140
- interactions with cations/proton 139
- energy minima searching approaches 174
- enhanced multiphoton ionization 226
- enzyme 625, 639, 669, 758, 769, 771, 873
- active site 771
- catalytic constant 647
- catalytic cycle, enzyme conformational dynamics, role 647
- catalytic performance 643, 660
- catalytic power 644
- catalytic triad arrangement 758
- deactivation 769
- efficiency 644
- inhibition 772
- near-attack conformation (NAC) hypothesis 646
- reaction mechanism 644
- role in cascade pathway 628, 639
- systems 520, 521
- transition state 644
- variable time 769
- Zn(II)-dependent, acid-base mechanism 658
- enzyme catalysis, mechanistics 187–191, 578, 643, 644, 648, 652, 661, 764, 890
- enzymes catalytic performance, influencing factors 643–648
- enzymology, computational modeling, categories 648–650
- FMM, application 190
- oxidation cycles 652
- parts 187
- potential energy surface (PES), Gibbs energy diagram 644, 645
- quantum mechanics/molecular mechanics (QM/MM) methods 187
- MFEP method, application 190
- small to large active-site models 643, 650
- enzyme-inhibitor complex (EI) 764, 766
- association-disassociation 764
- dissociation equilibrium constant 766
- enzyme-inhibitor structure-activity relationships 772–777
- enzyme-substrate complex 644, 646
- 8-epiconfertifin 631–633, 638
- depletion of charge 638
- formation, energetic pathway 633

- {1,2}-methyl shift 638
- reaction intermediates 631, 632
- TS structures 631, 632
- β-epimer 630, 636
- trajectory generation 636
- epoxide, isolation 637
- 4α,5β-epoxyinunolide, isolation 627
- cyclization, steps 627
- ¹⁵D_{5,1}D¹⁴ conformation 620, 627, 628
- conformational barrier 629
- conformational properties 628
- reaction intermediates 631, 632
- Samek's nomenclatures 627
- transformation mechanism 630
- TS structures 631, 632
- from *Stevia tephrophylla* 627
- transformation mechanism 627
- Escherichia coli* 525, 712
- Escher's development II 427
- Espinosa–Molins–Lecomte (EML) empirical topological formula 510
- European Medicines Agency (EMA) 810
- evolutionary theory 839, 843, 844, 853, 857
- deterministic forces, selection 839
- direction 853–855
- mutation 839, 853–855
- excitation energy transfer (EET)
 - processes 164, 165
- excited-state methods, E-states 95
- excited-state proton transfer (ESPT) 109, 114
- expansion coefficients 428

f

- farnesyl pyrophosphate, biosynthesis 625, 626
- fast marching method (FMM) 171, 173–187, 189, 191
- algorithm 190
- application 182–187
- four-well analytical PES 182
- ionized O-methylhydroxylamine, dissociation 185
- S_N2 reaction 184
- calculation 186
- heapsort technique 176
- add to heap process 177
- binary min-heap 177
- in-place sorting algorithm 176
- update heap process 177
- introduction 175
- least-squares method, interpolation 179
- procedure 182
- program 180
- calculation, initialization 181
- energy cost surface 181

- heap update 181
- setup, definitions/notation 180
- QM/MM-MFEP methods, incorporation 189
- Shepard interpolation 177
- upwind difference approximation 176
- wavefront propagation method 174
- fertile deterministic program 891
- finite difference methods (FDMs) 150
- finite element method (FEM) codes 150
- firefly bioluminescence 117, 118
- schematic reaction mechanism 117, 118
- first-principle statistical mechanical theory 425
- five-atom reaction system 177
- fluorescent proteins (FPs) 85
- Food and Drug Administration (FDA) 810
- force constant calculations 61
- Fourier series 428, 747
- Fourier transform relation 7
- fragmentation, definition 137
- free energy 144, 449, 452, 608, 609
- difference 189
- experimental vs. calculated 452
- expression 144
- transfer from gas to aqueous phase 448, 449
- free energy perturbation (FEP) technique 649
- free energy potential surface (FEPS) 152
- free methyl triphosphate, electron density envelopes 490
- Fukui's frontier orbital approximation 884

g

- gas-phase harmonic frequency analysis 651
- Gaussian approximation 672
- Gaussian basis set 587
- Gaussian calculation 177
- Gaussian 03 program 507, 629
- Gaussian-type orbital (GTO) 553, 554, 579
- basis sets 553–556, 579
- performance 555
- series 554
- Gay–Lussac prediction XI
- GC, base pairing, atomic contributions 351
- GC-AT adduct 377
- frozen geometries 377
- graph 377
- GC-GC adduct 375, 376
- graph 376
- GC-WC dimer 434
- generalized gradient approximations (GGA) functionals 552, 561
- generalized valence bond (GVB) orbitals 93

- genetic algorithm techniques 685, 705, 751
- genetically encoded amino acids 411, 413, 444, 456, 463
 - calculated vs. experimental partial molar volumes 444
 - crystallographic determinations, literature references 463–467
 - energy magnitude vs. mass 411
 - side chains 456
 - quantum theoretical classification 413
 - side chains 411
- genetic code 423, 424, 454, 456
- resilience 456
- ribosomal translation 424
- genetic information 423
- trans,trans*-germacradiene 625
 - from *trans,trans*-FPP, direct cyclization process 625
- germacranolide, biomimetic transformations 626
- GIAO method 596
- Gibbs energy 143, 495, 648
- Glu43 530, 532
 - carboxylic oxygen 530
 - water bridging position 532
- glucose, structures 158
- glucose-6-phosphate dehydrogenase activity 791
- glutathione (GSH) cofactor 588
- glutathione peroxidase (GPx) enzyme 585, 599
 - activity 596
 - antioxidant activity 589
 - catalytic cycle, steps 588, 596, 599
 - cytosolic GPx (cGPx) 588
 - gastrointestinal GPx (giGPx) 588
 - GPx-like mechanism 591
 - human plasma GPx (pGPx) 588
 - active site 588, 589
 - catalytic activity 589
 - crystal structure 588
 - peroxidase activity 589
 - natural GPx proteins 589
 - natural vs. mutant murine 586
 - phospholipid hydroperoxide GPx (PHGPx) 588
 - role 589
 - sulfur substitution for selenium 586
 - theoretical studies 600
- glutathione peroxidase (GPx) mimics, computational studies 589–600
 - GPx-like activity of ebselen 589–596
 - effect of the molecular environment 598–600
 - substituent effects 596–598
 - GPx-like catalytic cycle 599
 - thiol exchange reactions 596, 600
- glycine 228, 232
 - proton transfer process, energy profile 232
 - radical cation species 228
 - optimized geometries 228
 - relative energies 228
- glycosaminoglycan 750
- glycosyl ring 530
- Gly radical cations 231
 - lowest conformer, single occupied molecular orbital 231
- Gly residue, carbonyl oxygen 788
- gold 245, 247, 248, 282, 285, 288, 289
 - anchoring 288, 289
 - Au₂₀ fullerene-type cages, lower-energy 249
 - Au₆ cluster bridges, WC G-C pair 296
 - Au-O/Au-N type, gold-base anchoring bond 285
 - auride anion 285
 - base anchoring bond 285
 - catalysts 247
 - computational bonding 283–285
 - electron affinity 282
 - electron charge 285
 - electronic configuration 246
 - heterogeneous catalysts 248
 - Mulliken charges of 282
 - nanodimensions 246–253
 - nanoparticles 247, 282
 - activity 247
 - charge state 282
 - nitrogen, charge-transfer effect 276
 - oxygen anchorings 276
 - structure 281
 - thin films 253
- gradient extremal following (GEF) method 173
- gram-negative bacteria, antigenic lipopolysaccharides 658
- green fluorescent protein (GFP) 109
- growing string method (GSM) 172
- guaianolides 625, 626, 628, 639
- guanine-cytosine triply hydrogen-bonded DNA base pair 222, 431, 433, 434, 437, 459, 460
 - electron density, relief map 431
 - gradient vector field map, superimposition 433
 - isodensity 437
 - molecular graph 434
 - zero-isodensity surface 460

