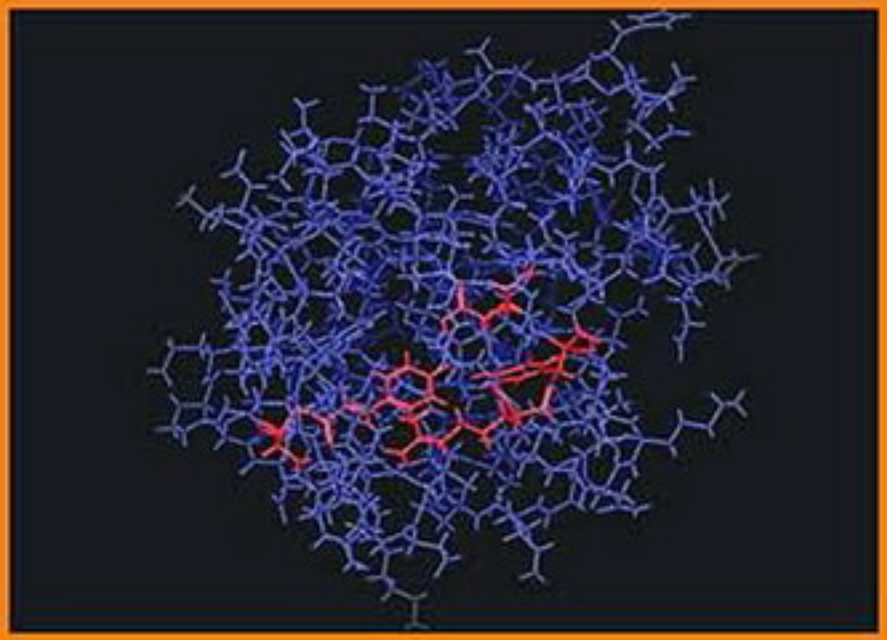
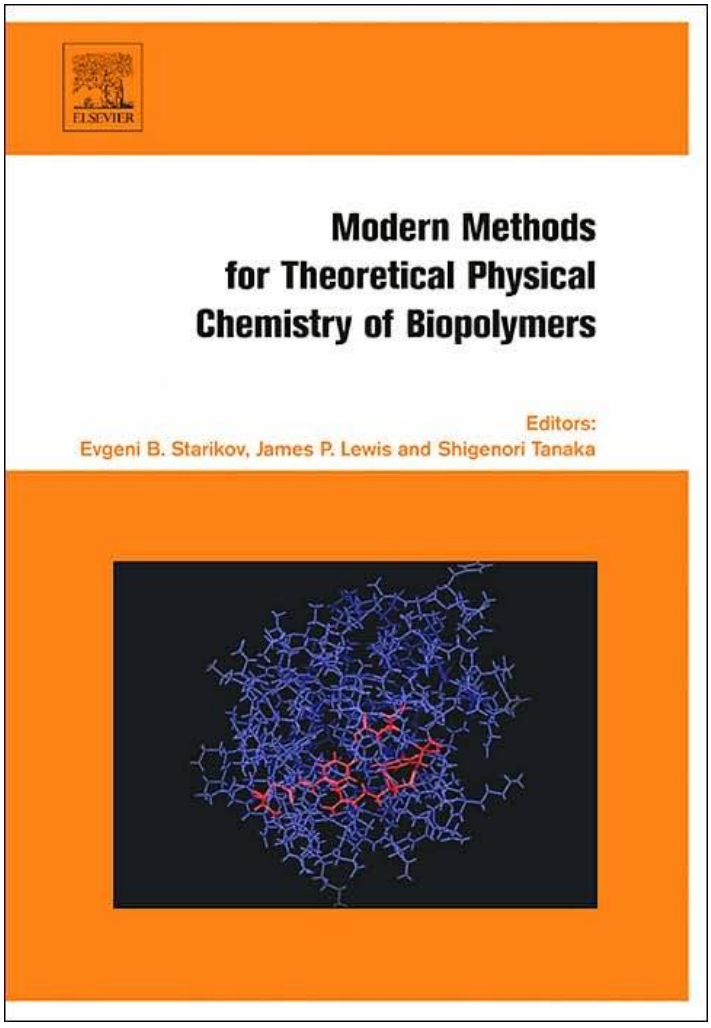
Modern Methods for Theoretical Physical Chemistry of Biopolymers

Editors:
Evgeni B. Starikov, James P. Lewis and Shigenori Tanaka



Modern Methods for Theoretical Physical Chemistry of Biopolymers

by Evgeni Starikov (Editor), James P. Lewis (Editor), Shigenori Tanaka (Editor)



- Pub. Date: August 2006
- ISBN: 0444522204
- Publisher: Elsevier Science & Technology Books

List of Contributors

Shinji Amari
Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan

J. F. R. Archilla
Group of Nonlinear Physics, Departamento de Física Aplicada I, ETSI Informática, University of Sevilla, Avda Reina Mercedes, s/n. 41012, Sevilla, Spain

David N. Beratan
Departments of Chemistry and Biochemistry, Duke University, Durham, NC 27708, U.S.A.

F. Matthias Bickelhaupt
Afdeling Theoretische Chemie, Scheikundig Laboratorium der Vrije Universiteit, De Boelelaan 1083, NL-1081 HV Amsterdam, The Netherlands

Johan Bredenberg
Karolinska Institutet, Department of Biosciences, Center for Structural Biochemistry, Hälsovägen 7, SE-141 57 Huddinge, Sweden

Ralf Bulla
Theoretische Physik III, Universität Augsburg, D-86135, Augsburg, Germany

Arrigo Calzolari
National Centre on nanoStructures and bioSystems at Surfaces (S3) of INFM-CNR, c/o Dipartimento di Fisica, Università di Modena e Reggio Emilia, Via Campi 213/A, 41100 Modena, Italy

Hiroshi Chuman
Institute of Health Biosciences, University of Tokushima, Tokushima, 770-8505, Japan

Peter Claiden
Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK

Daniel L. Cox
Department of Physics, University of California, Davis, CA 95616, U.S.A.

Tobias Cramer
Institut für Physikalische Chemie, Universität Freiburg, Albertstrasse 23a, D-79104 Freiburg im Breisgau, Germany

Jane Crawshaw
Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK

Gianaurelio Cuniberti
Institut für Theoretische Physik Korrelationen und Magnetismus, Universität Regensburg, D-93040, Regensburg, Germany

Martin Dahlberg
Division of Physical Chemistry, Arrhenius Laboratory, Stockholm University, S 106 91, Stockholm, Sweden

Owen R. Davies
School of Physics and Astronomy, Cardiff University, Cardiff, CF24 3YB, UK

James Elliott
Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK

Robert G. Endres
Department of Molecular Biology, Princeton University, Princeton, NJ 08544-1014, U.S.A; and NEC Laboratories America, Inc., Princeton, NJ 08540 U.S.A.

Dmitri G. Fedorov
National Institute of Advanced Industrial Science and Technology, 1-1-1 Umezono, Tsukuba, 305-8568 Ibaraki, Japan

Rosa Di Felice
National Centre on nanoStructures and bioSystems at Surfaces (S3) of INFM-CNR, c/o Dipartimento di Fisica, Università di Modena e Reggio Emilia, Via Campi 213/A, 41100 Modena, Italy

B. Fischer
Forschungszentrum Karlsruhe GmbH, Institut für Nanotechnologie, Postfach 3640, D-76021 Karlsruhe, Germany

Kaori Fukuzawa
Division of Science and Technology, Mizuho Information & Research Institute, Inc., 2-3 Kanda Nishiki-cho, Chiyoda-ku, Tokyo 101-8443, Japan; and CREST, Japan Science and Technology Agency (JST), Japan

Yi Qin Gao
Department of Chemistry, Texas A&M University, College Station, TX 77845, U.S.A

Heather L. Gordon
Department of Chemistry and Centre for Biotechnology, Brock University, St. Catharines, Ontario, L2S 3A1, Canada

Hitoshi Goto
Department of Knowledge-based Information Engineering, Toyohashi University of Technology, Toyohashi, 441-8580, Japan

Célia Fonseca Guerra
Afdeling Theoretische Chemie, Scheikundig Laboratorium der Vrije Universiteit, De Boelelaan 1083, NL-1081 HV Amsterdam, The Netherlands

Rafael Gutierrez
Institut für Theoretische Physik Korrelationen und Magnetismus, Universität Regensburg, D-93040, Regensburg, Germany

Michiaki Hamada
Mizuho Information and Research Institute, Inc., Tokyo, 101-8443, Japan

D. Hennig
Freie Universität Berlin, Fachbereich Physik, Institut für Theoretische Physik, Arnimallee 14, 14195 Berlin, Germany

Arnd Hübsch
Department of Physics, University of California, Davis, CA 95616, U.S.A.

Yuichi Inadomi
Grid Technology Research Center, National Institute of Advanced Industrial Science and Technology, Tsukuba, 305-8568, Japan

Yuichiro Inagaki
Mizuho Information and Research Institute, Inc., Tokyo, 101-8443, Japan

John E. Inglesfield
School of Physics and Astronomy, Cardiff University, Cardiff, CF24 3YB, UK

Masakatsu Ito
Grid Technology Research Center, National Institute of Advanced Industrial Science and Technology, Tsukuba, 305-8568, Japan

Toshiyuki Kamakura
Department of Knowledge-based Information Engineering, Toyohashi University of Technology, Toyohashi, 441-8580, Japan

Martin Karplus
Deparment of Chemistry and Chemical Biology, Harvard University, MA 02138, U.S.A; and Laboratoire de Chimie Biophysique, ISIS, Université Louis Pasteur, 67000 Strasbourg, France

Kazuo Kitaura
National Institute of Advanced Industrial Science and Technology, 1-1-1 Umezono, Tsukuba, 305-8568 Ibaraki, Japan

Thorsten Koslowski
Institut für Physikalische Chemie, Universität Freiburg, Albertstrasse 23a, D-79104 Freiburg im Breisgau, Germany

Petras J. Kundrotas
Department of Biosciences at Novum Research Park, Karolinska Institutet, SE-141 57 Huddinge, Sweden

Noriyuki Kurita
Department of Knowledge-Based Information Engineering, Toyohashi University of Technology, Toyohashi, Aichi 441-8580, Japan

Aatto Laaksonen
Division of Physical Chemistry, Arrhenius Laboratory, Stockholm University, S 106 91, Stockholm, Sweden

Victor D. Lakhno
Institute of Mathematical Problems of Biology, Russian Academy of Sciences, Pushchino, Moscow Region, 142290, Russian Federation

K. H. Lee
Korean Institute for Science and Technology, Supercomputing Materials Laboratory, P.O. Box 131, Cheongryang, Seoul 130-650, Korea

J. P. Lewis
Department of Physics and Astronomy, Brigham Young University, Provo, UT 84602-4658, U.S.A.

Jianping Lin
Departments of Chemistry and Biochemistry, Duke University, Durham, NC 27708, U.S.A.

Raimo A. Lohikoski
University of Jyväskylä, Department of Physics, FIN - 40014 University of Jyväskylä, Finland

Pekka Mark
Karolinska Institutet, Department of Biosciences, Center for Structural Biochemistry, Hälsovägen 7, SE-141 57 Huddinge, Sweden

R. Marsh
Department of Physics and Astronomy, Brigham Young University, Provo, UT 84602-4658, U.S.A.

H. Merlitz
Forschungszentrum Karlsruhe GmbH, Institut für Nanotechnologie, Postfach 3640, D-76021 Karlsruhe, Germany

Yuji Mochizuki
Advancesoft and Institute of Industrial Science, Center for Collaborative Research, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan; Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan; and CREST, Japan Science and Technology Agency (JST), Japan; Department of Chemistry, Faculty of Science, Rikkyo University, 3-34-1 Nishi-ikebukuro, Toshima-ku, Tokyo 171-8501, Japan

Umpei Nagashima
Grid Technology Research Center, National Institute of Advanced Industrial Science and Technology, Tsukuba, 305-8568, Japan

Yoshihiro Nakajima
Center for Computational Sciences, University of Tsukuba, Tsukuba, 305-8577, Japan

Tatsuya Nakano
Division of Safety Information on Drug, Food and Chemicals, National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan; Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan; and CREST, Japan Science and Technology Agency (JST), Japan

Naofumi Nakayama
Department of Knowledge-based Information Engineering, Toyohashi University of Technology, Toyohashi, 441-8580, Japan

Takayuki Natsume
Department of Knowledge-Based Information Engineering, Toyohashi University of Technology, Toyohashi, Aichi 441-8580, Japan

Lennart Nilsson
Department of Biosciences at NOVUM, Center for Structural Biochemistry, Karolinska Institutet, S-141 57 Huddinge, Sweden; Karolinska Institutet, Department of Biosciences, Center for Structural Biochemistry, Hälsovägen 7, SE-141 57 Huddinge, Sweden

Takeshi Nishikawa
Grid Technology Research Center, National Institute of Advanced Industrial Science and Technology, Tsukuba, 305-8561, Japan

Sigeaki Obata
Department of Knowledge-based Information Engineering, Toyohashi University of Technology, Toyohashi, 441-8580, Japan

Miroslav Pinak
Japan Atomic Energy Agency, Shirakata, Shirane 2-4, 319-1195 Tokai, Japan

Dragan M. Popovic
Department of Chemistry, University of California, One Shields Avenue, Davis, CA 95616, U.S.A.

Jason Quenneville
Department of Chemistry, University of California, One Shields Avenue, Davis, CA 95616, U.S.A.

Aina Quintilla
Research Center Karlsruhe, Institute for Nanotechnology, D-76021 Karlsruhe, Germany

R. A. Römer
Department of Physics and Centre for Scientific Computing, University of Warwick, Coventry CV4 7AL, U.K.

Stuart M. Rothstein
Departments of Chemistry and Physics, Brock University, St. Catharines, Ontario, L2S 3A1, Canada

Mitsuhisa Sato
Center for Computational Sciences, University of Tsukuba, Tsukuba, 305-8577, Japan

Matthias Schmitz
Theoretische Biophysik, Lehrstuhl für Biomolekulare Optik, Ludwig-Maximilians Universität München, Oettingenstraße 67, 80538 München, Germany

A. Schug
Forschungszentrum Karlsruhe, Institute for Nanotechnology, P.O. Box 3640, D-76021 Karlsruhe, Germany

Yasuo Sengoku
Department of Knowledge-Based Information Engineering, Toyohashi University of Technology, Toyohashi, Aichi 441-8580, Japan

Rajiv R. P. Singh
Department of Physics, University of California, Davis, CA 95616, U.S.A.

Spiros S. Skourtis
Department of Physics, University of Cyprus, Nicosia 1678, Cyprus

Evgeni B. Starikov
Department of Chemistry, Freie Universität Berlin, Germany; Department of Structural Biology at NOVUM, Karolinkska Institute, Huddinge, Sweden; and Research Center Karlsruhe, Institute for Nanotechnology, D-76021 Karlsruhe, Germany

Alexei A. Stuchebrukhov
Department of Chemistry, University of California, One Shields Avenue, Davis, CA 95616, U.S.A.

Shigenori Tanaka
Graduate School of Science and Technology, Kobe University, 1-1, Rokkodai, Nada-ku, Kobe 657-8501, Japan; and CREST, Japan Science and Technology Agency (JST), Japan

Paul Tavan
Theoretische Biophysik, Lehrstuhl für Biomolekulare Optik, Ludwig-Maximilians Universität München, Oettingenstraße 67, 80538 München, Germany

Nadine Utz
Institut für Physikalische Chemie, Universität Freiburg, Albertstrasse 23a, D-79104 Freiburg im Breisgau, Germany

A. Verma
Forschungszentrum Karlsruhe, Institute for Scientific Computing, P.O. Box 3640, D-76021 Karlsruhe, Germany

Alexander A. Voityuk
Institució Catalana de Recerca i Estudis Avançats, Institute of Computational Chemistry, Universitat de Girona, 17071 Girona, Spain

H. Wang
Department of Physics and Astronomy, Brigham Young University, Provo, UT 84602-4658, U.S.A.

Toshio Watanabe
Grid Technology Research Center, National Institute of Advanced Industrial Science and Technology, Tsukuba, 305-8568, Japan

Wolfgang Wenzel
Research Center Karlsruhe, Institute for Nanotechnology, D-76021 Karlsruhe, Germany; Forschungszentrum Karlsruhe, Institute for Nanotechnology, P.O. Box 3640, D-76021 Karlsruhe, Germany

Alan Windle
Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK

Wei Yang
Deparment of Chemistry and Chemical Biology, Harvard University, MA 02138, U.S.A

Foreword

Biopolymers, such as nucleic acids, proteins and polysaccharides, are the functional basis of all forms of life on Earth. Their three-dimensional structures provide the robustness needed to form templates for parts that constitute biochemical function. At the same time, the dynamics of biopolymers are also crucial, first for folding them into their active conformations and, second, dynamics itself may often play an active role in function. While the connection between structure and function is widely recognized among biochemists, as indicated by the frequently used 'key-in-lock' metaphor, the dynamics counterpart is less well explored and understood – much due to the lack of sufficiently selective and accurate experimental methods and results. Here computational chemistry has for some time played a major role in testing and visualizing models of conformational molecular dynamics and also processes involving proton or electron transfer reactions, and has become an invaluable complement to experimental structural tools, such as NMR or X-ray crystallography, for assessing equilibrium structures in solution and solid state, respectively.

In addition to the traditional scientific areas connected with the study of biological macromolecules, there are a number of young, yet actively expanding fields, involving biopolymers: nanoscience, biotechnology and molecular medicine. In nanoscience and molecular biotechnology, biopolymers and related compounds may be used as templates or scaffolds for various miniaturized technologies – a field which requires detailed knowledge about the structure, dynamics and interactions of the molecules. For instance, hybridization schemes of nucleic acids may be used for designing addressable molecular node assemblies on soft-matter surfaces, which in turn may have applications for making molecular electronics, microscopic molecular machines, diagnostic devices and so on. In the area denoted molecular medicine, specific proteins, whose conformational variability may be related to serious health disorders like Alzheimer's, Parkinson's, Huntington's and Creutzfeld–Jacobs diseases, are dealt with. Also here knowledge about structure, dynamics and interactions of these molecules is needed for understanding the mechanistic background of these deficiencies.

What is clear throughout, and widely accepted, is that biopolymers are extremely complicated physical objects. In order to meet the needs of the above-mentioned scientific branches, broad approaches including several parallel experimental and theoretical techniques are generally necessary. Very frequently, however, the existing experimental methods, even at their modern, tremendously high level of sophistication, are not sufficiently selective to embrace all the interesting aspects of the biopolymer-related phenomena under study. Indeed, this is usually the case when it comes to studies on ultra-fast kinetics of enzymatic reactions, of charge transport through biopolymers, etc. Instead, one is referred to computational approaches for getting a more complete description and understanding of such complex systems. Also, the biopolymers are themselves very complicated dynamical systems, with lots of degrees-of-freedom at a variety of time scales, from femto-seconds up to seconds range. To untangle all these dynamics, for example, to be able to characterize just one dynamic variable at a time, within the frame of some particular experimental study on biopolymers or their complexes, is in most cases even in principle impossible.

Bearing all this in mind, the importance of theoretical studies on biopolymers, their components and related compounds for fully fledged scientific research on them may hardly be exaggerated. Nowadays, theoretical and computational methods in this field are quite well developed, being vigorously stimulated by the dramatic improvements and increase in modern computing power.

This multi-author book, edited by Starikov, Tanaka and Lewis, provides an interesting selection of contributions from a wide international team of high-class researchers representing the main theoretical areas useful for tackling the complicated scientific problems connected with biopolymers' physics and chemistry. It exemplifies the applications of both the classical molecular-mechanical and molecular-dynamical methods, as well as the quantum chemical methods needed for bridging the gap to structural and dynamical properties dependent on electron dynamics. It also provides nice illuminations on how to deal with complex problems when all three approaches need to be considered at the same time. The book gives a rich spectrum of applications: from theoretical considerations of how ATP is produced and used as 'energy currency' in the living cell, to the effects of subtle solvent influence on the properties of biopolymers, and how structural changes in DNA during single-molecule manipulation may be interpreted. As an experimental physical chemist active in the biopolymers field I do appreciate this effort to give an interesting introduction to the currently available theoretical methods.

Bengt Nordén

President of the Fourth Class (Chemistry) of the Royal Swedish Academy of Sciences and former Chairman of its Nobel Committee for Chemistry

Preface

While working out the general concept for the present publication, we never thought of it as an exhaustive handbook, nor did we plan to write a detailed course book. Instead, its main purpose is to show the modern successes and trends in theoretical physical chemistry/chemical physics of biopolymers. Hence, in our minds, the main readership should consist of senior pre-graduates, doctoral students, and younger postdocs in the latter fields, although it would definitely be an utmost honor to us, if those experienced, professional theorists–who are involved in biophysical, structural biological, bionanotechnological, or other related studies–could still find several interesting and novel aspects here. With this in mind, we have encountered a formidable task of carefully selecting the most important contributions to the field, so that the result is inevitably more or less a product of our subjective choice. Nevertheless, we hope that colleagues, whose important and interesting work on biopolymers has not found its proper reflection here, will not get angry with us, the Editors.

For the sake of convenience, we have divided the book into several logical parts, each of which includes a number of chapters written by renowned specialists in their corresponding fields. The succession of these parts represents the actual guideline showing the recommended direction for reading the book, although those colleagues who may be interested in only one aspect or several particular aspects might easily skip all unnecessary material without hindering their proper understanding of that particular area.

The first and foremost part of the book deals with quantum chemical studies on biopolymers and their models, since the most general theory describing properties of any molecule is quantum mechanics. Quantum chemistry is a well-known and well-tried way of approximating a quantum mechanical description of molecular electronic structures. However, when it comes to studying very long oligo (or poly)mers, the *conventional* quantum chemical 'machinery' often fails, owing to complex computational tasks not quite amenable even to the newest computers. This poses a problem of attracting further likely approximations, which could make the quantum chemical studies of large biomolecules more feasible.

One of the ways to solve the latter problem is discussed in detail in Chapter 1, describing the mathematical and computational foundations of the so-called FMO (fragment molecular orbital) method. Interestingly, the latter has already been proven to provide a reliable tool for investigating the electronic structure, intermolecular interactions, and even electron absorption spectra of large biomolecules. Interested readers can find a more detailed description of this in Chapter 2.

While the FMO method pretends to preserve the fully fledged quantum chemical level of molecular description, a more popular approach tends to select a relatively restricted and small region within a biomacromolecule, which can be treated in a rigorous quantum chemical way, whereas the vast majority of the biopolymer in question, together with its solvent surrounding, is considered a classical molecular–mechanical/dynamical system and/or a source of electrostatic field acting on the quantum chemical subsystem. Such 'hybrid' approaches, usually dubbed QM/MM (quantum mechanical/molecular mechanical), are proven to be extremely useful. For example, this is useful when studying

chemical processes in reactive pockets of enzymatic proteins, and also in one of the examples given in Chapter 3.

Along with this, it must be stressed that the usefulness of the conventional quantum chemical approaches for solving modern biomolecular problems should in no way be underestimated, as clearly demonstrated in Chapter 4 (which discusses the details of the physical–chemical nature of Watson–Crick hydrogen bonds in nucleic acids) and Chapter 5 (which discusses the quantum chemical evaluation of charge transfer parameters in nucleic acids).

Still, a great number of biologically relevant problems involving biopolymers can be solved by assuming solely classical mechanical (purely MM) pictures and avoiding any thorough quantum chemical description, with conformational preferences/dynamics, rheology and folding of, docking small molecules to, structural defects in biomacromolecules, biomolecular motors, etc. among them. In our opinion, good examples of such studies are presented in Chapters 6–13. On the one hand, the solution of large-scale problems, like drug discovery, requires involvement of the modern computational technologies (for grid technologies, see Chapter 12). The models of the latter sort must in many cases be used in clever combinations with the molecular mechanical and/or molecular dynamical approaches in the all-atom representation (like in investigations of molecular motors, where molecular dynamics are combined with kinetic equations, see Chapter 14).

Of course, the intricate complexity of biopolymers also allows us to sometimes directly treat them as statistical/stochastic systems and/or to use statistical methods in description or elucidation of their properties. One of the areas of effective and successful applications of statistical methods, like Monte Carlo (Chapter 14), principal component (factor) analysis (Chapter 15), as well as stochastic optimization techniques (Chapter 16), is connected with elucidating the detailed mechanisms of protein folding.

Aside from all the above-mentioned methods and concepts of biophysical chemistry, there is one unique theme which belongs to the field of biochemical physics, namely, charge transfer/transport in and electrical properties of biomacromolecules, with special reference to DNA. The latter phenomenon is still fiercely debated, and thus stimulates intensive and numerous physical–chemical theoretical studies.

The general physical basis of such studies is the so-called *'tight-binding'* approximation, where large biomolecular fragments like, for example, Watson–Crick base pairs in DNA, are declared to be just abstract 'sites' carrying only one orbital which describes electron motions within each 'site'. Chemically, these orbitals correspond to either HOMO (the highest occupied molecular orbital) or LUMO (the lowest unoccupied molecular orbital). The sites involved are then characterized by their 'site energy,' which is either ionization potential (the energy required to withdraw one electron from the site if the site orbital is HOMO) or electron affinity (the energy required to add one electron to the site if the orbital is LUMO).

In real biopolymers, these sites are coupled to each other, and it is this electronic coupling that promotes the motion of the added electron, or a positively charged 'hole' produced by withdrawing one electron. Since there is always a *discrete* set of such sites, the motion of the charged particle is just *hopping* from one site to another. This is why the strength of electronic coupling (the term preferred by chemists) between the neighboring sites is frequently called 'hopping parameter' or 'hopping integral' (the term preferred by

physicists), and in most cases it is silently or explicitly assumed that the hopping can only occur between the nearest-neighboring sites. To establish a connection between the above oversimplified model and the actual bipolymer systems, one needs to have likely estimates of the site energies and hopping integrals (alias electron couplings). In solving this problem, quantum chemistry is very effective (as we have already seen in Chapter 5, but see also in Chapters 18–24). Model Hamiltonians are very important not only in describing electronic structure of biopolymers, but also their classical mechanical and dynamical properties. This is clearly demonstrated in the Chapter 17, where phenomenological, analytically solvable models (mechanical soliton models) are described.

Meanwhile, to provide one with a completely valid description of the biomacromolecular charge transfer/transport, the rigid *in vacuo* set of the above-mentioned 'sites' is by far not enough. Indeed, the conformational degrees-of-freedom, as well as the water–counterion environment, of biopolymers must also be taken into account by any correct physical theory (see an excellent review of this theme in Chapter 18). There are basically *two* ways to achieve this, namely, *either* to assume some functional dependence of the site energies and hopping parameters on the intra-biopolymer degrees-of-freedom and dynamical modes of the surrounding, or to treat the the dynamics of the biopolymer together with its environment explicitly, by using classical molecular dynamics in the all-atom representation.

The first approach enables one to formulate so-called *polaron* theories of the charge transport, if site energies and hoppings can be considered *linear* functions of the conformational degrees-of-freedom, so that the charge motion can be viewed as a propagation of a charged particle, which permanently 'drags' the biopolymer 'lattice' deformation with itself (the picture most popular among physicists). The parameters of coupling between the charged particle motion and the biopolymer 'lattice' deformation can also be estimated using a clever combination of quantum chemistry and classical molecular dynamics in the all-atom representation (see also Chapter 5, as well as Chapters 19–24).

The second approach makes it possible to *direct* simulation of the charge transfer/transport process, which helps us to visualize it (the picture most popular among chemists, see Chapters 20–22 and 24).

More exciting details on how all of the above approaches and ideas work in real physical studies can be found in Chapters 18–24.

The last, but not least, area of modern biopolymers theory is concerned with their electrical properties, which are of primary importance for modern bionanoscience. This theme is related to model Hamiltonians from the previous book section, but there are still some very important differences. Namely, the tight-binding Hamiltonians are very good at describing the propagation of a *single* charge particle through biopolymers, which is completely relevant to experiments carried out by physical chemists. However, if we wish to consider the full physical–technical representation of electrical properties of biomacromolecules, we must also consider charge flows in them, created by attaching them to electrodes and applying voltage. The latter situation requires a different electric transport theory, compared with the former case. Chapter 25 shows in detail, how density-functional theory (DFT) can be used in trying to solve the latter problem. Chapters 26 and 27 present detailed accounts of such theories and algorithms in the all-atom representation. The difference between the two approaches–outlined in the last two chapters–consists of the accuracy and efficiency of biomolecular description.

Meanwhile, the Hamiltonians used in both of these chapters are actually the extended versions of the above-mentioned 'tight-binding' Hamiltonian. Specifically, the 'sites' here are 'real' atoms containing all the pertinent valence electron orbitals, and the possibility of a non-zero overlap between these orbitals in biomolecules is taken into account. This corresponds to the so-called 'extended Hückel' approach. The latter is a typical 'tight-binding' technique and, at the same time, the most primitive method of quantum chemistry, where one uses a set of fixed parameter values and empirical formulas to evaluate atomic ionization potentials and inter-atomic electron hopping in molecules, while more elaborate quantum chemical methods also make use of self-consistent iterations to reveal the most energetically favorable electron charge distributions in molecules.

To summarize, our book illustrates the vast majority of directions within modern biopolymers theory and can hopefully be a useful illustration for standard senior graduate or postgraduate courses in biophysics and/or biophysical chemistry.

DISCLAIMER

The Editors of the present book are fully aware that the theoretical approaches most recently advocated in the special literature and presented here can also contain methods and techniques which are still not unanimously accepted by the scientific professional community. Since our book should be considered just an *illustrative guide* to the physical–chemical theory of biopolymers and their components, the Editors do not take any responsibility for the merits or demerits of any of the particular methods and results published here *not* by the Editors themselves, nor does the publication of any particular chapter reflect the Editors' subjective acceptance or denial of its contents in any way. Every authors' team of this book carries full scientific responsibility for the credibility, validity and presentation of *solely their own* approaches and results.

Ewgeni B. Starikov
Shigenori Tanaka
James P. Lewis

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 1

Theoretical development of the fragment molecular orbital (FMO) method

Dmitri G. Fedorov and Kazuo Kitaura

National Institute of Advanced Industrial Science and Technology, 1-1-1 Umezono, Tsukuba, 305-8568 Ibaraki, Japan

Abstract

The fragment molecular orbital (FMO) method has been introduced in detail, with the emphasis on physical ideas constituting the foundation of the method. The recent theoretical development incorporating electron correlation at various levels has been covered. The means to compute the amount of interactions between fragments, including polarization and charge transfer, have been elucidated, with practical applications to a model system and a biological problem. The absolute and relative accuracy of the FMO total energies have been scrupulously established by comparison with *ab initio* methods. The practical issues of applying the method to real-life problems have been dealt with, including fragmentation, the choice of basis sets and parallelization. It is hoped that this very detailed description enables a general reader to apply the method to practical problems, as well as to extract a physical picture of the interactions therein.

1.1 INTRODUCTION

Molecular simulations have been used as a tool to study structures and functions of biomolecules for a number of years. Early studies of real biomolecules were mostly limited to molecular mechanics (MM) and the size of the systems was too large to permit applications of more reliable *ab initio* quantum mechanical (QM) methods. Although the MM methods can be successfully applied to a wide range of systems, they have serious if not disastrous problems describing some types of phenomena, most remarkably, those including electron density changes (charge transfer and chemical reactions) and excited state properties.

A hybrid method combining the two approaches known as QM/MM [1] has become popular, where the important part such as substrate and binding pocket from the enzyme is described by a QM method and the remainder is treated with MM. The number of atoms in

the QM region, however, often exceeds hundreds and even these methods become difficult to apply. While such a hybrid approach often describes the QM region satisfactorily, the environment is still treated with MM, inheriting its problems. It may be expected that as the system size grows, the energy contribution of the environment becomes larger too, so for large biological molecules reasonable accuracy is desirable in describing the *whole* of the system.

Much work has been done during the last decade, in order to extend the applicability of QM methods. Owing to the development of computers and the state-of-the-art implementations of the conventional *ab initio* QM method, Sato *et al*. [2] have succeeded in calculating a whole protein containing 1 738 atoms. To overcome the steep growth of *ab initio* computations with system size (known as scaling), linear scaling algorithms have been proposed [3], in which the amount of computations increases linearly with system size. Scuseria has reported linear-scaling density functional calculations of an RNA piece containing more than 1 000 atoms [4] and a number of linear-scaling methods have been proposed specifically to treat electron correlation [5,6].

On the other hand, a variety of fragment-based methods has emerged aimed at electronic structure calculations of large molecules. In these methods, a molecule is divided into fragments, then calculations are carried out on the fragments, and finally the total energy and properties of the whole molecule are evaluated using those of the fragments. Such approaches have a long history and can be classified according to different criteria. Perhaps two classification criteria are the most basic: (a) building scheme (including elongation and fragment-pair consideration) and (b) the way the environment (everything not included in the fragment or fragment conglomerate) is treated. It is the combination of the two that is required to do the proper classification.

According to the building scheme criterion, fragment-based methods can be classified into three categories; (a) divide-and-conquer approaches, (b) transferable approaches, and (c) fragment-interaction approaches.

(a) The divide-and-conquer method proposed by Yang [7] divides the system into small subsystems, and calculations are performed on each subsystem with its surroundings (buffer) using a local Hamiltonian. The density matrix of the whole molecule is obtained by combining the density matrices of the subsystems. Using the total density matrix, the total energy (and other properties) is obtained as the expectation value of the Hamiltonian operator of the whole molecule. The elongation method proposed by Imamura *et al*. [8] and the *ab initio* fragment-based theory suggested by Das *et al*. [9] belong to this category.

(b) Transferable approaches are based on the well-known additivity property of heat of formation, which is approximately equal to the sum of bond (or other subunit) energies. In these methods, the total energy of a molecule is estimated by the addition and subtraction of the fragment (functional group) energies obtained from independent calculations, and their accuracy largely depends upon the fragmentation technique. Various methods have been proposed along this line for a long time [10,11]. The ONIOM method [12] uses the same idea to transfer the effect of environment in the additive fashion.

(c) The fragment interaction approach is based on the theory of molecular interactions; the total energy E of a molecular cluster is given as a sum of monomer energies and

intermolecular interaction energies obtained from monomer pairs (dimers) and, possibly, larger conglomerates:

$$E = \sum_I E_I + \sum_{I>J} \Delta E_{IJ} + \sum_{I>J>K} \Delta E_{IJK} + \cdots \quad (1)$$

where E_I are monomer energies, and ΔE_{IJ} and ΔE_{IJK} are two- and three-body interaction energies, respectively. This series expansion was applied to calculate the total correlation energy of a molecule in the incremental method proposed by Stoll [13]. The molecular fractionation with conjugated caps (MFCC) approach originally proposed by Zhang and Zhang [14] has been recently extended by Li *et al*. [15] into the energy-corrected formalism (EC-MFCC), where a similar series expansion of molecular total energy in terms of interfragment interaction energies is used. The FMO method by Kitaura *et al.* [16,17] falls into this category.

As far as the other criterion is concerned, most fragment-based methods ignore the environment completely or include just adjacent parts in the form of caps. Among the methods that do include *all* environment, one can name the FMO method and its recent closely related variant proposed by Hirata *et al.* [18]. The incremental method by Stoll, frequently used in terms of excitations in orbitals groups, is not formally a fragment-based method but it shares the many-body expression in Eq. (1) and also incorporates the environment (through restricted Hartree–Fock (RHF) calculations of the whole system).

The QM/MM approach (that can be thought of as introducing two fragments) has the environment due to MM in the QM region, but generally no effect of QM in the MM region (with the exception of polarizable MM methods, such as that by Dupuis *et al.* [19]). The so-called electronic embedding in ONIOM [20] adds MM charges to the QM region. However, if more than two layers are employed (so that ONIOM becomes different from QM/MM), then interaction with the environment (from lower QM layers) is ignored during higher-layer QM calculations (as well as the above-mentioned lack of QM environment in the MM region). Thus, ONIOM in its present form does not include full interaction with the environment in each 'fragment' (layer) calculation.

Just because atoms are defined in the *ab initio* methods, these methods cannot be classified as 'divide-and-conquer', since the influence of the whole system is considered. Similarly, in the FMO method one defines and handles fragments submerged in the whole system. Drawing a political analogy, divide-and-conquer policy results in independent countries, whereas the fragments in the FMO method correspond to states or prefectures in the same country and instead of 'divide-and-conquer' it is much more appropriate to classify such methods under the name of *'e pluribus unum'*.

We thus propose to classify methods to the '*e pluribus unum*' category if they satisfy two conditions: (a) during each subsystem calculation, the influence of the whole system is included (e.g., through the external Coulomb field applied to each fragment or fragment pair in the FMO method), and (b) the properties of the total system are obtained. Incremental correlation and FMO methods (including the electrostatic (ES) potential (ESP) approximation method by Hirata *et al*. [18]) fall into this category, whereas the majority of other fragment methods not accounting for all of the environment do not. Despite its name, the divide-and-conquer method by Yang [7] belongs to this category, as it contains

the projection of the total Hamiltonian upon local subsystems, thus retaining the effect of the whole system upon each subsystem. Elongation methods [8] seem to drift toward this category and can be assigned to it, if the elongation technique is repeated after reaching the last unit, so that the effect of the whole system is purported back to each fragment.

1.2 THE THEORY OF THE FMO METHOD

1.2.1 Basic ideas

The mathematical definitions are given in the subsections below and now we introduce the general concepts. The fragments in the FMO method are computed while submerged in the rigorous Coulomb field computed for the whole system, followed by explicit many-body corrections obtained from fragment pairs and triples, called dimers and trimers, respectively. Owing to such a 'Coulomb bath', there is no need to add unphysical caps to define fragments, and bonds can simply be fractioned electrostatically, by dividing electrons between fragments.

In fact, the FMO method can be viewed as an *ab initio* method with strongly enforced orbital localization and a many-body expansion to compensate for the limitations of such enforcement. Compared with the *ab initio* case, fragment-based Fock matrices contributions are exact for each subsystem, contain the exact Coulomb interaction due to the rest of the system (environment) and neglect only the electron exchange with the environment.

As shown below, dimer and trimer calculations serve several very important purposes. First, they allow for charge transfer between fragments. Second, the addition of such higher body corrections greatly improves the overall accuracy. Third, valuable quantitative many-body interaction information is obtained.