- guanine-cytosine Watson–Crick DNA
 dimer 430, 432
 – chemical structure 430
 – electron density, relief map 432
 – gradient field 432
 guanine-gold interaction 263–268
 – anchoring 265
 – B3LYP/RECP 265
 – bond lengths 267, 282
 – PES 263
- h**
- H₄B cofactor 104
 Hamiltonian kinetic energy densities 702
 Hamiltonian operator 877
 Hammond's postulate 638, 885
 Handy and Cohen's optimized exchange functional 552
 hard-soft acid-base (HSAB) principle 740
 harmonic frequency data, analysis 796
 harmonic functions 136
 HF approximation 32
 Hartree–Fock (HF) methods XIV, 220, 676, 695, 726, 728
 – calculation 21, 24, 48, 877
 – density 18
 – energy 7
 – exchange 558, 562
 – orbitals 17
 – polarizabilities 392
 – ZD00 model 397
 – theory 482
 Heisenberg–Dirac–Van Vleck Hamiltonian 539
 Heisenberg equivalent matrix approach 891
 Heisenberg exchange constant 544
 Heisenberg exchange Hamiltonian 541
 – eigenstates 541
 helanolides, stereogenic center, stereochemistry 626
 Hellmann–Feynman theorem 479, 480, 483, 520
 heme
 – degradation, reaction mechanisms 96
 – α -meso carbon 96
 – propionate groups, role 566
 heme-containing enzyme, *see* nitric oxide synthase (NOS)
 hemerythrin 542, 543
 – diiron-oxo proteins 542
 hemibond 2-center-3-electron interactions 239
 Hermitian operator 437, 887
 Hessian matrix, eigenvalues 174
 hetero-molecular complexes 371
 – π - π interactions 371–374
 HHQK-BBxB 750, 751
 – motif, histidine/lysine residues 751
 – receptor 750
 high-energy phosphate bond 473, 489
 Hilbertian quantum approach 887
 Hilbert space 890, 892
 histidine
 – basic amino acid 201
 – complexes 792
 – radical cations 231
 – carbonyl group 788
 – dimmers, interaction energies, structures 329
 – ligands 785
 – lowest conformer, single occupied molecular orbital 231
 – residues 793, 794
 Hodgkin–Richards indices 698
 Hohenberg–Kohn theorem 86, 337, 338, 426
 holistic approach, biology 877
 HOMO–LUMO 248, 371, 702, 710, 732
 – energy 710
 – energy gap 248, 702, 732
 – overlap 371
 homo-molecular complexes, *see* catechol
 – π - π interactions 378–381
 homotaurine, *see* 3-aminopropane-1-sulfonic acid (3-APS)
 Hooke's law, harmonic potentials 747
 Huntington's disease 745
 hybrid DNA sensors 252
 hybrid Hartree–Fock DFT functional B3LYP level of theory 814
 hybridization, *ad hoc* concept 875
 hybrid meta-GGA exchange-correlation functionals 520
 hydrated hexapeptide, kernel calculations 15, 17, 18
 hydrated Leu1–Zervamicin 18–22
 – fragment calculations 18–21
 hydrogen
 – atom transfer reactions 232
 – bonded systems 219
 – bridge 260
 – capped amino acid side chains 450
 – electrode 796
 – elimination models 392, 395, 397
 hydrogen bond 140, 219, 259, 278, 315–318, 344, 354, 359, 374, 376, 432, 433, 507, 510, 729, 882
 – anchoring 278
 – A-site and P-site 510

- calculations 54
- classes 315
- conventional 259
- definition 260
- distances 222
- formation 507
- interaction 140, 219, 259, 315–318, 432
- DNA backbone 315, 316
- DNA nucleobases 315, 317, 318
- protein side chains 316–318
- Linus Pauling's model 852
- networks, rearrangements 661
- NMR shift 261
- nonconventional 259, 260, 261, 278
- properties 376
- QTAIM characterization 344
- surface 433
- system, uracil-glycine dimer 317
- hydrogen peroxide, reduction reaction 591
- atomic electronic energy changes 595
- by selenolate 598
- catalytic cycles, summary 590
- Gibbs energy barriers, summary 592, 593
- proton transfer 600
- schematic reaction mechanisms 593, 594
- thiol exchange reactions 596, 600
- 1,3 hydrogen shift, *see* tautomerization
- hydrolase enzymes 757
- hydroquinone, atomic properties 370, 371
- i**
- imidazole framework 598
- independent atom model (IAM) 429
- indoleamines dioxygenase (IDO) 98–100
- alternative mechanistic pathway 98–100
- potential energy surface 100
- tryptophan dioxygenase (TDO) 99
- dioxygen activation process 99
- insulin 36–39
- KEM 36–39
- calculations, comments 38, 39
- chain, KEM Hartree–Fock energies 36
- integral equation formalism polarizable continuum model (IEFPCM) 650, 651, 797, 815, 816
- integrated molecular orbital + molecular orbital (IMOMO) method 599
- calculation 599
- interactions
- types 140
- potential 141, 143
- intermolecular bond critical points 373, 379, 380
- molecular graphs 379
- properties 373, 380
- interpret structure-activity relationships 769
- physical parameters, computation of 769–772
- intramolecular chemically initiated electron-exchange luminescence 118
- intrinsic reaction coordinate (IRC) axis analysis 204, 505, 634, 637
- in vivo* anti-tumor activity 738
- ionized system
- Ramachandran surface 237
- N-glycylglycine conformations 235
- representative structures, optimized geometries 235
- O-methylhydroxylamine 187
- dissociation reaction, energy profile 187
- iron complexes, benchmark 563
- iron-porphyrin 572
- axial chloride ligand 572
- Collins' Fe(IV)-complex 572
- iron-porphyrazine-chloride, spin ground-state 571
- iron-porphyrin 560, 571, 572
- axial chloride ligand 572
- Collins' Fe(IV)-complex 572
- iron-porphyrin-chloride 571
- iron-porphyrin-imidazole (FePorIm) system 568, 569, 572–574
- Fe-ligand distances 573
- spin-state splittings 569, 573
- isodensity polarizable continuum model (IPCM) 596
- isoelectronic side chains, rotamers 407
- isomerization reaction 190, 191, 233
- catalyzed by 4-oxalocrotonate tautomerase 190
- k**
- kernel energy method (KEM) 22, 25, 29, 36, 46, 47, 49–53
- approximation 50
- fourth-order 51–53
- X-ray crystal structure 52
- calculation 49
- comparison 30, 31
- fourth-order 50–53
- interaction energy 49
- results 36–41
- time comparison 29–31
- quantum models 29–36
- use 46