The strength of the FMO method lies in the many-body interactions computed in the 'Coulomb bath'. Only the subjective fragmentation is what makes it depart from first principles. Otherwise, the total Coulomb field, fragments and their dimers and trimers are computed with *ab initio* methods.

1.2.2 Fragmentation

Before going into the mathematical description, the issue of fragmentation has to be addressed. In order to fragment a molecule connected by covalent bonds, a way to fraction bonds has to be devised.

The bonds in the FMO methods are fractioned electrostatically (no caps whatsoever). One fragment is assigned two electrons from the fractioned bond and the other none. The simplest scheme that has a certain drawback is given in Fig. 1.1(a), where the C–C bond in C_2H_6 is fractioned. The fragment on the right gets two electrons from the fractioned bond, so 18 electrons in C_2H_6 are divided as eight and ten between the two fragments. The electron density of the left fragment (CH_3) is determined by the electron distribution of eight electrons in the total Coulomb field created by the whole system containing 18 electrons and 18 protons. The exchange interaction is limited to within the fragment. Likewise, the electron density of the right fragment is determined by the distribution of

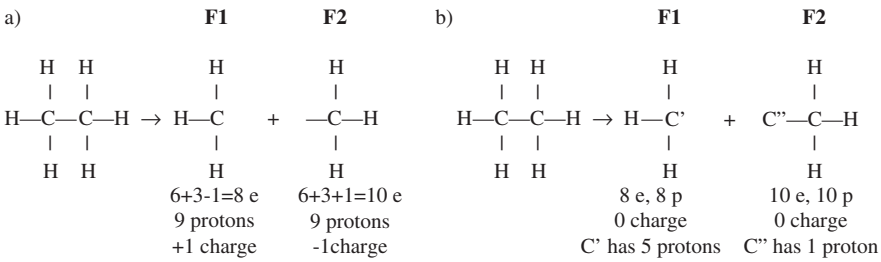


Fig. 1.1. Illustration of details of bond fractioning in the FMO method: (a) the simple scheme, and (b) the actually used scheme. C_2H_6 is divided into two CH_3 fragments, denoted by F1 and F2.

10 electrons in the total Coulomb field. Therefore, the only restriction compared with *ab initio* treatment is the neglect of exchange interaction between fragments. Such contribution is added on the second stage when dimer calculation is performed.

In the FMO method, one is interested in obtaining pair interaction energies, which are defined and discussed in greater detail later. If the above described fractioning were used as is, the left and right fragment would be assigned formal charges $+1$ and -1, respectively. The fragment–fragment pair interaction would then correspond to attraction of charged electron densities and be not very useful. Therefore, another step is taken and one proton is reassigned. It will be shown below that such reassignment does not change the total properties and only individual monomer and dimer energies are redefined. Assigning one electron and one proton from the C atom to the right fragment, the left fragment retains eight electrons and eight protons, while the right fragment has 10 electrons and 10 protons. The basis sets are left as is, that is, the carbon atom whose protons are assigned to the left (C′with five protons) and right (C″ with one proton) fragment has the carbon atom basis set.

With such division, molecular orbital space in each fragment contains the atomic orbitals on the atom at which the bond is fractioned. A simple technique of orbital projections can be used to divide not only the protons but also the atomic orbital space. For carbon atoms one can define a set of sp^3 hybridized orbitals, one of which points along the fractioned bond. The fragment on the right needs only to keep the latter sp^3 orbital, and to have the other four orbitals (three sp^3 and core 1s) projected out, while the left fragment has to retain four orbitals and to have one orbital projected out. Projections are performed on the orbital space and the exact definition is given below.

It should be noted that in the FMO method, the fragmentation is performed at an atom in a bond, and it makes a difference at which of the two atoms the bond is fractioned. This can be seen in Fig. 1.1(b), where the bond was fractioned at the carbon atom that is denoted by C′ and C″. An atom at which a bond is fractioned is called the *bond-detached atom*. It is possible, but in the great majority of cases not recommended, to fraction more than one bond at the same atom.

Next, a question arises as to where the fragmentation should be done to achieve the best accuracy. In general, the larger the resultant fragments, the higher the accuracy and the larger the computational cost. There is no specific need to have fragments of the same size. On the contrary, chemical knowledge should be employed, as chemically defined units are often the best choice both in terms of accuracy and for interaction analysis (*vide*

infra). For proteins, the natural unit is amino acid residues, so frequently one or two residues per fragment division are used. It was found that the best location is to fragment proteins at C_α. The issue of choosing the fragment size is addressed in detail below, it should suffice to say here that typical fragment sizes are 10–40 atoms.

The most important criterion where to fragment is to avoid fractioning bonds that involve delocalized electron densities. The archetypal example is benzene, and it should be obvious that aromatic bonds must not be fractioned. In general, multiple bonds frequently possess a delocalized character to some degree and it is not recommended to fraction a bond at an atom that is involved in a multiple bond. Some typical desired and undesired fragmentation examples are shown in the top and bottom parts of Fig. 1.2, respectively. Some numeric comparison of various fragmentation schemes can be found in [21].

The other issue to keep in mind is that fragmentation should not be performed too close to the region of importance, that is, to the reaction centre. A small buffer zone of atoms included in the same fragment along with the reaction centre can help significantly improve accuracy.

Fragmentation in the FMO method is not a formal mathematical exercise, and should be conducted based upon chemical knowledge. Certain types of system pieces that are not bound by a covalent bond may be desirable to put into one fragment, if it is known that a very strong interaction occurs between them, such as within ferrocenes or in salt bridges. Hydrogen bonding is taken into account by dimer calculations, so there is no need to put pieces connected by a hydrogen bond into the same fragment.

In some cases the choice of fragmentation has little effect upon accuracy but plays an important role in computational efficiency. A typical example is water clusters. If one is interested in the total energies of water clusters, then as discussed below it is better to place two water molecules in one fragment to improve the accuracy. While the total energy is not very sensitive to the particular way of pairing water molecules, the best approach from the computational efficiency point of view is to put two geometrically closest water molecules into the same fragment. With such fragmentation, the interfragment separation and electrostatic potential calculations are most efficient.

In the applications of the FMO method, so far only C–C and C–O bonds have been fractioned (in the latter case at carbon atoms). There is no restriction in the method as to what type of bonds can be fractioned, and using the projection operators described below

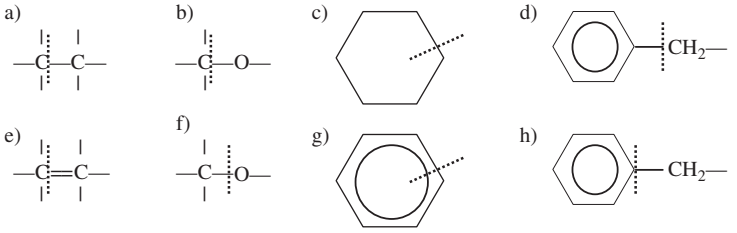


Fig. 1.2. Examples of appropriate (a)–(d) and inappropriate (e)–(h) bond fractioning in the FMO method. Double bonds (e) and even more so aromatic rings (g) should not be fractioned. Fractioning at oxygen atoms (f) requires a set of corresponding oxygen sp^3 hybridized orbitals and it is recommended to use scheme (b) instead. Single bonds should not be fractioned at sp^2 carbon atoms (h).

most single bonds fractioned at non-carbon atoms can be handled. A simple comparison of some small system's total energies obtained with the FMO and *ab initio* methods may be used to make sure no large accuracy loss takes place if new bond types are fractioned.

A legitimate question arises whether any system can be fractioned and used with the FMO method. Formally, the answer is yes, yet in practice a few certain system types may be expected to have unacceptably poor accuracy. Such systems all share the same feature of strong electron delocalization, and typical examples are: metallic crystals, large metallic clusters, and single molecule fullerenes. Clusters of fullerenes or superclusters composed of metallic clusters should not present a problem though. Other examples that may be possible to handle with very careful choice of fragmentation: carbon nanotubes with ring-wise fragmentation and organic systems with mostly sp^2 carbon atoms not permitting the desired fractioning at sp^3 atoms.

1.2.3 Mathematical formulation

We expand the total energy of a molecule or a molecular cluster divided into N fragments into the following series [22]:

$$\begin{aligned} E = &\sum_{I}^{N} E_I + \sum_{I>J}^{N} (E_{IJ} - E_I - E_J) + \sum_{I>J>K}^{N} \{(E_{IJK} - E_I - E_J - E_K) \\ &- (E_{IJ} - E_I - E_J) - (E_{JK} - E_J - E_K) - (E_{KI} - E_K - E_I)\} + \cdots \end{aligned} \tag{2}$$

The FMO expansion including trimers as given in Eq. (2) is denoted by FMO3 (the three-body expansion), and Eq. (2) without the last sum involving trimers defines the two-body expansion of FMO2. The former has higher accuracy and is more expensive, but both find their uses. The monomer sum only (the one-body expansion) as explained below is not useful in general.

Before we define monomer (E_I), dimer (E_{IJ}) and trimer (E_{IJK}) energies, let us consider two simple cases, where the number of fragments N is two and three.

$$\begin{aligned} E(N=2) &= (E_1 + E_2) + E_{21} - (E_1 + E_2) = E_{21} \\ E(N=3) &= (E_1 + E_2 + E_3) + (E_{21} - E_2 - E_1) + (E_{31} - E_3 - E_1) + (E_{32} - E_3 - E_2) \\ &\quad + E_{321} - (E_1 + E_2 + E_3) - (E_{21} - E_2 - E_1) - (E_{31} - E_3 - E_1) \\ &\quad - (E_{32} - E_3 - E_2) \;=\; E_{321} \end{aligned} \tag{3}$$

It is immediately seen that for two and three fragments, the energy expression in Eq. (2) becomes exact if two and three body expansions are used, correspondingly. That is, the total energy is exactly equal to the energy of dimer 21 and trimer 321, respectively. Since the dimer and trimer energies for $N = 2$ and 3, correspondingly, are identical to *ab initio* total energies, the FMO method based on the n-body expansion is exact if the number of fragments N is equal to n. If more fragments are present, Eq. (2) is a systematically improvable approximation to the total properties.

The energies of monomer, dimers and trimers (called *n*-mers, $n = 1,2,3$) appearing in Eq. (2) are obtained as follows. In the case of RHF, the FMO equations can be written as:

$$\tilde{\mathbf{F}}^X \mathbf{C}^X = \mathbf{S}^X \mathbf{C}^X \boldsymbol{\varepsilon}^X \tag{4}$$

$$\tilde{\mathbf{F}}^X = \tilde{\mathbf{H}}^X + \mathbf{G}^X \tag{5}$$

$$\tilde{H}^X_{\mu\nu} = H^X_{\mu\nu} + V^X_{\mu\nu} + B\sum_{i\in X} P^i_{\mu\nu} \tag{6}$$

$$V^X_{\mu\nu} = \sum_{K\neq X}^{N} \left\{ \sum_{A\in K} \left\langle \mu \left| -\frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} \right| \nu \right\rangle + \sum_{\rho\sigma\in K} D^K_{\rho\sigma} (\mu\nu|\rho\sigma) \right\} \tag{7}$$

$$P^i_{\mu\nu} = \left\langle \mu \middle| \varphi^h_i \right\rangle \left\langle \varphi^h_i \middle| \nu \right\rangle \tag{8}$$

$$E_X = \frac{1}{2}\mathrm{Tr}\left\{ \mathbf{D}^X \left(\tilde{\mathbf{H}}^X + \tilde{\mathbf{F}}^X \right) \right\} + E^{NR}_X \tag{9}$$

$$E'_X \equiv E_X - \mathrm{Tr}(\mathbf{D}^X \mathbf{V}^X) \tag{10}$$

where $X = I$ for monomers, $X = IJ$ for dimers and $X = IJK$ for trimers, μ, ν, ρ and σ run over atomic orbitals, K runs over external fragments, E^{NR}_X is the nuclear repulsion energy of X, Z_A and $\mathbf{R}_A$ are atomic charges and coordinates, respectively. **D** is the density matrix, **F** denotes the Fock matrix that is made of standard one- (**H**) and two-electron (**G**) contributions, with the addition of the environmental potential $\mathbf{V}^X$ and the projection operator matrices $\mathbf{P}^i$ built upon the hybridized orbitals φ^h_i (B is a universal constant). The physical meaning of E'_X energies is explained below.

In other words, in the FMO method *ab initio* monomer, dimer and trimer RHF calculations are performed in the global electrostatic field created by the whole system. The electrostatic field is composed of the contribution from within a given *n*-mer, as included in $\mathbf{H}^X$, and the external Coulomb field $\mathbf{V}^X$.

The projection operators $\mathbf{P}^i$ are placed on bond-detached atoms to divide the basis functions along the fractioned bonds, as described above. In the case of fractioning C–C bonds, the sp^3 hybridization orbitals φ^h_i are obtained from a RHF calculation of CH_4 ($R_{C\text{–}H} = 1.09$ Å) for a given basis set. These orbitals are used unchanged in FMO calculations (only rotated for each bond to match its direction). The parameter B is chosen to be sufficiently large to remove the corresponding orbitals out of variational space, that is, normally $B = 10^6$ a.u. (the term in Eq. (8) contributes to the total energy of the order of B^{-1}).

The projection operators for atoms other than carbon can be easily generated from appropriate calculations. For example, for fractioning single Si–Si bonds at sp^3 silicon one may take SiH_4. The influence of the particular choice of the localization scheme is numerically small ($<<B^{-1}$), however, one should not forget to rotate the model molecule (SiH_4) in such a way so that the model bond (Si–H) be put along the z-axis. A sample file to aid in making projection operators is included in the distributed version of GAMESS [23,24].

1.2.4 Computational scheme and property calculation

By looking at Eq. (7), it is apparent that the external electrostatic potential $\mathbf{V}^X$ depends upon all monomer densities $\mathbf{D}^K$. Thus, the monomer calculations have to be repeated self-consistently, performing *ab initio* RHF calculations of monomers in the external field until all monomers converge (which also includes convergence of the external field).

Next, we describe how the FMO scheme is applied in practice. A molecular system is partitioned into fragments. This division is fixed for all calculations. Monomer calculations are repeated until all monomer energies converge, which involves running each monomer typically about 20 times. Then dimer and, optionally, trimer calculations are performed in the external Coulomb field determined by the monomer electron densities during the previous step. Each n-mer RHF ($n>1$) computation is performed only once. The obtained total intramolecular dimer (E_{IJ}) and trimer (E_{IJK}) energies are combined by means of Eq. (2) and the total energy of the system is obtained. The diagram showing the various steps in the FMO method is given in Fig. 1.3.

Other properties, linear in electron density, are obtained following the density expansion analogous to energy in Eq. (2) and given below for the two-body case:

$$\mathbf{D} = \sum_I^{\oplus} \left[\mathbf{D}^I\right] + \sum_{I>J}^{\oplus} \left[\mathbf{D}^{IJ} - \mathbf{D}^I \oplus \mathbf{D}^J\right] \tag{11}$$

where the sum signs should be taken in the tensor (block) sense, so that each block I or IJ contributes to the supermatrix $\mathbf{D}$ in the appropriate location. The dimer contributions

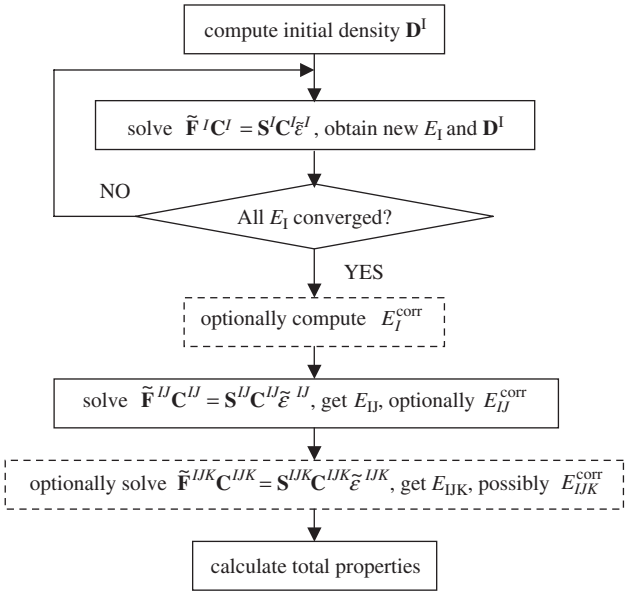


Fig. 1.3. The FMO computational scheme. Optional blocks are shown as dotted rectangles.

are defined as in:

$$\Delta\mathbf{D}^{IJ} \equiv \mathbf{D}^{IJ} - \mathbf{D}^{I} \oplus \mathbf{D}^{J} = \begin{bmatrix} \mathbf{d}^{II} & \mathbf{d}^{IJ} \\ \mathbf{d}^{JI} & \mathbf{d}^{JJ} \end{bmatrix} - \begin{bmatrix} \mathbf{d}^{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{d}^{J} \end{bmatrix} \tag{12}$$

The general structure of the total density matrix **D** in the FMO method is the same as in *ab initio* methods, as can be seen below for the case of three fragments.

$$\mathbf{D} = \begin{bmatrix} \tilde{\mathbf{d}}^{A} & \mathbf{d}^{AB} & \mathbf{d}^{AC} \\ \mathbf{d}^{BA} & \tilde{\mathbf{d}}^{B} & \mathbf{d}^{BC} \\ \mathbf{d}^{CA} & \mathbf{d}^{CB} & \tilde{\mathbf{d}}^{C} \end{bmatrix} \tag{13}$$

where off-diagonal blocks are equal (such as $\mathbf{d}^{AB}$=$\mathbf{d}^{BA}$) and obtained directly from dimer densities in Eq. (12). The diagonal blocks in the case of three fragments are given by:

$$\tilde{\mathbf{d}}^{A} = \mathbf{D}^{A} + (\mathbf{d}^{AA(AB)} - \mathbf{D}^{A}) + (\mathbf{d}^{AA(AC)} - \mathbf{D}^{A}) \tag{14}$$

In general, the corresponding expression is:

$$\tilde{\mathbf{d}}^{I} = \mathbf{D}^{I} + \sum_{J \neq I} (\mathbf{d}^{II(IJ)} - \mathbf{D}^{I}) \tag{15}$$

where $\mathbf{d}^{II(IJ)}$ denotes the *II* block in the *IJ* dimer density $\mathbf{d}^{IJ}$. The sum in Eq. (15) represents the density (charge) transfer corrections to monomer densities.

Dipole moments and Mulliken charges can be obtained using the density expression in Eq. (11), with the result that such properties themselves are expanded similarly to energy in Eq. (2).

The energy gradients can be obtained by taking an analytic derivative of Eq. (2). The present FMO gradient implementation [25] avoids solving couple-perturbed Hartree–Fock equations needed to obtain strictly analytic gradients (to find the derivative of dimer MO coefficients with respect to the atomic coordinates of external monomers). Two other small contributions are also omitted: the derivative of the external monomer potentials and the derivative of projection operators. For an *n*-mer gradient only, derivates with respect to its own atomic coordinates are included.

The energy gradient in the FMO method is computed for the whole molecule, thus optimization methods developed for *ab initio* quantum chemistry can be used without modifications, for example, by using the steepest descent or Newton–Raphson methods. It is conceivable to use the additional information (derivatives of fragment energies) in an FMO-specific way, for example, for partial optimizations, but to the best of our knowledge, such usage has not been reported yet and all geometry optimizations construct the gradient for the total system and optimize its structure as a whole.

Molecular orbitals and orbital energies can also be computed using the FMO method. Sekino *et al.* [26] compared these properties as obtained from individual *n*-mer calculations with the corresponding values for *ab initio* methods. Inadomi *et al.* [27] proposed computation of the molecular orbitals for the whole system, diagonalizing the Fock matrix computed from the density expansion in Eq. (11).

1.2.5 Many-body treatment

The FMO method was built upon the ideas coming from the energy decomposition analysis (EDA) of Kitaura and Morokuma [28]. The relation between the one-body expansion of the FMO method and the EDA is described in detail in [22]. Briefly, in the FMO method and the EDA scheme, the orbital mixing is restricted, both produce polarized intramolecular monomers, and polarization is fully accounted for at the *N*-body level, where *N* is the number of fragments (monomers). In the case of clusters of polar molecules, this type of many-body effect is major.

The schematic representation of the two-body FMO method is demonstrated for the case of three monomers in Fig. 1.4. The full *ab initio* energy (*E*) is obtained by allowing all orbital mixing and all Coulomb interaction. Any expansion that leaves invariant the number of solid lines (orbital mixing) and dotted arcs (the ESPs) does not introduce any artificial interaction. The *n*-body FMO expansion ($n > 1$) satisfies this condition, which can be verified by adding and subtracting the numbers of lines (or arcs) according to the sign in front of the corresponding energies. We note in passing that while the number of lines is the same for the regular RHF energy and the two-body FMO energy, it is the omitted *simultaneous* orbital mixing (involving three monomers) that makes the difference between the two.

Some of the major types of three-body effects are shown in Fig. 1.5. While FMO2 takes into account the three-body effect shown in Fig. 1.5(a), only FMO3 considers three-body effects shown in Fig. 1.5(b) and 1.5(c). For an *n*-mer, *n*-body interaction is fully included and the majority of higher-body interaction (up to *N*-body) is recovered through the Coulomb field (external potential) of other monomers. However, just placing the potential does not recover *all* higher-body effects and thus the necessity arises to proceed from the two or three-body terms in Eq. (2) if higher accuracy is desired.

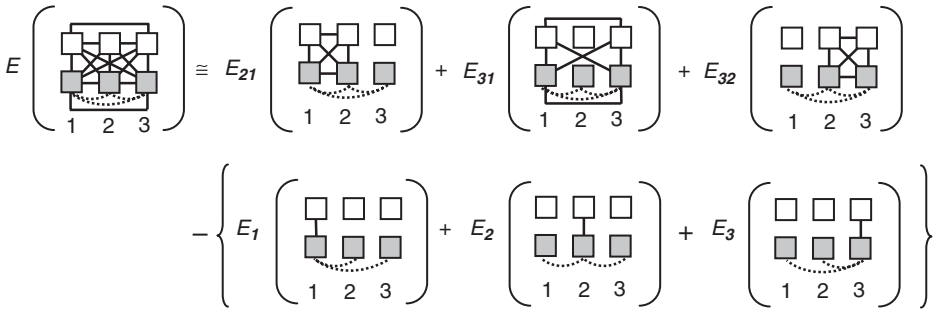


Fig. 1.4. Energy expansion for the two-body FMO method when three fragments are present:

$$E^2 = \sum_{I}^{N} E_I + \sum_{I>J}^{N} (E_{IJ} - E_I - E_J) = E_{21} + E_{31} + E_{32} - E_1 - E_2 - E_3.$$

The orbital mixing is shown with solid lines and the external electrostatic potential is depicted by dotted arcs. Shaded and empty squares above monomer indices (1, 2 and 3) represent occupied and virtual orbitals, respectively. Both the orbital mixing and the potential exactly match the *ab initio* diagram *E*, as obtained by summing the number of lines of each type on the left (*ab initio*) and right-hand sides (FMO2).

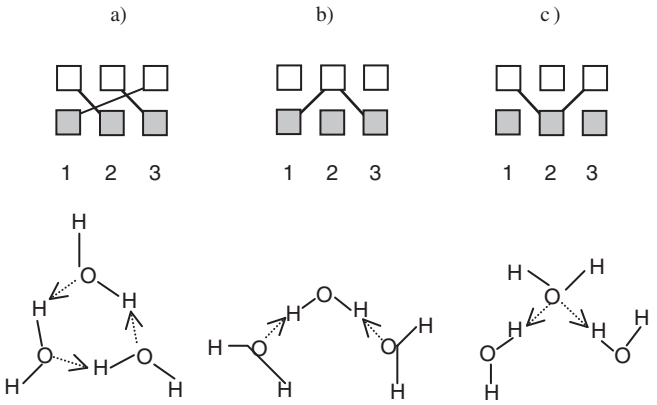


Fig. 1.5. Some of the important examples of three-body interaction in a system with three monomers 1, 2 and 3: (a) cyclic charge transfer that can be described by two-body terms in FMO2, (b) and (c) purely three-body charge transfer taken into account only by FMO3. Dotted lines show hydrogen bonds, and arrows point in the direction of charge transfer.

Finally, it is clear why one-body FMO methods are not useful. If one sums all arcs representing the ESPs in Fig. 1.4, the number of arcs is twice as large as it should be. This means that the electrostatic interaction is double counted if only one-body properties are added (with the only exception being the case when there is just one fragment).

1.2.6 Electron correlation

The incorporation of electron correlation is straightforward and follows three main routes: (a) self-consistent treatment of electron correlation in density functional theory (DFT) [29,30] and multiconfiguration self-consistent field theory (MCSCF) [31], (b) perturbative treatment (that is, with RHF-optimized orbitals): second-order Møller–Plesset perturbation theory (MP2) [32–34] and coupled-cluster (CC) theory [35], and (c) excited states in configuration interaction (CI) methods, such as CI with single excitations (CIS) [36].

For DFT and static electron correlation (MCSCF), the same formal expression of Eq. (2) is used. For DFT, exchange-correlation functionals are added to each *n*-mer Hamiltonian, whereas for MCSCF the multiconfigurational wave function is used and the corresponding MCSCF equations are solved. In the latter case, one fragment is chosen to be of MCSCF type and all other fragments are treated with RHF. The dimer calculations use the MCSCF wave function only if one of the monomers composing the dimer is of MCSCF type. In all other cases, dimers are handled using RHF. The MCSCF active space definition is the same in MCSCF monomers and dimers.

In single-reference-based dynamic correlation methods, such as MP2 or CC, molecular orbitals are obtained from *n*-mer RHF calculations. Then electron correlation energy is computed for *n*-mers and is consequently combined into the total correlation energy

following Eq. (2).

$$E^{\text{FMO1-corr}} = \sum_{I}^{N} E_I^{\text{corr}}$$

$$\Delta E^{\text{FMO2-corr}} = \sum_{I>J}^{N} \left(E_{IJ}^{\text{corr}} - E_I^{\text{corr}} - E_J^{\text{corr}}\right)$$

$$\Delta E^{\text{FMO3-corr}} = \sum_{I>J>K}^{N} \begin{Bmatrix} \left(E_{IJK}^{\text{corr}} - E_I^{\text{corr}} - E_J^{\text{corr}} - E_K^{\text{corr}}\right) - \left(E_{IJ}^{\text{corr}} - E_I^{\text{corr}} - E_J^{\text{corr}}\right) - \\ \left(E_{JK}^{\text{corr}} - E_J^{\text{corr}} - E_K^{\text{corr}}\right) - \left(E_{KI}^{\text{corr}} - E_K^{\text{corr}} - E_I^{\text{corr}}\right) \end{Bmatrix} \quad (16)$$

$$E^{\text{FMO2-corr}} = E^{\text{FMO1-corr}} + \Delta E^{\text{FMO2-corr}}$$

$$E^{\text{FMO3-corr}} = E^{\text{FMO1-corr}} + \Delta E^{\text{FMO2-corr}} + \Delta E^{\text{FMO3-corr}}$$

The correlated energy $E^{\text{FMO}n\text{-corr}}$ is added to $E^{\text{FMO}n\text{-RHF}}$ (n = 2 or 3). The n-mer contributions E_X^{corr} to the total correlation energy, where X is I, IJ or IJK, are computed using integrals and Fock matrices computed in the preceding RHF. That is, they contain the external electrostatic potential and the projection operators added to one-electron integrals, otherwise usual *ab initio* expressions for individual n-mer correlation energies are used.

Although the formal equation is shared for uncorrelated and correlated total energies, the accuracy of the two is quite different. The errors in the former arise from the neglect of exchange and restricting density distribution within n-mers; the errors in the latter are mostly due to restricting the short-ranged correlation within n-mers.

In the case of CI, in the present state of development one has to follow the multilayer treatment described below, and the excited states are computed just for one fragment in the external field of other fragments, and no expression similar to Eq. (2) is used.

The following notation for the FMO method is introduced: FMO*n*-*M*, where *M* is the wave function based on the *n*-body expansion. For example, FMO2-RHF implies using the two-body expansion with the RHF wave function and FMO3-CCSD(T) is the three-body method with the CCSD(T) wave function. The fragment size is sometimes specified as FMO*n*-*M*/*m*, where *m* is the number of residues or water molecules per fragment.

1.2.7 Interfragment separation

The interfragment separation can be easily defined. The unitless distance between n-mer A and monomer L is defined as the closest distance between all pairs of atoms in A and L, divided by the sum of their van der Waals radii W.

$$R_{AL} = \min_{i\in A,\, j\in L} \{R_{ij}\},\ R_{ij} = \frac{|\mathbf{R}_i - \mathbf{R}_j|}{W_i + W_j} \quad (17)$$

where $\mathbf{R}_i$ are atomic coordinates.

Thus defined distances can be used to greatly improve the computational efficiency in several ways. First, the external two-electron Coulomb potential u_L^2 exerted by fragment L upon n-mer A is computed as:

$$u_L^2(\mathbf{r}) = \int \frac{\rho_L(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\, \mathrm{d}\mathbf{r}'$$
$$\langle \mu | u_L^2 | \nu \rangle = \sum_{\rho\sigma \in L} D_{\rho\sigma}^L (\mu\nu|\rho\sigma) \tag{18}$$

where μ,ν are atomic orbitals of n-mer A.

The Mulliken atomic population approximation [37] denoted by RESPAP is used in computing the ESP in Eq. (18) if the interfragment distance R_{AL} is larger than the RESPAP threshold.

$$\sum_{\rho\sigma \in L} D_{\rho\sigma}^L (\mu\nu|\rho\sigma) \approx \sum_{\rho \in L} P_\rho^L (\mu\nu|\rho\rho) \tag{19}$$

where P_ρ^L are orbital populations (usually, Mulliken) for fragment L. Computationally speaking, this approximation reduces the expensive two-electron calculation by a factor of N_{BF} (the number of basis functions in A).

If the interfragment distances are larger than the RESPPC threshold, then the Mulliken point charge approximation [37] is used.

$$\sum_{\rho\sigma \in L} D_{\rho\sigma}^L (\mu\nu|\rho\sigma) \approx \sum_{\alpha \in L} Z_\alpha^L \left\langle \mu \left| \frac{1}{|\mathbf{r} - \mathbf{R}_\alpha|} \right| \nu \right\rangle \tag{20}$$

where Z_α^L are atomic populations (usually, Mulliken) for fragment L. This approximation is even more powerful than the previous one and the amount of calculations is reduced by roughly another factor of N_{BF}.

The total electron densities for far separated dimers can be described to a very good approximation by the tensor sum of their monomer densities, that is, there is no need to reconverge the density within the dimer, since converged monomer densities are available. This approximation is denoted by RESDIM [37] and is applied to dimer calculations if an interfragment distance is larger than RESDIM. The electron densities, however far separated, interact with each other due to the slow decaying nature of the Coulomb interaction. Hence, even though there is no need to converge the dimer density, the amount of electrostatic interaction should be evaluated and added to the monomer energies as follows:

$$E'_{IJ} \cong E'_I + E'_J + \mathrm{Tr}(\mathbf{D}^I \mathbf{u}^{1,J(J)}) + \mathrm{Tr}(\mathbf{D}^J \mathbf{u}^{1,J(I)}) + \sum_{\mu\nu \in I} \sum_{\rho\sigma \in J} D_{\mu\nu}^I \, D_{\rho\sigma}^J (\mu\nu|\rho\sigma) \tag{21}$$

where $\mathbf{u}^{1,K(L)}$ are one-electron Coulomb potentials exerted by fragment L into fragment K.

In a similar way, separated trimer approximation is introduced. SCF computation of a trimer IJK is not performed if the separation S_{IJK} is larger than a threshold, denoted by RITRIM [22]. The separation is defined as follows. Supposing that for a trimer IJK monomers I and J are the closest among all monomer pairs, then the separation S_{IJK} is

defined as $S_{IJK} = \min(R_{IK}, R_{JK})$, where R_{IK} is the distance between fragments I and K. That is, S_{IJK} is the distance between the closest dimer and the remaining monomer composing the trimer. There is a difference from the dimer case, since the electrostatic interaction is pairwise, and the three-body correction for a far separated trimer is exactly zero, that is, the separated trimer energy can be computed from monomer and dimer energies without any additional trimer-specific calculations.

$$\begin{aligned} E'_{IJK} &\cong E'_I + E'_J + E'_K + (E'_{IJ} - E'_I - E'_J) + (E'_{IK} - E'_I - E'_K) + (E'_{JK} - E'_J - E'_K) \\ &= E'_{IJ} + E'_{IK} + E'_{JK} - E'_I - E'_J - E'_K \end{aligned} \tag{22}$$

Next, we introduce correlation-specific approximations. A correlated dimer calculation is not performed if the corresponding interfragment distance is larger than the threshold RCORSD [32], which is chosen to be not larger than RESDIM, so the correlated dimer calculations are performed only if dimer SCF calculations are done. Similarly, the trimer calculation is omitted if the separation S_{IJK} is larger than RCORST [35], which should not exceed RITRIM.

Practically speaking, the values we normally use in uncorrelated calculations are: RESPAP = 1.0, RESPPC = 2.0 and RESDIM = 2.0 for the two-body methods, although we usually raise them by 0.5 if diffuse functions are present or a better accuracy is required. In the three-body methods there is a current program limitation that RESPAP and RESPPC approximations should not be used (RESPAP = 0, RESPPC = 0) and the recommended value of RESDIM is twice as large as RITRIM (or no RESDIM approximation, RESDIM = 0), e.g., RITRIM = 2.0 and RESDIM = 4.0.

Since correlated calculations are much more expensive than uncorrelated ones, we usually raise uncorrelated thresholds by 0.5, compared with the corresponding correlated ones. So, for two-body correlated methods the usual values are RESPAP = 1.5, RESPPC = 2.5, RESDIM = 2.5, RCORSD = 2.0 and for three-body methods we use RESPAP = 0, RESDIM = 0, RESDIM = 5.0, RITRIM = 2.5, RCORSD = 4.0 and RCORST = 2.0.

1.2.8 The multilayer approach

It has long been recognized that in many cases it is possible to select one part of a system where the phenomenon under investigation occurs and apply the most accurate methods for such a part, while cheaper methods can be used to describe the remaining (environment). The QM/MM method is an example of this approach, where the reaction centre is described with usually *ab initio* quantum chemical methods and the rest with a force field. Another example is the ONIOM method by Svensson *et al.* [12]. Despite certain similarities, the two approaches are significantly different. In the former, the whole system is always handled, with two different ways to treat its parts. In the latter, however, the higher layers are computed with a complete neglect of the remaining environment, and such omission is hoped to be compensated for by an additive scheme of summing a separate environmental contribution with the higher-layer energetics.

We have proposed the multilayer FMO (MFMO) method [21], where all fragments are divided into layers and one can assign a different basis set and/or a different wave function for each layer. Layer boundaries coincide with some fragment boundaries, which are treated as in the original unilayer approach. If the essence of the MFMO method is to be

given in a few words, the only differences from the unilayer method are: (a) the ESPs are computed using monomer electron densities from the appropriate layer, and (b) dimer and trimer calculations are performed at the lowest level of all monomers in a dimer or trimer. The basic MFMO equation is:

$$E = \sum_{I}^{N} E_I^{L_I} + \sum_{I>J}^{N} \left(E_{IJ}^{L_{IJ}} - E_I^{L_{IJ}} - E_J^{L_{IJ}}\right) + \sum_{I>J>K}^{N} \left\{ \begin{array}{l} \left(E_{IJK}^{L_{IJK}} - E_I^{L_{IJK}} - E_J^{L_{IJK}} - E_K^{L_{IJK}}\right) - \left(E_{IJ}^{L_{IJK}} - E_I^{L_{IJK}} - E_J^{L_{IJK}}\right) - \\ \left(E_{JK}^{L_{IJK}} - E_J^{L_{IJK}} - E_K^{L_{IJK}}\right) - \left(E_{KI}^{L_{IJK}} - E_K^{L_{IJK}} - E_I^{L_{IJK}}\right) \end{array} \right\} + \ldots \quad (23)$$

where L_I denotes the layer to which fragment I belongs; $L_{IJ} = \min(L_I, L_J)$ and $L_{IJK} = \min(L_I, L_J, L_K)$. Monomer ($E_I^{L_1}$), dimer ($E_{IJ}^{L_2}$) and trimer ($E_{IJK}^{L_3}$) energies are obtained for corresponding n-mers ($n = 1,2,3$) at the level defined by layers L_1, L_2 and L_3, respectively. Eq. (23) reduces to Eq. (2) for the case of one layer. Monomer energies E_I enter Eq. (23) at the level L_I whereas dimer and higher n-body corrections enter at the lowest level of all monomers, making the corresponding n-mer. It can also be seen that monomer energies E_I must be obtained for all layers $L = 1 \ldots L_I$ and not just for layer L_I (as needed for many-body corrections).

Higher layers correspond to higher computational levels, which are listed in ascending order. For example, FMO2-RHF/3-21G:MP2/6-31G* implies two-body MP2 with the 6-31G* basis set for the higher layer and RHF and the 3-21G basis set for the lower layer (environment).

Thus, in the computational scheme that is described in detail in [21], the calculation proceeds by computing all layers L starting from the lowest level $L = 1$, and only monomers with $L_I \geq L$ are reconverged (the densities for the other monomers I' needed to compute their ESPs are taken from the already calculated layer $L_{I'}$). Then within each layer L only those dimer and trimer calculations are performed where at least one fragment belongs to L. All monomers I are converged L_I times, all dimers and trimers are computed exactly once. The scheme can be visualized in Fig. 1.6.

Although the primary reason for development of the MFMO method is to reduce computational costs and permit more nearly complete basis sets and more accurate electron correlation methods for the important part of the system, there is another usage that is unique to the multilayer approach in one specific case, when there is only one fragment in the highest layer. In such a case, only monomer and no dimer and trimer calculations need be performed for the highest layer. In particular, computing excited states using the CI approach is in general restricted to such method.