kernel 14–17, 21, 27
 – idea, applications 17–22
 – calculations 11, 14
 – comments 21, 22
 – definition 23
 – density matrices 14–17, 22–55
 -- biological activity 24
 -- KEM 24–29
 -- peptide 24
 – double 23, 24
 -- interaction energies 43
 – total energy 27
 kinetic isotope effect (KIE) 530
 Klopman–Salem equation 884
k-nearest neighbor method 708
 Kohn–Sham orbitals 478, 505, 520, 558
 Kohn–Sham approximation 362
 Kohn–Sham highest occupied molecular orbital 370
 Kohonen nets 687
 Kolmogorovian theory of probability 892
 Koopman's theorem 713

I
 Laboratory of Molecular Structure and Spectroscopy (LMSS) 877, 891
 – projects 877
 Lagrangian approach 520
 – kinetic energy densities 702
 – multipliers 7
 latent variables (LVs) 679
 Lee–Yang–Parr correlation functional 502, 650
 Lewis acid–base model 362, 458
 linear combination of atomic orbitals (LCAO) 876, 877
 – approximation 876
 linear-free energy relations (LFERs) 669
 – isolated molecule, intrinsic features 669
 linear regression equation 448
 linear regression models 450, 454
 linear response approach 165
 Lineweaver–Burk double reciprocal plot 766
 link atoms (LAs), approach 63, 188
 lipopolysaccharides (LPSs) 658
 – components 658
 – membrane 658
 – metabolic pathway 658
 liquid systems, computer simulations 140
 – interaction potentials 140–142
 – properties 140
 liver alcohol dehydrogenase (LADH) 647
 localized orbitals (LOs) 133, 134
 – degree of transferability 134
 locally dense basis set approach (LDBS) 495

local virial theorem 435
 low-barrier hydrogen bond (LBHB) 647
 – proton transfer 648
 Lowdin's theory 851
 LpxC deacetylase, amide bond hydrolysis 659
 LpxC metalloenzyme deacetylation 658
 – acid–base catalytic mechanism 658–660
 – *N*-methylacetamide deacylation, potential energy surface 660

m

machine learning method 688
 – algorithms 710
 -- genetic 688
 – partial least-squares 688
 macrocycle, properties 560
 major histocompatibility complex–peptide (MHC-P) 416–418
 – class II-peptide complex 415, 417
 -- peptide binding region (PBR) 415
 – interaction 414
 – quantum study, diagram 416–418
 maleic anhydride 10
 – electrons per atom 10
 – energies 10
 Marcus ET theory 164, 647
 Max Born's probabilistic interpretation 875
 Mayr's two-step process 861
 metallo β -lactamases (M β Ls) 607, 609, 616
 – chemical feature 617
 – computational details 608
 – DCH/DHH site 616
 -- role 616
 – di-Zn B1 M β Ls 617
 – di-Zn/mono-Zn based M β L hydrolysis 617
 -- DCH/DHH-bound Zn₂, role 617
 – enzymatic reaction mechanism 607
 – GOB M β L 616
 – metal occupancy 609
 – Michaelis complexes 616, 617
 -- B2 mono-Zn M β L subclass 616
 -- B3 M β L subclass 616
 – metal site conformations 607
 – rate-determining steps 609
 – structural information 607
 -- atomistic structures 607
 -- DCH site 607
 -- DHH site 607
 -- 3H site 607
 -- X-ray crystallography 607
 – subclasses 605, 616
 – theory vs. experiment, preliminary comment 609

- mean absolute deviation (MAD) 562
- mean absolute percent deviation (MAPD) 391, 394, 398
- methyl elimination (ME) models 397
- molecular electrostatic potential (MEP) maps, calculations 133
- METAGGA scheme 558
- meta-heuristic method 705
- metal drugs, metabolism 739
- metalloenzymes 85, 91–108, 564, 650, 658
 - active site 564
 - cobalamin-dependent enzymes 105–108
 - methylmalonyl-CoA mutase 105–108
 - computational strategies 86–90
 - active-site model 88
 - QM/MM methods 88–90
 - LpxC 658
 - reaction mechanisms 85
 - heme-containing enzymes 91–105
 - heme oxygenase 95–97
 - indoleamines dioxygenase 97–100
 - nitric oxide synthase 101–105
 - tryptophan dioxygenase 97–100
- metalloporphyrins 560
- metalloprotein 781
- metal organic chemical vapor deposition (MOCVD) 585
- methane-benzene complexes 381, 382
 - atomic properties, relative values 382
- methane monooxygenase (MMO) 537
- methyl elimination (ME) model 392, 395
- methyl gallate (MG) molecule 372
 - caffeine adduct 371–375
- Michaelis complex 610, 611
 - architecture 611
- Michaelis–Menten kinetics 766
- Miertuš–Scrocco–Tomasi code (MST) 148
- Miller–Urey synthesis 201
- minimum energy conical intersection (MECI) 119
 - derivative coupling vector (DCV) 119
 - gradient difference vector (GDV) 119
- minimum energy pathways (MEPs) 171, 172, 204
 - algorithms for 172
- minimum energy reaction pathway (MERP) 186, 505
- model systems 64, 74
 - calculation, MM bonded terms 74
- modified neglect of differential overlap (MNDO) 694
- modified partial equalization of orbital electronegativity (MPEOE) method 391
 - (μ -OH)bis(μ -acetato)-bridged complexes 539, 540
 - broken symmetry calculations, results 540
- Møller–Plesset perturbation theory 310, 694, 726
- Møller–Plesset polarizabilities 391
- Mössbauer simulation 543
- Mössbauer spectroscopy 560
- molecular calculations 27, 134
 - *ab initio* approach 134
- molecular complementarity 456–461
 - Lewis 457
 - van der Waals 457
- molecular complex 887
 - physicochemical properties 887
- molecular correlation kinetic energy 483
- molecular dynamics technique 649
 - simulations 157, 532
- molecular electron density lego assembler (MEDLA) 696
- molecular electrostatic potential (MEP) 132–137, 492, 493
 - as component of intermolecular interaction 134
 - Coulomb interaction term, definition 135
 - E_{es} expression, simplifications 135, 136
 - atomic charges 136
 - multipolar expansions 136
 - point charge descriptions 135
 - isovalue envelopes 493
 - semiclassical approximation 133
 - use 133
- molecular maps 686
- molecular mechanical description 142, 649
- molecular mechanics methods XVIII, 72, 403
 - bonded terms 73
 - energy components 75
 - energy function 72
- molecular orbital 132, 873, 875, 882, 893
 - approach 878
 - approximation, role 882
 - contributions 132
 - dispositions 873
 - holistic model 875
 - predictive structures 873
- molecular orbital self-consistent field (MO-SCF) approach 873
- molecular quantum holism 882
 - philosophical implications 882–893
 - molecular landscapes and process 882–885
 - predictive structures 886–893

- realism of disposition 886–893
- molecular quantum self-similarity measures (MQS-SMs) 702
- molecular quantum similarity (MQS) 698
- molecular quantum similarity measures (MQSMs) 687, 688, 695
- molecular similarity index, *see* Carbó index
- molecular spin states, Kohn–Sham energies 541
- molecular structure theory, factors 133
- molecular virial theorem 481, 482
- molecular visualization tool, EVolVis 461
- molecular wavefunctions 131, 891
- *ab initio* calculation 131
- Monte Carlo multiple minimum (MCMM) conformational search 227, 234
- Monte Carlo pseudo-experiments 705
- Monte Carlo simulations XVI
- Morokuma's integrated method 148
- mRNA 424, 457
 - genetic code 457
- Mulliken analysis 288
- Mulliken charges 277, 287, 288, 293, 711, 714, 817
 - distributions 817
 - DNA base-gold complexes, PAs and DPEs 287
- Mulliken-derived electrostatic multipoles 416
- Mulliken population analysis 15, 729
- multi-center expansion, adoption 137
- multicenter proton relay mechanism 485
- multiple linear regression (MLR) 705, 713
 - based model 713
 - equation 705
- multipolar coefficients 136
- multi-reference methods 86
 - multi-reference perturbation theory (MRPT) 86
 - multi-reference self-consistent field (MRSCF) 86
- mutations 838, 845, 848, 852, 855
 - causes 848
 - Lowdin model 852
 - order 855–857
 - importance 856
 - quantum indeterministic basis 845–853
 - aqueous thermal motion 852, 853
 - proton tunneling 849–852
 - tautomeric shifts 845–849
 - role 855
 - Watson–Crick model 848
- MutY catalysis mechanism 530
- crystal structure, data 531
- Mycobacterium tuberculosis* 525
- myoglobin 94
 - photodissociation, schematic presentation 94
- n**
- N-acetylproline amide (NAP), ball stick model 157
- n*-alkanes, bond dissociation enthalpies 476
- N*-allylminium ions 75
 - C–N bond activation reaction, oxidative addition step 75
- nanobiotechnology, applications 250
- nanodimensions 246
 - DNA 248–253
 - gold 246–253
- nano-sized gold clusters 247, 248
 - catalytic activity 247
 - chemical reactivity 248
- nanotechnology, types 248
- naphthalen-1-yl(10*H*-phenothiazin-10-yl) methanone 767
 - Lineweaver–Burk plot 767
- natural bond orbital (NBO) method 134, 261, 596, 730
- natural nucleobases, structure 327
- natural population atomic (NPA) charge 261
- natural selection 838, 854, 857, 860
 - nature 857–863
- natural sesquiterpenic lactones 626
 - biogenetic hypotheses 626
 - double bonds, configurations 626
- near-attack conformation (NAC) hypothesis 646
 - substrate, Gibbs energies 646
- Nernst equation 796
- neurodegenerative disorders 760
 - Alzheimer's disease, treatment 760
- neurotoxic inflammatory 750
- neurotransmitter dopamine 787
- neutral selenol, interconversion 593
- Newton's laws 245
- Newton trajectory (NT) 174
- N*-glycoside hydrolysis 530
- N*-glycylglycine 236
- N*^ω-hydroxy-L-arginine (NHA) 101, 104
- N*^ω-hydroxy-L-arginine, substrate model 652, 653
- nickel complexes 574, 575
 - high-spin 574
- nickel-ligand distances 575
- NiFe hydrogenase 555, 574–576
 - catalytic cycle 575, 576
 - model 574