1.2.9 Pair interaction analysis

For the purposes of the interaction analysis, the two-body energy expression in Eq. (2) can be rewritten as:

$$\begin{aligned} E &= \sum_{I}^{N} E'_I + \Delta E^{\text{int}} \\ \Delta E^{\text{int}} &= \sum_{I>J}^{N} \left(E'_{IJ} - E'_I - E'_I\right) + \sum_{I>J}^{N} \text{Tr}\left(\Delta \mathbf{D}^{IJ} \mathbf{V}^{IJ}\right) \end{aligned} \quad (24)$$

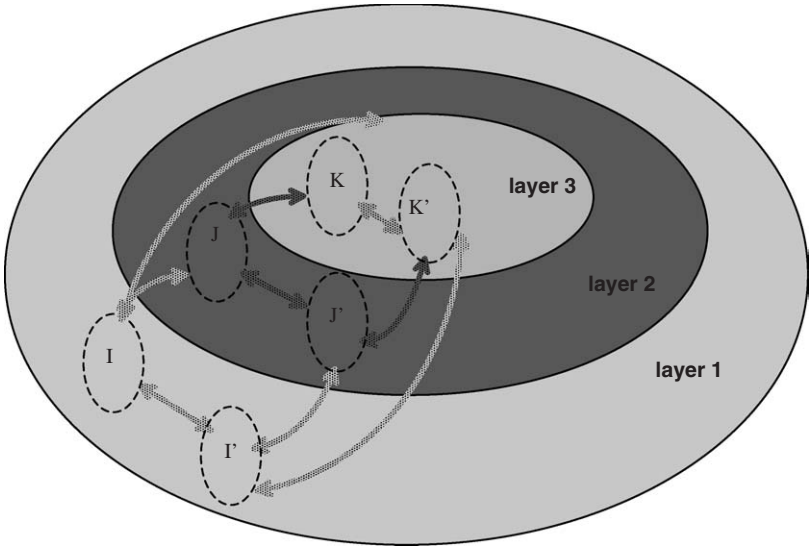


Fig. 1.6. The multilayer FMO scheme in the case of three layers. Fragments are denoted by *I*, *J* and *K* (the same letter in each layer). The level of dimer calculation is shown as arcs with the layer shade: for instance, *I*-*I*′and *I*-*J* dimers are computed using the wave function and the basis set for layer 1 (light shade of grey). Arc shades also coincide with the level at which the ESPs are computed, but only in the ascending order of layers, e.g., the ESPs due to fragments *I* and *J* are added to the Hamiltonian for fragment *K*, using the corresponding level (light and dark shades, respectively).

This is an entirely equivalent formulation [37] to the two-body expansion in Eq. (2) in the case when approximations to the ESPs (RESPAP and RESPPC) are not used. With ESP approximations, the two expressions in Eqs (2) and (24) result in a somewhat different total energy E, and the form given in Eq. (24) was found to be the better one due to a better balance between ESP approximations in monomers and dimers. A similar expression can be derived for the three-body expansion, however, since the present implementation does not allow using ESP approximations, for the three-body methods the original Eq. (2) is used. The correlation energy is always expanded according to Eq. (16).

The physical meaning of the energies E' in Eq. (24) is that they represent the internal energies of n-mers within the total system, polarized by the environment, but with the energy of the environment subtracted. For example, if one considers a water cluster, then E_I for each water molecule placed in its own fragment rapidly grows if the cluster size is increased, due to larger Coulomb interaction. The E_I', on the other hand, differ by a small amount from the energy of individual water molecules. The difference between E_I' (the energy of 'fragments in molecule') and E_I^0 (the energy of free molecules) is the amount of destabilization caused by disturbing fragment densities from their optimum distribution in individual molecules. This destabilization is called the one-body polarization, and the gain that offsets such destabilization comes from pair interaction energies, divided into two contributions.

The first contribution to ΔE^{int} given by $E'_{IJ} - E'_I - E'_J$ contains interfragment electrostatic interaction and two-body polarization (due to interfragment charge transfer). The second contribution given by $\text{Tr}(\Delta\mathbf{D}^{IJ}\,\mathbf{V}^{IJ})$ defines the density relaxation energy within dimers (due to interfragment charge transfer) and is often one order smaller in magnitude than the former one.

The charge transfer energy thus is split into two contributions: $E'_{IJ} - E'_I - E'_J$ terms contain the amount of electrostatic interaction due to transferred charge, while $\text{Tr}(\Delta\mathbf{D}^{IJ}\,\mathbf{V}^{IJ})$ terms contain the energy required to redistribute the density due to charge transfer. The monomer calculations in the FMO method are always performed with the total electron count fixed, and the charge transfer phenomenon is accounted for in dimer and trimer calculations, when the electron density is allowed to relax within corresponding *n*-mers.

Now we address the issue of reassigning one proton for the bond-detached atoms, mentioned earlier. By construction, a dimer that includes both fragments between which a bond is fractioned has this bond intact, so the difference that arises due to this proton reassignment only appears in the monomer one-electron Hamiltonian $H_{\mu\nu}$ (Eq. (6)) and the external one-electron electrostatic field $V_{\mu\nu}$ (Eq. (7)). If no proton reassignment took place (Fig. 1.1(a)), the total Hamiltonian for the left fragment (F1) has the following additive contributions:

$$h^{\text{F1}}_{\mu\nu} = \sum_{\substack{A=1 \\ A\in\text{F1}}}^{9} \left\langle \mu \left| -\frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} \right| \nu \right\rangle \tag{25}$$

$$v^{\text{F1}}_{\mu\nu} = \sum_{\substack{A=1 \\ A\in\text{F2}}}^{9} \left\langle \mu \left| -\frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} \right| \nu \right\rangle \tag{26}$$

where $\mathbf{h}^{\text{F1}}$ is a part of the one-electron Hamiltonian $\mathbf{H}^{\text{F1}}$. The sum $h^{\text{F1}}_{\mu\nu} + v^{\text{F1}}_{\mu\nu}$ of the one-electron Coulomb attraction due to atoms in fragment F1 ($h^{\text{F1}}_{\mu\nu}$) and the external one-electron field due to atoms in fragment F2 ($v^{\text{F1}}_{\mu\nu}$) is added to the total Hamiltonian of F1: $\tilde{H}^{\text{F1}}_{\mu\nu}$, $\mu,\nu,\in$ F1. If one proton is assigned to F2 (Fig. 1.1(b)), then simply one attraction integral term is moved from $\mathbf{h}^{\text{F1}}$ in Eq. (25) to $\mathbf{v}^{\text{F1}}$ in Eq. (26), leaving their sum invariant. To complete the perfect agreement one has to add a ghost atom and corresponding projection operators to the scheme described in Fig. 1.1(a). Thus, proton reassignment does not change the Hamiltonians and electron densities. The *n*-mer energies E_X are only modified because the nuclear repulsion term E^{NR} is redefined in Eq. (9). The total energy E is, however, invariant, since the sum of nuclear repulsion energies E_{NR} is invariant.

On the other hand, the definition of E' energies that are used in the pair interaction analysis excludes the environment so it is not the same depending on which scheme in Fig. 1.1 is used. The E' from the scheme depicted in Fig. 1.1(a) are not very useful, because they would correspond to charged fragment attraction, which is an artefact of the fragmentation. However, using the scheme in Fig. 1.1(b) results in fragments with their original charge preserved.

The only demerit in the Fig. 1.1(b) fragmentation scheme lies in $\Delta E'_{IJ}$ terms for dimers that have a fractioned bond between monomers *I* and *J*. Such $\Delta E'_{IJ}$ terms include Coulomb interaction across the fractioned bond, their values are on the order of -14 Hartree and not

very useful. Work is in progress to provide for a physically reasonable way to correct for this artefact of fragmentation.

1.2.10 An example of the FMO interaction analysis: ethanol and water

As an example of interaction analysis, let us take a very simple system of ethanol and water, which is distributed in GAMESS under the title of the basic FMO tutorial. Ethanol is divided into two fragments, CH_3 and CH_2OH, so that there are three fragments, numbered ascendingly in the order CH_3, CH_2OH and H_2O. For simplicity, the STO-3G basis set is used (that is also known to produce good hydrogen-bond energies, due to error cancellation). The analysis below is applied to a simple pair interaction between ethanol and water, to illustrate the meaning of the corresponding interaction terms. The geometry and the corresponding interaction diagram are depicted in Fig. 1.7. The energetics is given in Table 1.1.

When ethanol and water are brought in close contact, they destabilize each other relative to their corresponding free states. This energy is called the (monomer) destabilization polarization energy. Strictly speaking, the optimum geometry of individual molecules is different from their geometry in the interacting system. The difference in energy due to such geometry differences is called the deformation energy, and for simplicity is not discussed here. In the FMO method, the destabilization polarization energy $\Delta E_I^{\mathrm{PLd}}$ can be defined as:

$$\Delta E_I^{\mathrm{PLd}} \equiv E_I' - E_I^0 \tag{27}$$

where E_I' is the 'fragment I in molecule' energy defined in Eq. (10) and E_I^0 is the free fragment I energy. In this case ethanol (subsystem A) is divided in two fragments, so we actually used the energy of dimer 12 to obtain ethanol polarization energy: $\Delta E_A^{\mathrm{PLd}} \equiv E_{12}' - E_A^0$, and

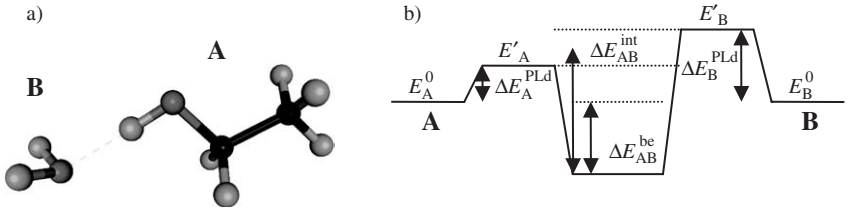


Fig. 1.7. Interactions in ethanol-water complex. Free molecules A and B are mutually polarized by the amount of $\Delta E_A^{\mathrm{PLd}}$ and $\Delta E_B^{\mathrm{PLd}}$ and the amount of pair interaction energy is shown as $\Delta E_{AB}^{\mathrm{int}}$. The binding energy $\Delta E_{AB}^{\mathrm{be}}$ is measured against the free subsystem energies $E_A^0 + E_B^0$ (see Color Plate 1).

Table 1.1 The energies (STO-3G) of subsystems A (C_2H_5OH) and B (H_2O): E^0 is free fragment energy, E' is 'fragment in molecule' energy and ΔE^{PLd} is the polarization energy

	E^0, a.u.	E', a.u.	ΔE^{PLd}, kcal/mol
A	−152.132596	−152.132329	0.167
B	−74.965849	−74.965781	0.043
A+B	−227.098445	−227.098109	0.210

for water (subsystem B) we have $\Delta E_B^{\mathrm{PLd}} \equiv E_3' - E_B^0$. As can be seen from Table 1.1, the polarization energies of ethanol and water are 0.167 and 0.043 (kcal/mol), respectively. E_A^0, and E_B^0 are obtained from separate calculations on free subsystems A and B.

The destabilization due to mutual polarization is compensated for by interaction. The interaction energy between ethanol and water is obtained as a sum of all pair energies between the two subsystems, that is, interactions between fragments 1,2 (ethanol) and fragment 3 (water):

$$\Delta E_{AB}^{\mathrm{int}} = (E_{31}' - E_3' - E_1') + (E_{32}' - E_3' - E_2') + \mathrm{Tr}(\Delta \mathbf{D}^{31}\mathbf{V}^{31}) + \mathrm{Tr}(\Delta \mathbf{D}^{32}\mathbf{V}^{32}) \qquad (28)$$

The sum of the first two terms is the dimer polarization and interfragment electrostatic energy and it is equal to -1.409 kcal/mol. The remaining two terms correspond to the density relaxation energy and their sum is equal to -4.410 kcal/mol. The total pair interaction energy -5.819 is thus obtained strictly from one single point FMO calculation of the total system. *The pair interaction energy is the amount of interaction between 'fragments in molecule', it is not the same as the traditional binding energy measured relative to non-interacting systems*.

The pair interaction energy is, however, close to the binding energy. The binding energy $E_{AB}^{\mathrm{be}} = E_{AB} - E_A^0 - E_B^0$ as obtained by subtracting free A and B energies from their complex energy is equal to -5.911 kcal/mol. We note that in this case E_A^0 (and E_B^0) computed with *ab initio* and FMO methods are identical, since in A (and B) there are less than three fragments, in which case the FMO energy is exact. As a practical FMO application, Fukuzawa *et al*. [38] reported binding energies of human oestrogen receptor with its ligands.

There is ongoing work to clearly divide contributions to pair interaction energies. In the configuration analysis for fragment interaction (CAFI) [39], charge transfer and interfragment polarization energies can be computed. To complete the interaction energies, deformation and correlation energies should be taken into account, in the latter case the extra contribution is of the form $E_{IJ}^{\mathrm{corr}} - E_I^{\mathrm{corr}} - E_J^{\mathrm{corr}}$.

Interaction analysis for a real biological system was conducted by Nemoto *et al.* [40], who studied the complex of pheromone-binding protein of the silkworm moth, *Bombyx mori* (BmPBP) with its ligand Bombykol, also evaluating the deformation energy. In Fig. 1.8 we provide a typical example of the pair interaction analysis. In order to study details of the ligand–protein interaction, the ligand was divided into four small fragments, and in Fig. 1.8 all pair interactions of the hydrophobic tail piece (C_3H_7) of Bombykol with each residue in BmPBP are shown. Both attractive (negative) and repulsive (positive) interactions are observed, and based upon the residue serial numbers one can suggest protein structure mutations to alter the protein functionality. Two other examples of the FMO interaction analysis performed for biological systems can be found in [41,42].

It should be noted that in order to introduce monomer polarization energies, one has to define 'free monomer' energies, which is easy when monomers are stand-alone molecules but not so when monomers are connected by covalent bonds. Work is in progress to extend the definition of polarization energies into such cases. The pair interaction analysis, however, does not require any such definitions and can be used in its present state for arbitrary systems (pair interaction for dimers connected by a fractioned bond have only the $\mathrm{Tr}(\Delta \mathbf{D}^{IJ}\mathbf{V}^{IJ})$ term).

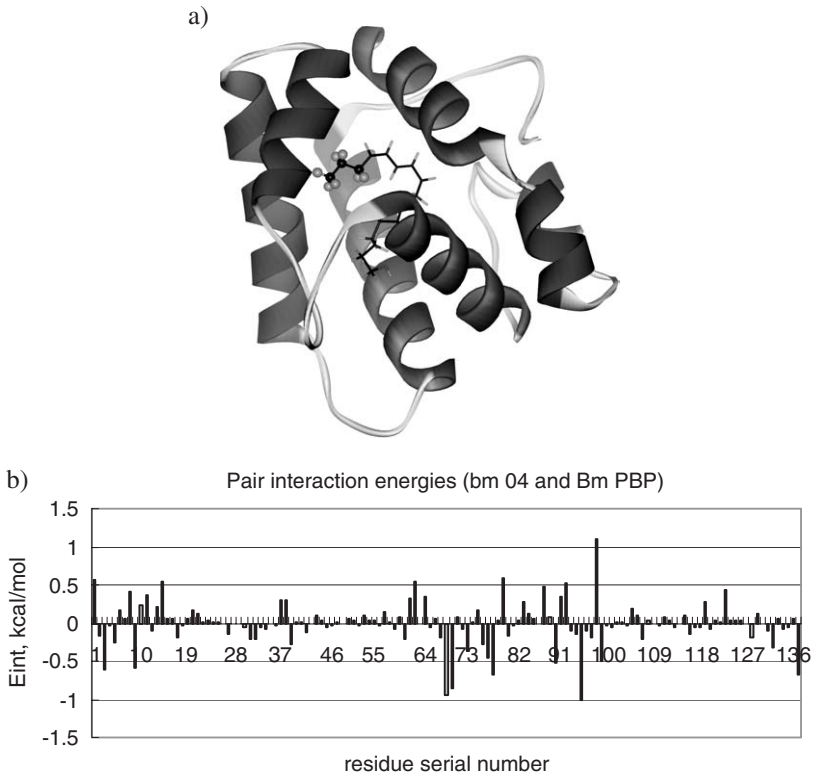


Fig. 1.8. Pair Interaction analysis in BmPBP: (a) the structure of the BmPBP complex, (b) pair interaction of a part of bombykol ligand (denoted by bm04) with each residue in BmPBP. The protein and ligand are treated on the same footing. The protein is shown schematically by its secondary structure units, the atoms of the ligand are shown as sticks, and, in addition, a part of the ligand for which the interaction analysis is given, is shown with balls and sticks (see Color Plate 2).

If one is interested in interaction energies between certain parts of the total system, then simply the sum of the corresponding pair interaction has to be computed.

Finally, the amount of charge transfer can be easily defined if one relies on Mulliken charges and defines the amount of charge transfer as:

$$\begin{aligned} \Delta q^{\mathrm{CT}}_{I\to J} &\equiv q^I_I - q^{IJ}_I \\ q^X_I &\equiv -\sum_{\alpha\in I} Z^X_\alpha + Z_I \\ Z^X_\alpha &= \sum_{\mu,\nu\in\alpha} D^X_{\mu\nu} S^X_{\mu\nu} + \frac{1}{2}\sum_{\mu\in\alpha}\sum_{\nu\notin\alpha} D^X_{\mu\nu} S^X_{\mu\nu} \end{aligned} \tag{29}$$

where X is I or IJ, α and β denote atoms, μ and ν run over atomic orbitals. q^{IJ}_I and q^I_I are the total charges of I in dimer IJ and monomer I, respectively; Z_I is the number of protons

in I, and $\mathbf{D}^X$ and $\mathbf{S}^X$ are the electron density and overlap integrals of X, respectively. The amount of charge transfer between fragments in Eq. (29) corresponds to charge transfer between 'fragments in molecule', although due to charge normalization in each n-mer monomer charges q_I^I are equal to the total charge on fragment I. Therefore, q_I^I are the same for 'fragments in molecule' and free fragments and the effect of environment comes through charge redistribution in dimers, which determines q_I^{IJ}.

The sum of all atomic Mulliken charges in each n-mer calculation is normalized to the number of electrons. Using the Mulliken charge of A that comes from 'fragments in molecule', one obtains $q_A^A = 0.000$ (which is equal to 0, the total charge of C_2H_5OH) and $q_A^{AB} = -0.049$ (the charge of C_2H_5OH within the complex), thus $\Delta q_{I\to J}^{CT} = +0.049$ and this is the amount of charge transferred from ethanol (A) to water (B). In terms of electron count (which has the opposite sign as electrons have -1 charge), 0.049 electrons are transferred in the opposite direction from water to ethanol.

Moreover, it is straightforward to see all atomistic details of charge transfer by looking at individual atomic populations Z_α^X. Oxygen and hydrogen atoms in water loose 0.012 and 0.037 electrons, respectively, which are transferred mostly to hydroxylic oxygen (0.038) in ethanol and the rest goes to other atoms (accompanied by charge redistribution in C_2H_5OH).

Finally, the effect of polarization can be briefly inspected on the atomistic level, e.g., comparing free water and water as 'fragment in molecule'. Looking at Mulliken charges, one obtains that due to polarization by ethanol, oxygen atoms in water draw additional 0.014 electrons from both hydrogen atoms (which corresponds to the 0.043 kcal/mol polarization energy discussed above).

1.3 ACCURACY OF THE FMO METHOD

Since the FMO method is designed to reproduce full *ab initio* properties, e.g., energies, electron densities and derived quantities, such as energy gradients and dipole moments, it is very important to determine how close a numeric agreement is reached. Fulfilling this need, we performed systematic studies comparing the FMO accuracy for each wave function type [22,30,31,32,35]. The accuracy is defined as the difference between the FMO properties and the *ab initio* ones, it has nothing to do with experiment whatsoever. If the underlying *ab initio* method or the basis set is inappropriate for a given problem, the same applies to the corresponding FMO method.