- nitric oxide, formation 101
 - nitric oxide synthase (NOS) 103, 104, 652, 653
 - catalytic mechanism 653
 - half-reaction, mechanism 103, 104
 - nitrogen oxide (NO) 652
 - effects 652
 - oxidation, potential energy surface 654
 - produced by nitric oxide synthases 652
 - N*-methylacetamide (AcNMe) 658
 - mechanistic pathways 658
 - oxidation 101
 - non-cisplatin drugs, properties 732
 - noncovalent interactions 132, 307, 308
 - computational approaches 308–314
 - non-DNA drug targets 734
 - nonnuclear attractors (NNA) 434
 - non-steroidal anti-inflammatory agents/analgesics (NSAIDs) 808
 - non-steroid anti-inflammatory drugs (NSAIDs) 807–823, 826, 827, 829
 - absorption spectra 820–823
 - acetylsalicylic acid 810
 - classification 808
 - definition 808
 - highest occupied molecular orbitals (HOMOs) 817
 - indomethacin 810
 - lowest unoccupied molecular orbitals (LUMOs) 817
 - orbital structures 817–820
 - pharmacological action 808, 809
 - redox chemistry 815–817
 - side effects 810, 811
 - uses 809, 810
 - pain relievers 807
 - photodegradation mechanisms 812, 813, 823, 827
 - steps 823
 - phototoxicity 811, 812
 - theoretical studies 812–815
 - chemical structures 813
 - computed dipole moment 820
 - decarboxylation/dechlorination 816
 - C–C bond lengths 816
 - C–Cl bond lengths 816
 - deprotonated species, absorption spectra 821
 - excited state reactions 823–827
 - singlet excited states, photodegradation 826, 827
 - T1 state, photodegradation 825, 826
 - neutral species, absorption spectra 821
 - photodegradation mechanisms 826
 - reactive oxygen species (ROS) 827, 828
 - suprofen (SUP) 829
 - non-zwitterionic amino acid models 405
 - N*-10-phenothiazine amides 762–764
 - inhibition constants 762
 - molecular volumes 762
 - N*-10-phenothiazine carbamates 765, 766
 - deactivation constants 765
 - inhibition constants 765
 - molecular volumes 765
 - N*-representability problem 4, 5
 - N*-representable matrix 18
 - nuclear DNA oxidation adducts 791
 - nuclear magnetic resonance (NMR) shifts 269
 - nuclear-nuclear repulsion potential energy 479
 - nucleic acid bases 423, 446, 447
 - calculated vs. experimental partial molar volumes 446, 447
 - nucleic-acid system 248
 - nucleobases 316
 - amino acid T-shaped interactions 325
 - hydrogen bonding 316
 - methylation 327
 - nucleophilic agent, HOMO 610
 - nucleophilic hydroxide 525
 - mechanism 526
 - oxygen atom 525
 - nucleophilic reaction, simulations 533
 - nucleophilic water addition 530
 - nucleosides, partial volume 445
 - nucleotide excision repair (NER) 517
 - nudged elastic band (NEB) method 172
- O**
- octanol-water partitioning coefficient 681
 - oligonucleotide-directed mutagenesis 451
 - olvatochromism 162
 - one-electron density matrix 339–343, 361
 - Onsager–Lorentz theory 159
 - ontological potentiality concept 889
 - OPBE, reliability 562
 - optimized H-bond distances 223
 - organo-selenium therapeutic agents 586, 596, 597
 - own *N*-layer integrated molecular orbital molecular mechanics (ONIOM) 61, 63–65, 76, 89, 649, 731
 - application, guidelines 65–72
 - cancellation problem 72–76
 - energy expression 63
 - potential energy surface 76

- scheme 64, 649
- components 64
- three-layer methods, schematic diagrams 89
- oxaliplatin 723, 725
- oxidase iron enzymes, catalytic mechanisms 652
- oxidative stress, 588, 791
 - human body, defense mechanism 588
 - metabolic signs of 791
- oxygen-coupled electron-transfer (OCET) 104

p

- parallel dissociation process 154
- para-substituted phenols 681
 - biodegradability/toxicity 681
 - molecular skeleton 682
 - substituents/properties 682
- Pariser–Parr–Pople (PPP) method XVII
- Parkinson's disease 743
- partial least-squares (PLS) procedure 679, 688, 706
 - regression method 706
 - use 679, 686
- partial molar volumes 439, 445, 446
 - experimental vs. calculated group contributions 445
 - principal contributions 440
- partitioning coefficient 683
- pattern recognition techniques 671
- Pauli exclusion principle 138
- Pauling configurations 93
- P450 enzymes 565, 566, 567
 - catalytic activity 567
 - hydrogen-abstraction chemistry 566
- peptide(s) 26, 28, 30, 234
 - bond formation pathway 508
 - bond transition 506
 - calculation time 30
 - energy calculation 26, 28
 - ionization 234–239
 - *N*-glycylglycine, ionization 234–236
 - Ramachandran maps, ionization, influence 236–239
 - structures 25
- peptide-bound copper (II) 781
 - electrostatic effects 783
 - enthalpic component 782
 - ligand 782
 - polarity 783
 - reduction potentials 781
- peptide-host interactions 414
 - quantum mechanical studies 414–419
- peptidyl transferase center (PTC) 501
- perturbation theory (PT) approaches 134, 505, 884
 - symmetry-adapted 134
 - twofold symmetry 505
 - use 134
- phenothiazine 761, 770, 772, 776, 778
 - AChE inhibition 778
 - amides 771, 778
 - 4-biphenylcarbamate 776
 - B3LYP/6-31G(d) level 770
 - butterfly angles 770
 - carbamate derivatives 775, 777
 - action 776
 - inhibitory properties 777
 - 3-*N,N*-dimethylaminophenyl derivative 775
 - structure-activity comparison 777
 - *N*-10-carbonyl derivatives 761
 - scaffold moiety 760, 770, 772
 - structure-activity relationships 761, 778
 - substituent size 771
 - synthesis 761
 - tricycle 773
 - ring system 769
- phenyl alkylamine hallucinogens 698
 - drug-receptor correlations 698
- phosphate backbone 317
 - hydrogen-bonding interactions 317
- photoactivatable fluorescent proteins (PAFPs), category 109
- photoactivation mechanisms 110
- photobiology 85, 109–119
 - fluorescent proteins (FPs) 109–117
 - photoconversion 115–117
 - green fluorescent proteins (GFP) 110, 111
 - reversible photoswitching fluorescent proteins (RPFs) 111–115
 - luciferases 117–119
- photodegradation mechanism 814, 828
 - ROS effects 828–830
- photosensitivity reactions 806
 - definition 806
 - pathways 806
 - photoallergic 806
 - phototoxic 806
- phototoxic response pathways 812
- physicochemical variables 404–408
- π -cation interaction 758
 - aromatic ring system 758
- ping-pong mechanism 105
- platinum
 - anti-cancer drug, diaminocyclohexane (DACH) 723