The general tendency observed in all FMO tests is that the error in absolute energy grows linearly with system size. The absolute values of the uncorrelated (RHF) and dynamic correlation (MP2, CC) energy tend to increase linearly as the system grows, and so do the errors in the corresponding FMO method. The static correlation (MCSCF) energy is different due to the very different nature of this quantity which changes little with system size (that is, if the active space is fixed and one only increases the number of atoms in the environment), and the FMO-MCSCF error is also nearly constant.

The energy gradient and dipole moment follow the energy trends and will not be discussed here. Numeric tests can be found in the corresponding FMO publications [22,30–32].

As discussed above, the two-body expansion for most cases has a very satisfactory accuracy. The only common exception is hydrogen bonds, where three-body charge transfer effects are fairly large (Figs 1.5(b) and 1.5(c)). Typical examples of such systems are α-helices (e.g., in polypeptides) and water clusters. While the FMO methods describe β-strand conformers with very high accuracy, the accuracy for α-helical conformers is somewhat lower.

The absolute errors in the RHF and DFT total energies for α-helices and β-strands of polyalanine $(ALA)_n$ are depicted in Fig. 1.9. The general comparative trend in the accuracy for FMO-DFT and FMO-RHF is that the former has two to three times larger errors. Otherwise, the accuracy trends in both are nearly identical, so we focus on DFT hereafter.

First, let us consider α-helices, shown for FMO-DFT in part Fig. 1.9(a). The error for two-body methods is not small, being as large as about 44 milliHartree (or about 27 kcal/mol) even if two residues per fragment are used. For the three-body method, however, the error is 1.5 milliHartree or 0.9 kcal/mol. This means that the absolute value of the total energy (about $-10\,135$ Hartree) for the system as difficult as α-helices, is reproduced in FMO3-DFT with the relative accuracy of 99.999999852 percent or 148 ppb error.

Let us consider now β-strands, shown for FMO-DFT in Fig. 1.9(b). Beyond the fastest case of two-body, one residue per fragment calculation, the errors are practically zero. If one uses two residues per fragment, even for the two-body method, the error is only 1.7 milliHartree or 1.0 kcal/mol. With the three-body method, the error drops further to 0.13 milliHartree. This means that in β-strands that represent an important part of biological systems (random coil) the accuracy for FMO3-DFT is 99.999999987 per cent or 13 ppb error.

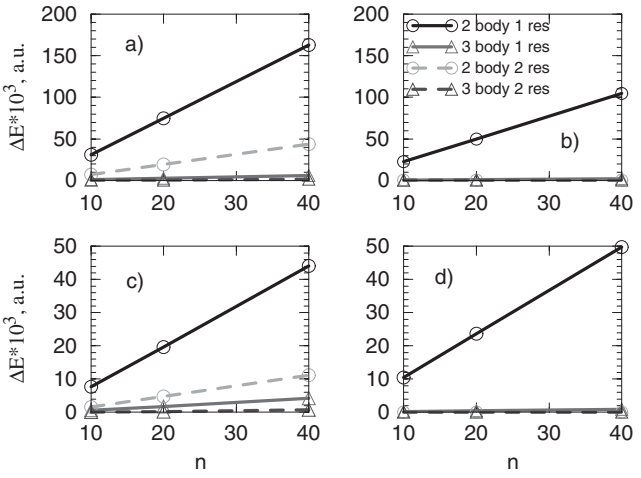


Fig. 1.9. The error in energy (milliHartree) for FMO-DFT: (a) α-$(ALA)_n$, (b) β-$(ALA)_n$ and for FMO-RHF: (c) α-$(ALA)_n$, (d) β-$(ALA)_n$, measured against the corresponding *ab initio* values. The basis set is 6–31G* throughout (see Color Plate 3).

An estimate of the FMO error in energy for a system containing only α-helices and β-strands is as follows:

$$\varepsilon(n_\alpha, n_\beta) \approx a_\alpha + b_\alpha n_\alpha + a_\beta + b_\beta n_\beta \tag{32}$$

where the values of a and b for alanine polypeptides are summarized in Table 1.2 and n is the number of residues of each secondary structure type. The values were determined for pure conformers and further study is needed to establish the validity of Eq. (28) if applied to mixed conformers, in particular whether the errors will be additive or not, and whether the α and β error contributions are to be added as absolute values.

The other typical part of biological systems (β-sheets) has a feature similar to α-helices, that is, a network of interacting hydrogen bonds, where explicit three-body effects (including charge transfer) are substantial and the accuracy of the FMO methods for β-sheets is similar to α-helices.

The following important observation should be made about the data in Table 1.2. By looking at slope coefficients b, one can observe that they are negative for two-body methods (with the one exception of b_β for FMO2-DFT/2), the two-body FMO total energies are below the *ab initio* energies and the gap grows with system size. On the other hand, the three-body methods have positive slopes (with the exception of b_β for FMO3/1), and the corresponding FMO energies are above the *ab initio* energy.

Water molecules, that are frequently included in crystallized structures of proteins, represent another example of systems with strong hydrogen bonds and three-body charge transfer, which is only properly accounted for at the three-body level (FMO3). In general, accuracy trends for water molecules resemble those of α-helices and have been considered in detail in previous studies. It should suffice to suggest placing two water molecules per fragment for better accuracy. If pure water clusters are to be computed, then three-body methods with one or two molecules per fragment should be preferred, depending upon the desired accuracy.

Previous discussion was concerned with RHF and DFT. It should be noted that although DFT does include electron correlation in the correlation potential, due to the

Table 1.2 Coefficients (in milliHartree) describing linear dependency of error in energy $\varepsilon \approx a + nb$ for α-helix and β-strand conformers of $(ALA)_n$, for the k-body FMO expansion and l residues per fragment (denoted by FMOk/l). The number of significant figures kept corresponds to the degree of observed regularity in linear dependence in each case

	Method	FMO2/1	FMO2/2	FMO3/1	FMO3/2
a_α	DFT	12.58	4.96	−1.00	−0.42
b_α	DFT	−4.37	−1.21	0.17	0.042
a_β	DFT	5.315	−0.059	0.072	−0.029
b_β	DFT	−2.757	0.0114	−0.0440	0.0040
a_α	RHF	4.35	1.50	−0.60	−0.20
b_α	RHF	−1.20	−0.31	0.12	0.023
a_β	RHF	2.550	0.0020	0.063	−0.0056
b_β	RHF	−1.308	−0.0012	−0.025	0.00077

self-consistent nature of the method the correlated and uncorrelated errors are not easily separable. In the case of other methods, such and MP2 or CC, it is easy to separate the correlation energy and investigate its accuracy. High accuracy of the FMO-based MP2, MCSCF and CC methods was established in previous studies, where the following largest systems were computed with FMO and *ab initio* methods: 10 residues or 32 water molecules for MP2, nine residues or phenol in 32 water molecules for MCSCF, four residues or eight water molecules for CCSD(T). For these systems, the errors in milliHartree did not exceed: 2.6 (MP2), 0.88 (MCSCF) and 1.3 (CCSD(T)) using the two-body expansion. To illustrate the correlated energy accuracy, the FMO-CCSD(T) data are summarized in Table 1.3.

Next, we address the issue of the accuracy in reproducing energy differences, where we compute the relative stabilities of α-helices and β-strands of polyalanine and its mutants as obtained by FMO and *ab initio* methods, based on data available from previous FMO publications. Summarizing the data in Table 1.4, the RHF relative stabilities are reproduced with the errors of at most 3.5, −7.0, 3.3 and 0.5 (kcal/mol) for FMO2/1, FMO2/2, FMO3/1 and FMO3/2, respectively. For DFT the corresponding values are 36.3, −27.5, 4.9 and 0.9 (kcal/mol). The structures were generated by Hyperchem and are not optimized with quantum chemical methods.

It should be noted that β-strands have negligible absolute errors and the error in the relative stabilities comes largely from the errors in describing α-helices. There is some error cancellation, however, so that the relative stabilities for RHF happened to be very small for the two-body method, which is not because of the wave function type per se, but due to error cancellation, as can be verified by comparing data for other basis sets (not shown).

Overall, the error in uncorrelated energy differences as computed by the two-body FMO method comes to a very large extent from the difference in the number of trilateral hydrogen bonds (Figs 1.5(b) and 1.5(c)). In other words, the error is very roughly proportional to the difference in the length of α-helices, β-sheets and the number of cross-links (where the residue chains come together forming such trilateral hydrogen bonds). For the three-body method, the errors are very much reduced and fall below the chemical accuracy.

Table 1.3 The absolute errors (in milliHartree) in the correlation energies $\Delta E^{\mathrm{corr}}_{\mathrm{FMO}n}$ for the FMO*n*-CCSD(T) method using *m* molecules/residues per fragment, relative to *ab initio* CCSD(T) energies (in Hartree). $(\mathrm{GLY})_n$ stands for polyglicine, cc-VDZ denotes the cc-pVDZ basis set without polarization functions

System	Basis set	*m*	$\Delta E^{\mathrm{corr}}_{\mathrm{FMO2}}$	$\Delta E^{\mathrm{corr}}_{\mathrm{FMO3}}$	$E^{\mathrm{corr}}_{ab\ initio}$
$(H_2O)_3$	cc-pVDZ	1	−0.268	0.000	−0.647995
$(H_2O)_4$	cc-pVDZ	1	−0.218	0.117	−0.866103
$(H_2O)_5$	cc-pVDZ	1	−0.027	0.142	−1.081992
$(H_2O)_8$	cc-pVDZ	1	−0.761	0.214	−1.744307
$(H_2O)_8$	cc-pVDZ	2	−0.214	0.011	−1.744307
$(H_2O)_3$	cc-pVTZ	1	−0.425	0.000	−0.831521
$(GLY)_3$	cc-VDZ	1	0.118	0.000	−1.421172
$(GLY)_4$	cc-VDZ	1	−0.156	−0.037	−1.852665

Table 1.4 Relative stabilities (in kcal/mol) of α-helix and β-strand conformers of polyalanine $(ALA)_n$ and its phenylalanine mutant PHE-$(ALA)_n$, obtained with FMO*n*/*m* and *ab initio* methods (negative values mean that α-helices are more stable). For MP2 and MCSCF only correlation energy difference is shown. FMO*n*/*m* stands for *n*-body expansion using *m* residues per fragment. MCSCF/1 and MCSCF/2 and uni- and bilayer FMO-MCSCF, correspondingly. The basis set is 6–31G* throughout. All energies are in kcal/mol

	System	FMO2/1	FMO2/2	FMO3/1	FMO3/2	*ab initio*
RHF	$(ALA)_{10}$	−20.9	−23.7	−22.2	−22.6	−22.7
RHF	$(ALA)_{20}$	−84.1	−89.7	−85.3	−86.5	−86.7
RHF	$(ALA)_{40}$	−221.9	−232.4	−222.1	−224.9	−225.4
DFT	$(ALA)_{10}$	-28.5	−27.4	−22.2	−22.9	−22.9
DFT	$(ALA)_{20}$	−100.6	−97.2	−83.0	−84.8	−85.0
DFT	$(ALA)_{40}$	−255.0	−246.2	−213.8	−217.8	−218.7
MP2	$(ALA)_{10}$	−32.2	−34.0			−33.8
MCSCF/1	PHE-$(ALA)_4$	−0.253	−0.262			−0.262
MCSCF/1	PHE-$(ALA)_8$	−0.156	−0.159			−0.155
MCSCF/2	PHE-$(ALA)_4$	−0.218	−0.192			−0.262
MCSCF/2	PHE-$(ALA)_8$	−0.121	−0.118			−0.155

The errors in the correlated contributions to relative stabilities are in general small, as can be seen from the available MP2 and MCSCF data. In the latter method, the error in the relative stabilities is in the microHartree regime in all cases (slightly larger for the computationally cheaper multilayer approach).

An important point to realise is that the relative errors discussed here are ultimately the most severe test of the accuracy of energy differences, as it involves paramount structure changes involving all atoms. This is not the case for chemical reactions, when only one localized part of the system (reaction centre) undergoes profound changes, while the rest of the system relaxes in a minor way. It should be expected that the relative errors for such cases are much smaller and it is sufficient to use FMO2/2, whereas for the studies of relative stability of conformers, at least FMO3/1 is recommended if high accuracy is required. As can be seen from the data in Table 1.4, FMO3/2 reproduces *ab initio* relative stabilities with chemical accuracy.

An example of the FMO accuracy and computational efficiency in describing chemical reactions is presented in Table 1.5, where the absolute errors in reaction heat and activation barrier were computed for the following decarboxylation reaction of phenylcyanoacetate anion (PCAA) , catalysed by cyclodextrin β-CD (glucose heptamer):

$$\mathrm{PhCHCNCOO^-} \xrightarrow{\beta\text{-CD}} \mathrm{PhCHCN^-} + \mathrm{CO_2} \tag{33}$$

The errors for the highest level of FMO calculations (single layer) are: 1.2 and 1.3 kcal/mol, for DFT and RHF, respectively. Using smaller basis sets and uncorrelated wave function for the catalyst results in errors of the order of 1–3 kcal/mol, while the FMO calculations are as much as nine times faster compared with *ab initio*. We note that in this reaction the catalyst was too close to the reactant and to obtain even better accuracy one would put the outer layer of atoms (in this case encompassing all β-CD) into the same layer as the reaction centre, or, if nearly perfect accuracy is sought, in the same fragment as the reactant.

Table 1.5 The errors in the activation barrier ΔE_{act} for PCAA decarboxylation reaction, catalysed by β-CD and timings T on 16 3.0 GHz P4 nodes. The values in brackets are the absolute barrier heights against which the errors are measured

Method	ΔE_{act}, kcal/mol	T, m
RHF/6–31(+)G*	[25.6]	219.0
FMO2-RHF/6–31(+)G*	−1.2	160.2
FMO2-RHF/3–21(+)G:RHF/6–31(+)G*	−1.9	22.8
DFT/6–31(+)G*	[16.8]	917.6
FMO2-DFT/6–31(+)G*	−1.3	529.5
FMO2-DFT/3–21(+)G:DFT/6–31(+)G*	−2.1	198.2
FMO2-RHF/6–31(+)G*:DFT/6–31(+)G*	3.0	168.6
FMO2-RHF/3–21(+)G:DFT/6–31(+)G*	1.0	25.5

Table 1.6 Singlet–triplet vertical excitation energies (in eV), computed using the *ab initio* MCSCF, FMO2-MCSCF/*m* and FMO2-RHF:MCSCF/*m* (*m* water molecules are used as one fragment) for PhOH + $(H_2O)_n$, 6–31G*

n	FMO2-MCSCF/1	FMO2-MCSCF/2	FMO2-RHF:MCSCF/1	FMO2-RHF:MCSCF/2	MCSCF
16	3.981	3.982	4.002	4.002	3.982
32	3.999	4.003	4.004	4.004	4.005
64	3.990	3.989	4.005	4.004	3.996

Finally, we address the computation of excited states in the FMO method. At present, two methods can be used for this purpose: MCSCF and CIS. In the former case, one optimizes the wave function (and electron density) for excited states, whereas in the latter case one optimizes the electron density for the ground state (using RHF) and then performs a CI singlet or triplet calculation with these orbitals.

The accuracy of the MCSCF-based excited state calculation can be seen in Table 1.6, where singlet–triplet excitations in phenol solvated in explicit water are given. The error for the highest level of calculations (single layer FMO-MCSCF) is not larger than 0.006 eV. For the much cheaper multilayer method, the error is not larger than 0.02 eV. These errors are much smaller than the actual precision of the MCSCF itself (as compared with experiment). Unilayer FMO2-MCSCF reproduces to a high degree not only the absolute values but also very fine features of electron correlation dependency upon the environment (number of solvent molecules), due to the inclusion of MCSCF phenol–water dimers, absent in the bilayer approach (where only phenol is put into the higher layer).

1.4 PARALLELIZATION

Ab initio quantum chemical methods are in general parallelized with fairly low scalability. There are several reasons for this. One is that the amount of parallel traffic grows more than linearly with the number of CPUs, the other is that it grows very steeply with

system size. In the RHF case, the growth is usually quadratic (Fock matrix, etc.), while for correlated methods it is quartic (transformed two-electron integrals) or steeper. The other problem is load balancing: the larger the system is, the more difficult it is to ensure that each CPU is assigned the same amount of work, and the more cost the scheduling scheme incurs.

Finally, in many, although perhaps not in all, *ab initio* methods there are parts of calculations that are not parallelized. The typical example is the Fock matrix diagonalization in RHF. Attempts to parallelize this step result in large parallel traffic (probably cubic), usually requiring very expensive and fast networking and are not very efficient if a large number of CPUs is present and the system size is not sufficiently large.

A fortunate consequence of fragmentation in the FMO method is the ease of parallelization. It is natural to recognize the fact that *n*-mer calculations can be performed nearly independently on separate CPU groups. A two-level parallelization scheme added to distributed data interface (DDI) is known as generalized distribited data interface (GDDI) [43] and was developed aiming at applications where independent subtasks can be assigned to groups of CPUs. Although up to now the primary target of GDDI is the FMO method, other usages have been reported (numeric gradients in [43]) and more can be conceived.

The scheme of GDDI is shown in Fig. 1.10(a). Nodes need not be of the same type, however, the present parallel library implementation does require that the same binary executable be run on all nodes, so that in practice one can only mix different CPU types within the same architecture.

There are several steps in the FMO method, which are parallelized differently, as determined by the number of subtasks. The main steps are: self-consistent monomer SCF calculation, where the number of subtasks is given simply by the number of monomers N; dimer and, possibly, trimer calculations. The number of dimer and trimer calculations is determined by the approximation values and instead of the upper limit of $N(N-1)/2$ and $N(N-2)/6$, respectively, both numbers are in reality proportional to N (with the

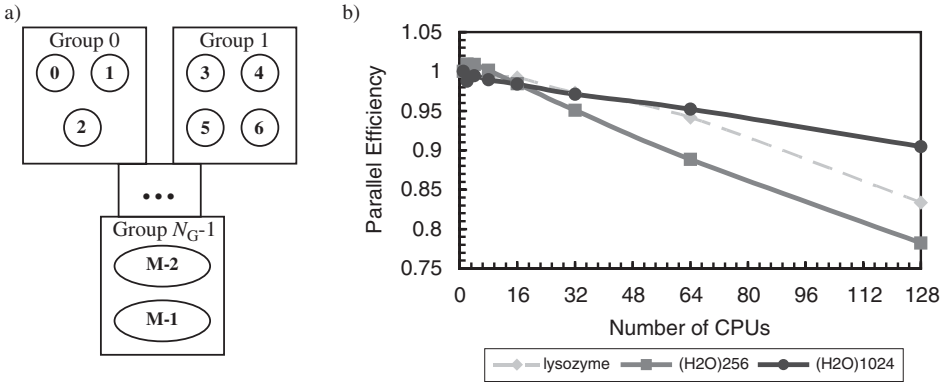


Fig. 1.10. (a) Division of M nodes (numbered from 0 to M-1) shown as ellipses into N_G GDDI groups (numbered 0 to N_G -1). (b) Parallel efficiency of GDDI as measured on a 128-node PentiumIII cluster connected by FastEthernet for water clusters $(H_2O)_{256}$, 6-31G* and $(H_2O)_{1024}$, STO-3G and lysozyme (2 036 atoms), STO-3G.

coefficient of roughly 2–8 for dimers and 4–100 for trimers). Linearly extended systems have a very small number of dimer and trimer calculations whereas spatially compact globular systems have a fairly larger number to calculate.