- bonding interactions 734
- chloroqua complexes 729
- DNA adducts, formation of 725
- drugs 723
- development 723
- structures 724
- moiety 733
- Poincaré-Hopf relationship 509
- topological relationship 506
- point charges 77–81
- use 77
- point mutation 846, 853
- tautomeric shift 846
- theory 853
- Poisson-Boltzmann equations (PBEs) 113, 145, 150
- self-consistent reaction field model (SCRF) 650
- polarizability effects 162, 389, 885
- decomposition methods 392
- definition 389
- models 389–392
- polarizable continuum model (PCM)
 - method 142, 143, 145–147, 150, 151, 204, 210
 - *ab initio* versions 151
 - applications 150–165
 - chemical equilibria 152–154
 - electronic transitions 162
 - energy transfers 164
 - environment on formation 161
 - PES 152
 - photoinduced electron 164
 - reaction mechanisms 154–156
 - relaxation of excited states effect 161
 - solvation energies 150–152
 - solvent effects on molecular properties/spectroscopy 156–161
 - spectroscopies 162
 - approach 164
 - codes 151
 - C-PCM/D-PCM 147
 - formulation 144
 - framework 160, 165
 - Hamiltonian 142
 - IEF-PCM scheme 147
 - integral equation methods 145
 - PCM-ZINDO version 147
 - solvation model 323
 - use 147
 - versatility 146
 - versions 146
- polarization 63, 89, 90, 133, 134, 138, 140, 141, 145, 148, 149, 151, 159–161, 163, 220, 276, 374, 380, 382, 383, 385, 406, 408, 410, 412, 554–556, 558, 650, 651, 656, 727, 729
- polychlorinated dibenzo-*p*-dioxins (PCDDs) 685
- polycyclic aromatic hydrocarbons (PAHs) 696, 712
- electronic structure of XVII
- polymer chains 714
- conformational properties 714
- polyunsaturated fatty acids (PUFAs) 790
- pople basis set 565, 592, 593
- GTO basis sets 555
- population genetics 843
- potential energy surface (PES)
 - reaction XXIII, 87, 100, 152, 172, 186, 204, 221, 312, 320, 635, 650, 651, 727
- concept XIV
- energy-cost surface 185
- four-well analytical, parameters 182
- energy cost surface, MEP on 183
- potential value, isosurface 186
- spin-singlet 653
- potential of mean force (PMF) 189, 521
- pragmatic transcendental approach 873
- prebiotic chemistry 199–201
- prebiotic compounds 201
- precursor, HCN 201
- prediction of acidity constant (pK_a) 706
- primordial conditions 200
- principal components analysis (PCA) 679–681, 704
- ANN approaches 712
- SIMCA-P package 681, 682
- principle of least action, expression 361
- prion diseases 781
- peptide-bound copper (II) 781
- reduction potentials of 781
- prion protein 787, 789
- copper binding 789
- octarepeat region 787–789
- reduction potential 787
- protein backbone 316
- hydrogen bonding 316
- protein chains triplet 47
- interaction energies 47
- protein data bank (PDB) 42, 315, 321, 414, 518, 751
- accession code 531
- protein folding 744, 745
- protein misfolding 743, 745
- Alzheimer's disease 747–750
- beta-amyloid 748–750
- neurodegenerative disease 747
- neuropathological hallmarks 748

- disorders 743, 745
 - quantum biochemistry 745
 - drug design 745, 750–753
 - molecular mechanics 746, 747
 - protein stability 451
 - genetic mutation, effect 451
 - protein synthesis 502
 - production line 502
 - proton acceptor 282, 286
 - gold atom 281
 - proton affinities (PA) 286
 - proton-coupled electron-transfer (PCET) 104
 - proton interactions 139
 - proton shift mechanism 594
 - proton shuttle mechanism 512, 513
 - proton transfer reactions 222
 - proton tunneling 851
 - Lowdin's mechanism 851
 - pseudoguaianolides 625
 - biogenetic origin 624, 639
 - generation 625
 - terminal biogenesis 628
 - transformation mechanism 639
 - pseudo-reversible cholinesterase inhibitors 769
 - P-site sugar 513
 - ribose sugar 509
 - π -systems 878, 884
 - molecular orbital 878
 - reactivity 884
 - pyrazine-pyridine biheteroaryls 708
 - pyrimidines, KP-photoinduced dimerization 829
- q**
- quadratic string method (QSM) 172, 173
 - PES, local quadratic approximation 173
 - quantitative structure-activity relationship (QSAR) models 404, 407, 425, 659, 674, 676–678, 693, 694, 698, 703, 710, 713–715, 733
 - approaches 712
 - biochemistry and molecular biology 710–712
 - 2D/3D QSAR 669
 - descriptors 678, 688, 710
 - quantum-chemical 710
 - selection 703–705
 - drug design 712–714
 - linear regression techniques 705, 706
 - machine-learning algorithms 706–710
 - material and biomaterial science 714, 715
 - mathematical technique 706
 - medicinal chemistry 712–714
 - polymeric materials, comparison 715
 - power/weakness 670
 - surging prominence 685
 - use of 714
 - validation 680
 - VIP plots 683
 - quantitative structure-property relationships (QSPR) models 407, 693, 694, 698, 703, 710, 713–715, 733
 - approaches 712
 - biochemistry and molecular biology 710–712
 - drug design 712–714
 - material and biomaterial science 714, 715
 - mathematical technique 706
 - medicinal chemistry 712–714
 - polymeric materials, comparison 715
 - quantum-chemical descriptors 710
 - use of 714
 - quantum approaches, evolution 876–882
 - quantum biochemistry (QB) methods XVII, 3, 133, 746, 841, 887
 - approximations 890
 - introduction XI–XXVI
 - spectrum 746
 - quantum biology 131
 - quantum chemical methods 554, 570, 694–697, 725, 879, 885
 - calculations 202, 331, 599
 - descriptors 695, 697–699, 703, 705, 713, 714
 - *ab initio* methods 695
 - classification of 697–703
 - parameters 693
 - use 879
 - quantum chemical topology (QCT) 670, 688, 689
 - descriptors 688
 - quantum computations 3
 - quantum-confined electronic transitions 247
 - quantum crystallography (QCr) 4, 13, 21, 501
 - comments 21, 22
 - KEM method XVIII
 - mathematical objective 13
 - origins 4–10
 - quantum electrodynamics (QED) theory 165
 - quantum fingerprint 671
 - quantum framework 890
 - quantum indeterminism 840, 853, 863
 - quantum jump 845, 850
 - quantum kernels 10
 - beginnings 10–22
 - formalism 11–14

- quantum mechanics/molecular mechanics
 - (QM/MM) schemes 61–63, 112, 132, 136, 142, 163, 171, 187–189, 192, 339, 403, 518, 585, 586, 599, 600, 608, 648–650, 841, 876, 879, 888, 889, 892
 - approach 213
 - Bohr’s interpretation 888
 - boundary 90
 - calculations 88, 90, 93, 102, 132, 135, 140, 502, 559, 560, 565
 - chemistry models 32
 - development 171, 188
 - expectation value 438
 - free-energy perturbation (FEP) method 189
 - Hamiltonian 519
 - Hilbertian structure 889
 - implementation 519
 - interaction energy 89
 - minimum free-energy path (MFEP) method 189
 - modeling 9
 - nondeterministic nature 888
 - partitioning scheme 730
 - potential, energy surface 72
 - procedure 142
 - region 567
 - semiempirical approach 151
 - simulations 523, 529, 532, 611
 - Endo IV 529
 - MD, 528, 734
 - structure 879
 - subsystem 188, 190
 - free-energy gradients 190
 - transition state 506
 - types 188
 - use 876
- quantum methods 34, 879, 882
- quantum models 874
 - aggregate 874
 - composite 874
 - origins 874–876
- quantum molecular similarity (QMS)
 - method 676
 - measurement 688
- quantum probabilities 840
- quantum system 670
- quantum theory 850, 888, 892
 - of probability 892
 - probabilistic algorithm 892
 - statistical algorithm 888
- quantum theory of atoms in molecules (QTAIM) methods XXII, 339, 344, 346, 360, 365, 366, 374, 404, 410, 415, 430–438, 462, 474, 477, 505, 594, 696, 735
 - analysis 370, 372, 484, 737
 - atomic moments, electrostatic potential 437
 - basic concepts 430
 - electron density analysis 367
 - framework 477
 - group energy 494
 - multipoles 438
 - properties 377, 383, 462
 - utility 462
- quantum topological molecular similarity (QTMS) 670, 676, 681, 685, 687, 697
 - ability 689
 - applications 679
 - chemometrics 679–681
 - computational modules 680
 - descriptors 686, 687
 - development 669, 672
 - equilibrium bond lengths 678
 - feature 675
 - general reflections 687
 - hopping center of action 681
 - hypothesis 683
 - leap 684–687
 - para-substituted benzoic acids generation 676
 - physical organic chemistry, anchoring in 671–678
 - steroid set 685
 - study 677, 686
 - work 686
- quantum transition, *see* quantum jump
- quartet-doublet splitting 557
- quartet-sextet splitting 555
- QUILD program 557, 579
- quinhydrone 367–369
 - intermolecular BCPs, properties 369
- molecular graph 369
- stacking energies 368
- quinone 370
 - atomic properties 370
 - hydroquinone complex, electron density transfer 371
- r**
 - radial basis function (RBF) kernel 708
 - neural network (NN) 708
 - radical-rebound-type mechanism 656
 - Ramachandran maps 237
 - Raman bands 732
 - Raman scattering 160
 - Raman spectrum 733
 - rational drug design 43, 44
 - interaction energy, importance 43, 44

- reaction coordinate (RC) 608, 613
 - constraint 613
- reactive oxygen species (ROS) 587
- real system, low-level calculation 64
- rebound pathway 564
 - vs. cationic pathway 566
- redox-active blue copper proteins 794
- redox catalytic mechanisms 652–657
 - AlkB family, oxidative dealkylation 654–657
 - nitric oxide synthase, NO formation 652–654
 - first half-reaction 652
 - second half-reaction 652, 654
- redox enzymes, catalytic mechanisms 661
- reduced gradient following (RGF) method 173
 - curves 174
- reduction scale spectrum 787
- relaxed spin-state splittings 554, 559
- RESP charges, sets 79
- RESP program 78
- restricted RCCSD(T) method 220
- ribonucleic acid (RNA) nucleobases 309
 - atomic numbering 309
 - structure 309
- ribosome 501, 505, 645
 - crystallography 501
 - peptide bond formation, transition state 645
 - ribonucleic acid (RNA) 245
 - 30S ribosomal subunit, aminoacyl site 44–46
 - thermodynamic parameters 505
 - tRNA 503
- ribozyme catalysis 501
- ring critical points (RCP) 434
- rotamers 404–408
- ruthenium complexes 736, 737
 - bond dissociation energies 737
 - treatment of cancers 738
- S**
- scaled hypersphere search (SHS)
 - methods 173
- scaled particle theory (SPT) 144
- Schizosaccharomyces pombe* 529
- Schlegel's synchronous transit-guided quasi-Newton (STQN) method 592
- Schrödinger equation XIII, 144, 688
- Schrödinger's wave-mechanical formalism 891
- Schwinger's principle 338, 435
- secondary α -deuterium kinetic isotope effect (sec-KIE) 97
- secondary interaction hypothesis (SIH) 343
- secondary metabolites 623
 - functions 623
 - structures 623
 - as taxonomic markers 623
- selenium 585, 587, 597, 598
 - biochemical applications 585
 - glutathione peroxidase (GPx) mimics, computational studies 589–600
 - chemical properties 586
 - discovery 585
 - electronic energy barriers 597
 - quantum mechanical approaches 585
 - role 585, 586
 - steric effect 598
 - vs. sulfur 586
 - treatment, quantum mechanical methods 586
- selenoenzyme 590
- selenol anion reaction 592
- selenoxide oxygen 595
- self-consistent reaction field (SCRF)
 - method XIX, 149, 650, 651, 729, 733, 736
 - solvation models 737
- self organized maps, *see* Kohonen nets
- semi-dynamic approach 152
- sesquiterpenes 623
 - reaction mechanism 627–639
 - synthetases enzyme 625
 - cyclic compounds, generation 625
 - terminal biogenesis, computational simulation 623
 - 8-epiconferin, case 623
- Shannon's sense, definition 423
 - information 424
- sigmoidal dielectric function 147
- simple linear equation 454
- single-determinant approach 13
 - *N*-representability 5–7
- single point mutations 450
 - $\Delta\Delta H$, correlation 455
- single proton transfer reactions 731
- singlet-quintet splittings 562, 570
- singlet-triplet splitting 556
- singlet-triplet states 576
- site-directed mutagenesis 98
- Slater-type orbital (STO) basis sets 553
 - ADF program 558
 - 3G basis functions 34
- SMx framework 149
- S_N2 reaction mechanism 184, 185, 190
 - energy profile 185
 - PES 184
 - in solvent 190