The optimal number of groups N_G in GDDI into which N_T subtasks (individual n-mer calculations) are directed is such that it is several times smaller than the number of subtasks.

$$N_G \leqslant N_T/\alpha,\ \alpha = 3 \sim 4 \tag{34}$$

With such a number of groups N_G the load balancing between groups minimizes the lag at the synchronization points (e.g., at the end of each monomer SCF iteration). The lag at the end of dimer and trimer calculations is usually smaller, and can use smaller α in the definition of N_G.

The computational efficiency is a balance of several factors. One is that the smaller the group, the better the parallelization scalability (larger N_G are better). The other is that the larger N_G, the more time is wasted at synchronization points (smaller N_G are better). The other additional factor is that if a group has more memory it can sometimes use a different algorithm, such as a fewer pass integral transformation for electron correlation or in-core two-electron integral storage in RHF (smaller N_G are better upto a certain limit). In most cases there is no lower required limit to the group size, except in applications making use of DDI shared memory, such as MP2. In the latter case, large matrices are distributed within groups and significant memory is required per group.

It should be noted that in practical applications on massively parallel computers there are frequently more CPUs than can be combined into N_G groups according to Eq. (34). In such a case, the number of groups in the monomer step has to be made smaller, e.g., $N_G = N$.

The most recent symmetric multiprocessing (SMP) implementation of DDI (SMP-DDI) [44] takes advantage of memory shared by CPUs within a node. Therefore, in practice it is nodes that are divided into groups and not CPUs (although calculations are of course performed on CPUs). The groups in GDDI need not have uniform size. In the majority of cases, it is appropriate to use nearly uniform division, dividing, for example, seven nodes into three groups with three, two and two nodes; however, some extreme cases with a few very large fragments require more sophisticated group usage [45].

Demonstration of the high GDDI scalability is depicted in Fig. 1.10(b), where the parallel efficiency is plotted against the number of CPUs (1.0 corresponds to M-fold speed-up on M CPUs). Depending on the system, efficiency of 0.77–1.01 was attained. Superlinear scaling can also be observed, further enhanced by recent in-core integral implementation, due to efficient cache and random-access memory usage, in addition to increasing the number of CPUs. In general, FMO code in GDDI does not require expensive network, and FastEthernet and Gigabit are sufficient for many applications (with the exception of methods making use of DDI-shared memory, such as MP2, in which case efficiency frequently drops to about 0.5 on such low-end networks).

The FMO modelling software FMOutil [46] automatically sets up group divisions based upon the number of available nodes, CPUs per node and the amount of physical memory. Since it is rather difficult to successfully predict the optimum division for a particular system, users are encouraged to gain experience by trying several divisions N_G for

some steps, especially if the same molecule has to be computed many times, such as during geometry optimization.

1.5 SCALING

It is well known that *ab initio* methods scale very steeply with system size N. RHF formally scales as N^4 (the two-electron integral step), however, scaling approaches N^3 for larger systems due to integral screening (Fock matrix diagonalization also scales as N^3). Scaling is even steeper when electron correlation is considered. MP2 is thought to scale as N^5, CCSD(T) as N^7 and full CI (FCI) combinatorially. There is ongoing work to reduce the scaling of *ab initio* methods, and most of the work is based on localized orbitals for correlated methods.

It should be noted that in the practical sense the issue of scaling is somewhat ambiguous. One usually considers the scaling of computationally most expensive terms and when system size is not large enough other terms can add on expenses leading to distortions from the expected dependence.

The actual scaling of the FMO method was investigated for the two-body expansions of FMO-MP2 [32], FMO-CC [35], FMO-MCSCF [31] and the three-body expansion of FMO3-CC [35]. In all cases, the scaling was found to be nearly linear, as determined by plotting calculation time as a function of system size. Below we investigate the issue of scaling from the theoretical viewpoint.

First, we consider the formal scaling of FMO-RHF. There are three major steps: computation of monomers, dimers and trimers. Let N denote the number of monomers. Each monomer calculation is accompanied by a computation of $N-1$ electrostatic potentials due to other monomers, thus the first step has a formal scaling of N^2. The dimer step consists of SCF dimers (for which SCF is performed) and ES dimers (for which Eq. (21) is used). The number of SCF dimers can be readily visualized by drawing a sphere of radius R_{ESDIM} (the threshold value) around each monomer. It is clear that the number of other monomers in this sphere does not depend upon N but depends upon the spatial distribution and the value of R_{ESDIM}. Each SCF dimer calculation is accompanied by $N-2$ monomer ESP calculations, thus the formal scaling of SCF dimers is N^2. The number of ES dimers is $N(N-1)/2 - N_{SCF}^{\mathrm{dim}}$, where N_{SCF}^{dim} is the number of SCF dimers. Each ES dimer calculation does not involve ESPs from other monomers, and there is no approximation to reduce their number, hence the formal scaling of the ES dimer step is N^2 as well.

Similarly, by drawing a sphere of radius R_{ITRIM} around each dimer it is obvious that the number of trimers for a given dimer is independent of N. Furthermore, the number of dimers themselves is again restricted by a threshold value and is thus independent of N. The number of trimers can be visualized by drawing a sphere around each monomer I of a radius that is practically chosen to be $2R_{\mathrm{ITRIM}}$, and including trimers containing I within the sphere that meet the threshold requirements. Therefore, the number of trimers is proportional to N, but with a much larger factor compared with that of dimers. Each trimer calculation is accompanied by $N-3$ monomer ESP computations, thus the scaling is formally N^2. Note that different from the dimer case, there are no ES trimer contributions

(they rigorously vanish, as the ES effect is exactly two-body in nature). The same will hold true for higher-body terms (provided that the thresholds are used).

Hence, we conclude that the formal scaling of FMO-RHF for any n-body expansion is N^2. However, the R_{ESPPC} approximation replaces two-electron integrals by one electron at all steps except the ES dimer step. The N^2 dependence remains in the $N-1$ calculations of one-electron integrals (for each of the N fragments). It should be noted that in the present state of development of FMO3 methods, we use no RESPAP and RESPPC approximations. However, two-electron integral screening reduces scaling from N^2 to nearly N during the calculation of the ESPs and the ES dimers.

Thus, the scaling of FMO-RHF is nearly linear. The term that slightly extends beyond linear scaling is the computation of one-electron integrals for the most part of the total system (within each n-mer SCF calculation). One-electron integrals receive much less attention in terms of their efficient calculation and some improvement can be made there. However, this extra term is not very significant and the FMO-RHF method is thus effectively linear in scaling.

Next, we consider correlated methods, such as FMO-MP2 or FMO-CC. The number of correlated dimers is restricted by the R_{CORSD} threshold (the correlated analogue of R_{ESDIM}) and is thus linear. If one leaves out FMO-RHF costs, then even the formal scaling of the two-body correlated FMO methods is linear. The number of trimers is restricted by R_{CORST}, which is the correlated analogue of R_{ITRIM}, and the formal scaling of the trimer step is linear, albeit with a much larger factor compared with that of dimers. The actual scaling of FMO2-MP2 is shown in Fig. 1.11.

Thus, the FMO methods are inherently of nearly linear scaling, irrespective of the n-body expansion and for both correlated and uncorrelated wave function. Practically speaking, however, the approximation thresholds are chosen conservatively and for very

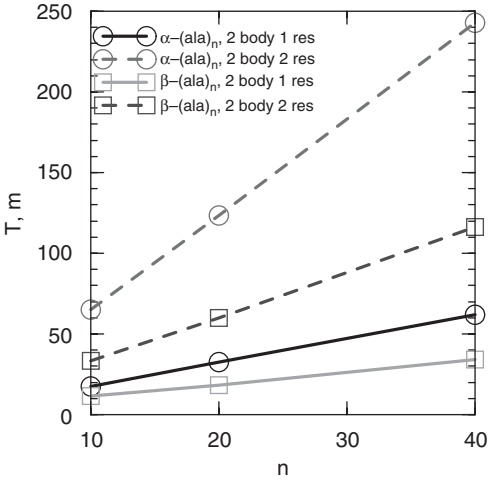


Fig. 1.11. The timings (in min) of FMO2-MP2 in the case of α and β conformers of $(\mathrm{ALA})_n$, using the 6-31G* basis set and one/two residues per fragment division. Nearly linear scaling is observed in all cases.

small systems the radii determining the number of n-mers (n = 2,3) are so large that nearly all possible n-mers are included, formally increasing the apparent scaling beyond linear.

Therefore, in order for linear scaling to be achieved, the system size expressed in terms of the van der Waals radii should be significantly (several times or more) larger than $(n-1)R_{\mathrm{ESDIM}}$ for the n-body FMO methods. Using the recommended value of $R_{\mathrm{ESDIM}} = 2.0$ and considering the values of the atomic van der Waals radii, the system size should extend beyond at least 5Å and 10Å in one direction, in order for linear scaling to be achieved in the two and three-body FMO methods, respectively.

Finally, a reasonable question to ask is: are the FMO calculations always faster than the corresponding *ab initio* ones? In order to attain high accuracy, elaborate calculations are performed in the FMO method and it is actually slower for very small problems. The system size, for which the FMO method becomes faster, depends on many factors, such as the basis set, wave function, fragment size, system compactness and the number of computer nodes. To give at least some idea for two very different cases, FMO2-RHF and FMO3-CCSD(T) were found to be faster than the respective *ab initio* RHF and CCSD(T) for molecules with 112 [$(ALA)_{10}$] and nine atoms [$(H_2O)_3$], respectively.

1.6 GUIDELINES TO APPLICATIONS

In the present stage of development, three major following ways to use the FMO method can be conceived:

1. optimizing molecular geometries,
2. computing single-point total energies, and
3. performing interaction analysis.

The first task is usually by far the most expensive. It has been our experience that it takes roughly N_{AT} steps to optimize a system with N_{AT} atoms. Even assuming linear scaling for a single-point calculation, the cost of performing geometry optimization is quadratic unless an unusually good initial guess was attained or the structure is very rigid. The basis set of good quality but perhaps a bit too expensive for practical optimizations is 6–31(+)G*, and we also use 3–21(+)G for this purpose. The FMO energy gradient is not strictly analytic, and there is a minor error of the order of $10^{-5}\ldots10^{-4}$ a.u./bohr with which gradient departs from being analytic. In practice, we have usually been able to converge geometries up to 1 000 atoms to the RMS gradient of 10^{-4} and for larger systems to slightly larger values, provided that we use a two residue per fragment division.

It can be expected, however, that if a solvent is present, the solute structures become more rigid and much fewer steps should be required to obtain optimization convergence. Next, in the case of large molecules, ultimately structures need not be optimized to a very high degree. Instead, one has to obtain a structure reasonably close to the global minimum (10^{-3} RMS convergence seems sufficient) and then perform dynamics simulations to sample the neighbourhood of the global minimum and determine the free energy.

Single-point calculations can be used to estimate reaction heats or activation barriers if the molecular geometries are known from experiment or other studies, e.g., QM/MM.

Another usage is to estimate relative stability of isomers, such as the native structure versus the α-helical one. In general, these two usages require different accuracy. The first one to estimate reaction heats usually handles two similar geometries, with perhaps significant changes to the reaction centre but minor structure relaxation of the remaining part. The other one, dealing with relative stabilities of isomers, often has to compare two very largely different structures. The amount of intramolecular interactions including dispersion in large systems is huge, and in order to obtain a meaningful difference of two huge numbers, high accuracy is needed in getting each factor.

Therefore, in most cases the recommended level of calculations is two-body FMO methods for chemical reactions and three-body FMO methods for isomer relative stability studies. In the latter case, it may suffice to apply the three-body expansion to obtain the uncorrelated (RHF) energetics, and two-body values for correlated corrections (MP2). In the case of two-body methods, two residues per fragment division should be used, whereas for three-body approaches it is possible to use one residue per fragment division. The use of a sufficiently accurate basis set is even more important for large molecules than for small ones and the smallest basis set recommended is 6–31(+)G*.

For the last major use, the interaction analysis, it is usually sufficient and more convenient to use one residue per fragment division and a two-body FMO method. Electron correlation does play an important role and in so doing FMO2-MP2 is a desired approach. The interaction analysis reveals the nature of interfragment interaction for 'fragments in molecule'. The quantitative information is obtained for interaction of each fragment with all others. This is in particular very important to:

1. doing a comparative study of mutants,
2. understanding the nature of repulsive and attractive interactions that have a direct affect, e.g., on catalytic activity of enzymes,
3. predicting mutations that can reduce the activation barriers and lead to large improvement of catalytic activity, and
4. analysing interaction with substrates and obtaining chemical insight into docking mechanisms.

A variety of basis sets has been used with the FMO method: STO-3G, 3–21G, 6–31G*, 6–311G*, cc-pVDZ, cc-pVTZ, cc-pVQZ and some others. While basis sets with diffuse functions (6–31++G**) have been used in several FMO studies [22,32], at the present level of development it is recommended that the diffuse functions be put only on negatively charged functional groups, e.g,. the carboxyl group COO^-, which we denote by adding (+), as in 6–31(+)G*. Anions have considerable spatial extension of electron density and their diffuse functions are appropriate and necessary. In the present state of development of the FMO method convergence problems and accuracy loss have been encountered when diffuse functions are put on all atoms, so such usage is in general discouraged.

With the exception of possibly employing 3–21G for structure optimization, low-quality basis sets should be used with great caution and avoided if possible, even in the multilayer FMO for the environment, especially if properties of the whole system are sought. That is, it may be possible to use small basis sets for the environment if one is interested in excited states of a single fragment or for chemical reactions. If one operates with the total

energy differences in other cases, sufficiently accurate basis sets should be used (6–31G* or better). As a quantitative warning, the relative stability of α-$(ALA)_{40}$ versus β-$(ALA)_{40}$ is −225.4 kcal/mol (Table 1.4) for *ab initio* RHF/6–31G*, but −370.3 kcal/mol for RHF/3–21G (from data in [22]).

One of the significant problems in studying large molecules is the modelling, that is, construction of molecular geometries. The experimental structures available from the PDB database have no hydrogen atoms, experimental resolution is frequently too low, and in many cases more work is necessary to complete the structure, e.g., add substrates. The absence of hydrogen atoms in crystallographic experimental structures presents a severe problem in deciding upon the protonation states of amino acid residues in proteins.

In the case of the FMO method, in addition to the above general problems, a minor difficulty to practical usage is the fragmentation and preparation of input data. To assist users, we have been developing the FMOutil software [46] that automatically converts PDB data into GAMESS input file format. Protonation, fragmentation and setting up parallel execution can be easily performed. Non-peptide components, if present, have to be fragmented manually (except water molecules). In addition to preparing input data, another set of utilities [46] is available to convert the results of calculations into common file formats so that post-processing of results is greatly simplified.

Only the main usages of the FMO method have been described above. Many novel applications are the subject of current research, such as calculation of molecular orbitals for the whole system and computing excited states. Visualizing the intramolecular electron density distribution and the electrostatic potential on surfaces surrounding the system can enhance the usefulness of the interaction analysis. Molecular dynamics applications of the FMO method [47–49] have been to some extent hampered by the lack of a strictly analytic gradient, however.

1.7 CONCLUSIONS

Summarizing the results, the FMO method brings to the chemical sciences what chemists always wanted: acknowledging the fact that molecules are composed of chemical functional units, such as residues, and such units and the interaction between them are computed on the sound basis of *ab initio* quantum chemistry.

Ab initio quantum chemical methods have been immensely successful, possessing the precious quality of being systematically improvable, so, given the system, one can select the required tools with which it can be handled. Two major problems are encountered, however: on enormous amount of computations and the lack of chemical insight.

Very often getting detailed chemical information beyond simple energetics is very difficult, available only to experts in the field, and in any case the quantities that can be extracted are of limited nature, mostly applying either to the whole molecule (such as molecular orbitals) or to atoms (e.g., atomic populations). What is missing is the detailed information about the quantitative amount of various types of interaction (polarization, charge transfer, etc.) between parts of the system.

As has been demonstrated for a large number of methods using common wave function types and basis sets, the FMO method is systematically improvable, partially because it is based on *ab initio* quantum chemistry so that better basis sets and wave function can be used, and partially due to its many-body expansion. The FMO properties are in close agreement with *ab initio* values, the computational efficiency is so high as to permitt 20000-atom calculations [45] with nearly linear scaling, and the interaction analysis information quantitatively substantiates and fosters chemical insight in practical applications.

The FMO method is not a panacea, and difficulties may be encountered in some cases where fragmentation is hard to perform due to the physical nature of the system. Much theoretical development has to be done to further enhance the usefulness of the method, such as working out fully analytic gradients, providing the means for free-energy calculations and including the solvent. Yet it is our hope that even in the present state of development, the FMO method should turn out to be of great use to real-life chemical problems.

1.8 ACKNOWLEDGEMENTS

This work was supported by a Grant-in-Aid for Scientific Research (JSPS, Japan) and NAREGI Nanoscience Project (MEXT, Japan), as well as CREST (JST, Japan).

1.9 REFERENCES

1 J. Gao and D.G. Truhlar, Annu. Rev. Phys. Chem., 53 (2002) 467.
2 F. Sato, T. Yoshihiro, M. Era and H. Kashiwagi, Chem. Phys. Lett., 341 (2001) 645.
3 S. Goedecker, Rev. Mod. Phys., 78 (1999) 997.
4 G. E. Scuseria, J. Phys.Chem. A, 103 (1999) 478.
5 P. Pulay, Chem. Phys. Lett., 100 (1983) 151.
6 M. Schütz and H.-J. Werner, J. Chem. Phys., 114 (2000) 661.
7 W. Yang, Phys. Rev. Lett., 66 (1991) 1438.
8 A. Imamura, Y. Aoki and K. Maekawa, J. Chem. Phys., 95 (1991) 5419.
9 G.P. Das, A.T. Yeates and D. S. Dudis, Int. J. Quant. Chem., 92 (2003) 22.
10 S.W. Benson, F.R. Cruickshank, D.M. Golden, G.R. Haugen, H.E. O'Neal, A.S. Rodger, R. Shaw and R. Walsh, Chem. Rev., 77 (1973) 1687.
11 V. Deev and M.A. Collins, J. Chem. Phys., 122 (2005) 154102.
12 M. Svensson, S. Humbel, R.J. Froese, T. Matsubara, S. Sieber and K. Morokuma, J. Phys. Chem., 100 (1996) 19357.
13 H. Stoll, Phys. Rev. B, 46 (1992) 6700.
14 D.W. Zhang and J.Z.H. Zhang, J. Chem. Phys., 119 (2003) 3599.
15 S. Li, W. Li and T. Fang, J. Am. Chem. Soc., 127 (2005) 7251.
16 K. Kitaura, E. Ikeo, T. Asada, T. Nakano and M. Uebayasi, Chem. Phys. Lett., 313 (1999) 701.
17 T. Nakano, T. Kaminuma, T. Sato, Y. Akiyama, M. Uebayasi and K. Kitaura, Chem. Phys. Lett., 318 (2000) 614.
18 S. Hirata, M. Valiev, M. Dupuis, S.S. Xanthneas, S. Sugiki and H. Sekino, Mol. Phys., 103 (2005) 2255.
19 M. Dupuis, M. Aida, Y. Kawashima and K. Hirao, J. Chem. Phys., 117 (2002) 1242.
20 T. Vreven, K. Morokuma, O. Farkas, H. B. Schlegel and M. J. Frisch, J. Comput. Chem., 24 (2003) 760.
21 D.G. Fedorov, T. Ishida and K. Kitaura, J. Phys. Chem. A, 109 (2005) 2638.
22 D.G. Fedorov and K. Kitaura, J. Chem. Phys., 120 (2004) 6832.