- solute accessible to the solvent (SAS) 150
- solute-solute interaction 439
- solute-solvent system 160, 161
- solvated system 131, 142–144, 146, 150–152, 157
 - continuum model 142–150
 - continuum solvation methods 148–150
 - development 157
 - free energy 149
 - from molecular electrostatic potentials 131
 - Hamiltonians 143
 - insulin, energy calculation 38
 - interaction, components 144
 - PCM, basic formulation 142–148
 - apparent charges, definition 147
 - cavity surface 147
 - dielectric function 146
 - solute description 147
 - with biomolecular photophysical processes 131
- solvatochromic shifts 162
- solvent effects 157, 160, 162
 - dynamic effects 160
 - kinetic isotope effect 97
 - polarization, assumption 160
 - reorganization energy 164
 - role 157
- somersault pathway 565
- space-filling density 436
- spin contamination 220
 - corrections 556–558
- spin-crossover phenomena 563
- spin density functional theory (SDFT) 537, 538
 - calculations 544
- spin states 551, 561
 - accurate description 551
 - basis set, influence 553–556
 - contamination 557
 - model complexes 559–564
 - self-consistency, influence 558
 - splittings 552, 553, 555, 575, 577–579
 - ECPBs, use 553
 - validation studies, iron complexes uses 561
 - vertical vs. relaxed 553
- stacking interactions 322, 365
 - computational method 366, 367
 - orientations 322
- standard hydrogen electrode (SHE) 781, 795
- staphylococcal nuclease stability 453
 - experimental vs. calculated change 453
- statistical mechanics 143
 - molecular partition function 143
- stepwise multiple regression (SMLR) 705
- Stevens, Basch and Krauss (SBK)
 - pseudopotential schemes 726
- Stuttgart–Dresden pseudopotentials 732
- substituent ring system 773
- substrate ester hydrolysis 768
- support vector machines (SVMs) 707
 - algorithm 708
 - schematic presentation 707
- surface and volume polarization for
 - electrostatic (SVPE) 148
- surface walking algorithms 173
- t**
- tautomeric shifts 850
 - roles 849
 - Watson–Crick model 850
- tautomerization 212
 - keto-enol 208
- taxonomic markers, 623. *see also* secondary metabolites
- T-cell receptor (TCR) molecules 414
- terpenes, origin 624
 - biogenetic hypothesis 624
- tesserae 145, 146
- thermodynamic integration 518, 526, 533
- thiocholine hydrolysis 761
 - kinetic analysis, using Ellman method 761
- thymine dimer 522
 - radical 522
 - splitting, catalyzed by DNA photolyase 521
- thymine-gold (T-Au) interactions 268–271
 - basic features 269
 - bond lengths 270, 271
 - hybridization 271
 - three conformers 268
- time-dependent B3LYP method 113
 - calculations 94
- time-dependent density functional theory (TD-DFT) 156, 158, 164, 394, 820
 - calculations 824
 - method 87
- transferability 337, 342
 - short-range nature 342, 343
- transferable atom equivalent (TAE)
 - method 697
- transition-metal complexes 559, 563, 579
- transition state (TS) model 501, 509, 510, 527, 613, 615, 628, 634, 637, 883
 - asynchrony 633
 - B3LYP/6-31+G(d,p) level 635
 - molecular graph 510
 - MP2/6-31+G(d,p) level 635
 - nature 627

- structure, SMM energies 76
- sugar moiety 505
- {1-2} transference of hydrogen 638
- transmissible spongiform encephalopathies (TSEs) 787
- transmission coefficient 648
- triarylmethane (TRAM) derivatives 711
- tricyclic ring system 770
- tRNA 41, 42, 503–505, 507, 511, 512
 - acceptor stem mimic (ASM) 504
 - analog 504
 - A-site 503, 511
 - crystal structure 42
 - ester carbonyl group 512
 - KEM 41–43
 - non-hydrolysable nitrogen 504
 - P-site 503, 505, 511
 - quantum mechanical molecular energy 41
 - stems 503
 - 1YFG picture 42
 - crystal structure 42
 - energy calculation 42
- tryptophan dioxygenase (TDO) pathway 98–100
 - alternative mechanistic pathway, 99, 100
 - potential energy surface 100
- T-shaped interactions 312, 313, 318, 329
- potential energy surface scans 313
- twisted intramolecular charge transfer (TICT) state 114
- two-body interaction operator 142
- tyrosine 782
 - amino acid structure 504
 - residue 551, 566

u

- ultraviolet (UV) radiation 806
 - UV-A/B 806
- united atom Hartree–Fock (UAHF) radii 797
- united atom topological model 651
- universal Darwinism 862
- universal force field (UFF) 651, 731
- unrestricted Hartree–Fock (UHF) method 556
- α,β -unsaturated ketone (*E*)-1-(4-hydroxyphenyl)but-1-en-3-one, antitumor agent 684
- unweighted pair group method with arithmetic mean (UPGMA) method 406
- uracil-glycine dimmers, hydrogen bonding 317
- uteroferrin
 - oxidized diferric form 545
 - purple acid phosphatase 545

v

- valence bond (VB) theory XIV, 881
 - advantages 881
 - approach 878
 - use 878
- valence shell charge concentrations (VSCCs) 460
- valproic acid (VPA) 713
- van der Waals envelope 491
- van der Waals interactions 524
- van der Waals molecular sizes 436
- van der Waals potential functions 746
- van der Waals radii 269
- van der Waals volume 441, 442, 444, 446, 451
- variable importance in projection (VIP) values 682
- vascular endothelial growth factor receptor-2, inhibitors 708
- vesicular stomatitis virus nucleoprotein 53
 - KEM 53–55
 - 2QVJ molecule 53, 54
 - Ser290Trp mutant (2QVJ), crystal-structure of XIX
- vibrational circular dichroism (VCD) spectroscopies 157
- vibrational Raman optical activity (VROA) 158
- viral theorem 341
- virial field 340–342, 435
 - short-range nature 342, 343
- VSEPR model 459

w

- Warshel's electrostatic hypothesis 646
- water molecule 77, 529, 614–616
 - ASP85 system 78
 - binding energy, calculation model 77
 - cisplatin complex 728
 - ligand, displacement of 793
 - proton transfer 213, 532
 - role 529, 616
- Watson–Crick (WC) model 315, 318, 352, 847, 848
 - atomic numbering 352
 - A-T pairing 292, 350
 - energy changes 350–355
 - CG/TA pairs 288, 728
 - DNA base pairs, interaction of 286
 - [A·T]·Au₃ complexes 289–293
 - Au₆ cluster bridges the WC G·C pair 296, 297
 - [G·C]·Au₃ complexes 293–296
 - general background 286–289
 - G·C duplex 294
 - intermolecular H-bonds 294

- stretching vibrational modes 292
- molecular graphs 352
- wavefunction theory (WFT) 649, 650
- weak molecular interactions theory 137
- weighted holistic invariant molecular descriptors (WHIM) 693
- Wilson's disease 782
- Woodward-Hoffmann rules 884

x

- X-ray crystallography 22
 - data 737
 - results 22
 - structure 732, 733
- X-ray diffraction 726
 - experiment 428
 - observed vs calculated diffraction pattern 428
- X-ray scattering data 8
 - experiment 21
 - N-representable density descriptions 8
- X-ray structure 525, 608, 731

z

- Zaib4 molecule 34-36

- KEM calculation 35
- quantum methods 34-36
 - calculations 34-36
- X-ray crystal structure 35
- zero-field splitting (ZFS) 542
- zero-flux surface 432, 433
- zero-memory Markov chain 424
- zero-point energy (ZPE) 729
 - electronic energies 819
 - vibrational energy 503, 815
 - -corrections 651
- zero-temperature string method 172
- zervamicin molecules 51
 - crystal structures 52
- zinc
 - enzymes 606, 609
 - catalytic mechanism 606
 - MβLs 609
 - flexibility 527
 - ligands 615
- zwitterionic structures 395
 - 3-APS 752
 - chromophores 115
 - glycine form 231
 - species 442