23 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. Su, T.L. Windus, M. Dupuis and J.A. Montgomery, J. Comput. Chem., 14 (1993) 1347.
24 GAMESS software can be downloaded for free from: www.msg.ameslab.gov.
25 K. Kitaura, S.-I. Sugiki, T. Nakano, Y. Komeiji and M. Uebayasi, Chem. Phys. Lett., 336 (2001) 163.
26 H. Sekino, Y. Sengoku, S.-I. Sugiki and N. Kurita, Chem. Phys. Lett., 378 (2003) 589.
27 Y. Inadomi, T. Nakano, K. Kitaura and U. Nagashima, Chem. Phys. Lett., 364 (2002) 139.
28 K. Kitaura and K. Morokuma, Int. J. Quant. Chem., 10 (1976) 325.
29 S.-I. Sugiki, N. Kurita, Y. Sengoku and H. Sekino, Chem. Phys. Lett., 382 (2003) 611.
30 D.G. Fedorov and K. Kitaura, Chem. Phys. Lett., 389 (2004) 129.
31 D.G. Fedorov, K. Kitaura and J. Chem. Phys., 122 (2005) 054108.
32 D.G. Fedorov, K. Kitaura and J. Chem. Phys., 121 (2004) 2483.
33 Y. Mochizuki, S. Koikegami, T. Nakano, S. Amari and K. Kitaura, Chem. Phys. Lett., 396 (2004) 473.
34 Y. Mochizuki, T. Nakano, S. Koikegami, S. Tanimori, Y. Abe, U. Nagashima and K. Kitaura, Theor. Chem. Acc., 112 (2004) 442.
35 D. G. Fedorov and K. Kitaura, J. Chem. Phys., 123 (2005) 134103.
36 Y. Mochizuki, S. Koikegami, S. Amari, K. Segawa,. K. Kitaura and T. Nakano, Chem. Phys. Lett., 406 (2005) 283.
37 T. Nakano, T. Kaminuma, T. Sato, K. Fukuzawa, Y. Akiyama, M. Uebayasi and K. Kitaura, Chem. Phys. Lett., 351 (2002) 475.
38 K. Fukuzawa, K. Kitaura, M. Uebayasi, K. Nakata, T. Kaminuma and T. Nakano, J. Comp. Chem., 26 (2005) 1.
39 Y. Mochizuki, K. Fukuzawa and A. Kato, S. Tanaka, K. Kitaura and T. Nakano, Chem. Phys. Lett., 410 (2005) 247.
40 T. Nemoto, D.G. Fedorov, M. Uebayasi, K. Kanazawa, Y. Komeiji and K. Kitaura, Comput. Biol. Chem., 29 (2005) 434.
41 S.-I. Sugiki, M. Matsuoka, R. Usuki, Y. Sengoku, N. Kurita, H. Sekino and S. Tanaka, J. Theor. Comp. Chem., 4 (2005) 183.
42 T. Ishida, D.G. Fedorov and K. Kitaura, J. Phys. Chem. B, 110 (2006) 1457.
43 D.G. Fedorov and R.M. Olson, K. Kitaura, M.S. Gordon and S. Koseki, J. Comp. Chem., 25 (2004) 872.
44 R.M.Olson, M.W. Schmidt, M.S. Gordon and A.P. Rendell, Proc. of Supercomputing, IEEE Computer Society, Phoenix, 2003.
45 T. Ikegami, T. Ishida, D.G. Fedorov, and K. Kitaura, Y. Inadomi, H. Umeda, M. Yokokawa and S. Sekiguchi, Proc. of Supercomputing, IEEE Computer Society, Seattle, 2005.
46 FMOutil is included in the distributed version of GAMESS and the latest version can be downloaded from:http://staff.aist.go.jp/d.g.fedorov.
47 T. Ishimoto, H. Tokiwa, H. Teramae and U. Nagashima, Chem. Phys. Lett., 387 (2004) 460.
48 Y. Komeiji, T, Nakano, K. Fukuzawa, Y. Ueno, Y. Inadomi, T. Nemoto, M. Uebayasi, D.G. Fedorov and K. Kitaura, Chem. Phys. Lett., 372 (2003) 342.
49 T. Ishimoto, H. Tokiwa, H. Teramae, U. Nagashima, J. Chem. Phys., 122 (2005) 094905.

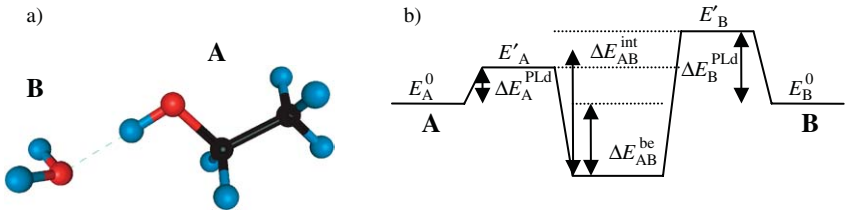


Plate 1. Interactions in ethanol-water complex. Free molecules A and B are mutually polarized by the amount of $\Delta E_A^{\mathrm{PLd}}$ and $\Delta E_B^{\mathrm{PLd}}$ and the amount of pair interaction energy is shown as $\Delta E_{AB}^{\mathrm{int}}$. The binding energy $\Delta E_{AB}^{\mathrm{be}}$ is measured against the free subsystem energies $E_A^0 + E_B^0$ (illustration appears on page 21 of this volume).

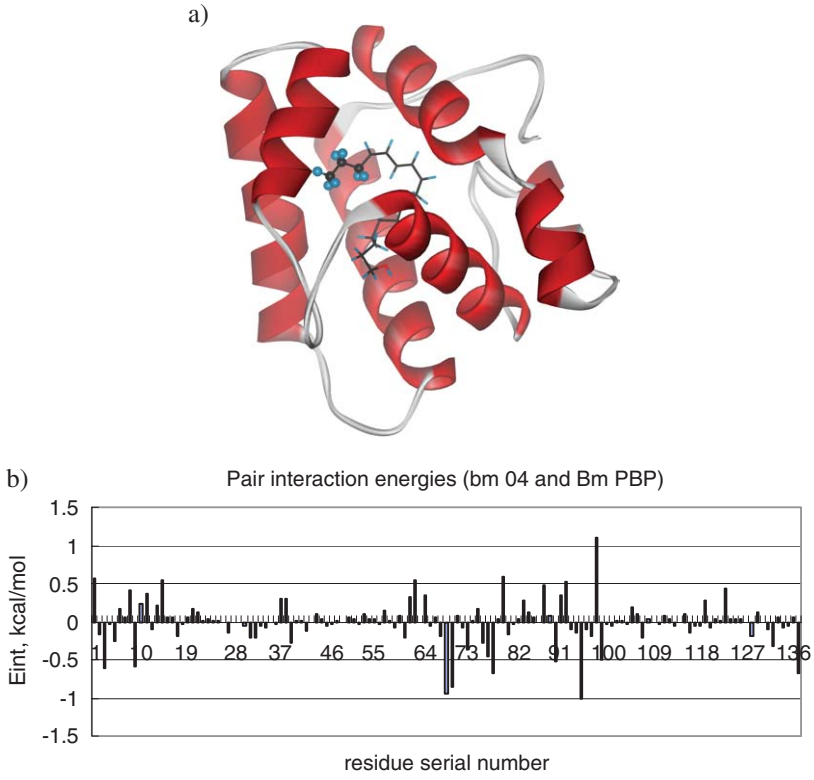


Plate 2. Pair Interaction analysis in BmPBP: (a) the structure of the BmPBP complex, (b) pair interaction of a part of bombykol ligand (denoted by bm04) with each residue in BmPBP. The protein and ligand are treated on the same footing. The protein in shown schematically by its secondary structure units, the atoms of the ligand are shown as sticks, and, in addition, a part of the ligand for which the interaction analysis is given, is shown with balls and sticks (illustration appears on page 23 of this volume).

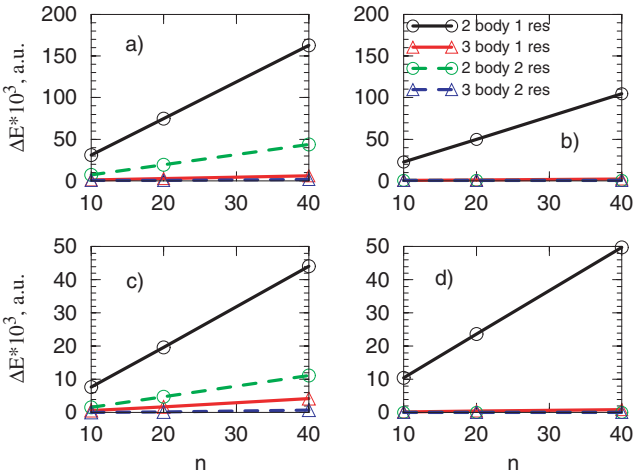


Plate 3. The error in energy (milliHartree) for FMO-DFT: (a) α-(ALA)n, (b) β-(ALA)n and for FMO-RHF: (c) α-(ALA)n, (d) β-(ALA)n, measured against the corresponding *ab initio* values. The basis set is 6–31G* throughout (illustration appears on page 25 of this volume).

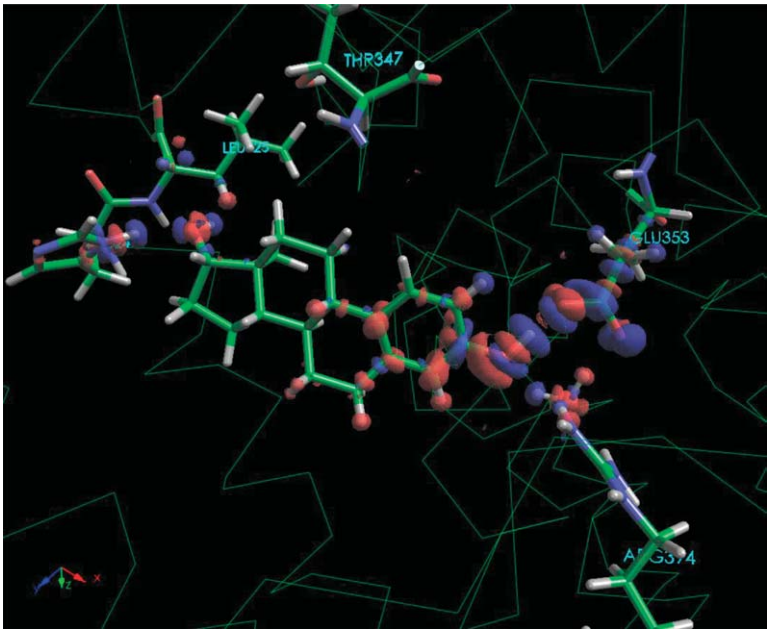


Plate 4. The difference of electron densities ($\Delta\rho$) between complexed and individual component molecules; $\Delta\rho = \rho_{complex} - (\rho_{ER} + \rho_{ligand})$. The stick representation refers to EST(1), a water molecule, and residues (Glu353, Thr347, Arg394, His524, and Leu525) with significant $\Delta\rho$ values. The thin line represents the $C\alpha$ backbone structure of EPα LBD except for the above residues. The blue (minus) and red (plus) represent the sign of the isosurface at ±0.002e/Bohr^3 drawn by BioStation Viewer [17] (illustration appears on page 45 of this volume).

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 2

Developments and applications of ABINIT-MP software based on the fragment molecular orbital method

Tatsuya Nakano[1,4,6], Yuji Mochizuki[2,4,6,7], Kaori Fukuzawa[3,6], Shinji Amari[4] and Shigenori Tanaka[5,6]

[1]*Division of Safety Information on Drug, Food and Chemicals, National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan*
[2]*Advancesoft and Institute of Industrial Science, Center for Collaborative Research, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan*
[3]*Division of Science and Technology, Mizuho Information & Research Institute, Inc., 2-3 Kanda Nishiki-cho, Chiyoda-ku, Tokyo 101-8443, Japan*
[4]*Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan*
[5]*Graduate School of Science and Technology, Kobe University, 1-1, Rokkodai, Nada-ku, Kobe 657-8501, Japan*
[6]*CREST, Japan Science and Technology Agency (JST), Japan*
[7]*Department of Chemistry, Faculty of Science, Rikkyo University, 3-34-1 Nishi-ikebukuro, Toshima-ku, Tokyo 171-8501, Japan*

Abstract

In this chapter we describe the development of ABINIT-MP code by which the fragment molecular orbital (FMO) calculations of HF, MP2, and CIS levels of theory are available in an efficiently parallelized fashion. We also present a variety of practical applications such as the virtual screening for estrogen receptor (ER) and the evaluation of excitation energy for photoactive yellow protein (PYP).

2.1 INTRODUCTION

The application of quantum chemical methods to large molecules such as proteins and nucleic acids remains a great challenge in theoretical chemistry. The calculations of electronic states for biomolecules are usually formidable due to their huge size, and some ingenious techniques are required to make them feasible. While an idea of the fragmentation of the whole system is straightforward and appealing for the reduction in computational cost, the difficulties always take place concerning the inaccuracies due to the

artificial breakage of chemical bonds and the reconstruction of observable properties. Among a number of these approaches, however, the fragment molecular orbital (FMO) method proposed by Kitaura *et al.* [1–5] affords a very promising way to retain the chemical accuracies in energy and other properties in the calculations for huge biopolymers.

The FMO method [1–5] relies on the division of a large molecular system into a collection of small fragments and on the molecular orbital (MO) calculations for the fragments (monomers) and their pairs (dimers) performed to obtain the total energy and other molecular properties. The total electronic energy of the molecules can then be calculated as

$$E = \sum_I E_I + \sum_{I>J} (E_{IJ} - E_I - E_J) \tag{1}$$

where E_I and E_{IJ} represent the energies of fragment monomer and dimer, respectively, calculated under the electrostatic potentials from other surrounding fragments (environmental electrostatic potentials). The FMO method thus avoids the MO calculation of the whole system of large molecules and reduces the computational time remarkably. Another advantage of this method is its ease in utilizing parallel processing, since the fragments and the fragment pairs can be treated independently.

A number of test calculations have been already performed [1–5] concerning the accuracy and efficiency of the FMO method, providing extremely promising results for application to biomolecules. Actual implementations of the FMO method in terms of software have also been carried out; examples include GAMESS [6–9], NWChem [10,11], and ABINIT-MP [12–17]. In this article, we illustrate recent developments of the ABINIT-MP software and some examples of its application to the MO calculations for proteins. The FMO calculations in the Hartree–Fock (HF), the second-order Møller–Plesset perturbation (MP2), and the configuration interaction with singles (CIS) approximations are now available in an efficiently parallelized fashion. The application examples are demonstrated for the virtual screening of ligand molecules for estrogen receptor (ER) and the evaluation of excitation energy of photoactive yellow protein (PYP).

2.2 ABINIT-MP SOFTWARE

The ABINIT-MP software [17] is a program code freely available on the web to perform the FMO calculations for biopolymers such as proteins and nucleic acids. The most time-consuming parts in the method, the calculations of environmental electrostatic potentials, are accelerated by employing the Mulliken approximation for two-electron integrals and a fractional point charge approximation without loss of accuracy [5]. The errors arising from the fragment approximation are thus suppressed within a few kcal/mol even for the medium-size proteins. A flexible parallelization is made possible by combining the fragment indices (upper level) and the two-electron integral indices (lower level) on distributed computational resources, leading to an enhancement of in-core processing. The MP2 level electron-correlation calculations for proteins can then be performed very efficiently by the ABINIT-MP code [12,13]. As for the excited states, a parallelized integral-direct solver for the CIS calculation has been developed in combination with the multi-layer FMO method [14], in which a high-level calculation could be performed for active sites. The functions of the software are successively extended so that more extensive basis sets

are available and the engines for the structural optimization and molecular dynamics [18] are introduced. The output files of the calculations can also be employed for the analysis of the molecular characteristics in terms of the inter-fragment interaction energies (IFIEs) and their configurations [16]. All the pre- and post-processors for the ABINIT-MP code can be integrated on a specialized graphical user interface called BioStation Viewer [17].

2.3 AN APPLICATION OF FMO-HF METHOD: PREDICTION OF LIGAND BINDING AFFINITIES OF ESTROGEN RECEPTOR α

The steroid hormone estrogens play important roles in the regulation of growth, differentiation, and homeostasis in a variety of tissues. These effects are induced by the binding of estrogens to the estrogen receptors (ER), which are members of the nuclear receptor (NR) superfamily and work as a transcription factor [19,20]. Because a variety of unknown compounds could bind to ER ligand-binding domains (LBD) and exert hormone-like effects on humans and wildlife, there is increasing interest in the *in silico* ligand screening of ER as a theme for drug discovery and the problem of endocrine disruptors [21]. The ligand-binding site of α-subtype of ER is constituted by the ligand and the surrounding polar and charged amino acid residues (Fig. 2.1). The hydrogen bonds between them can be thought of as stabilizing the ER-ligand binding [22,23]. Here, we show the FMO-HF results for the molecular interaction between the ERα and ligands to elucidate their binding mechanism [24,25].

The ligand molecules examined here are displayed in Fig. 2.2. We use three models for the ER-ligand complexes (Fig. 2.3). Along with the ligand and a water molecule, MODEL 1 contains the entire LBD which consists of 241 amino acid residues, MODEL 2 contains the first-layered α-helices of the ligand (50 residues), and MODEL 3 contains only four residues of Glu353, Arg394, His524 and Leu387. The large-scale calculations of MODEL 1 were performed only for four ligands, and most of the calculations were carried out using MODEL 2. MODEL 3 was used for geometry optimizations of the hydrogen bond networks on ER-ligand binding at the HF/6-31G(d) level using the Gaussian 98 program

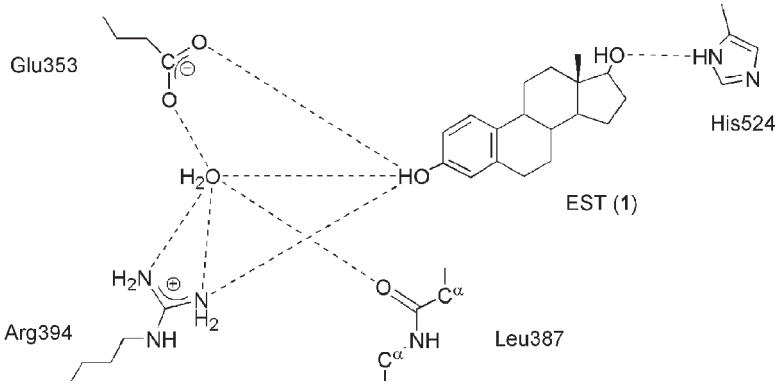


Fig. 2.1. Hydrogen bond network at the ligand binding site of ER complexed with 17β-estradiol, EST (1) [24]. Dotted lines indicate hydrogen bonds.

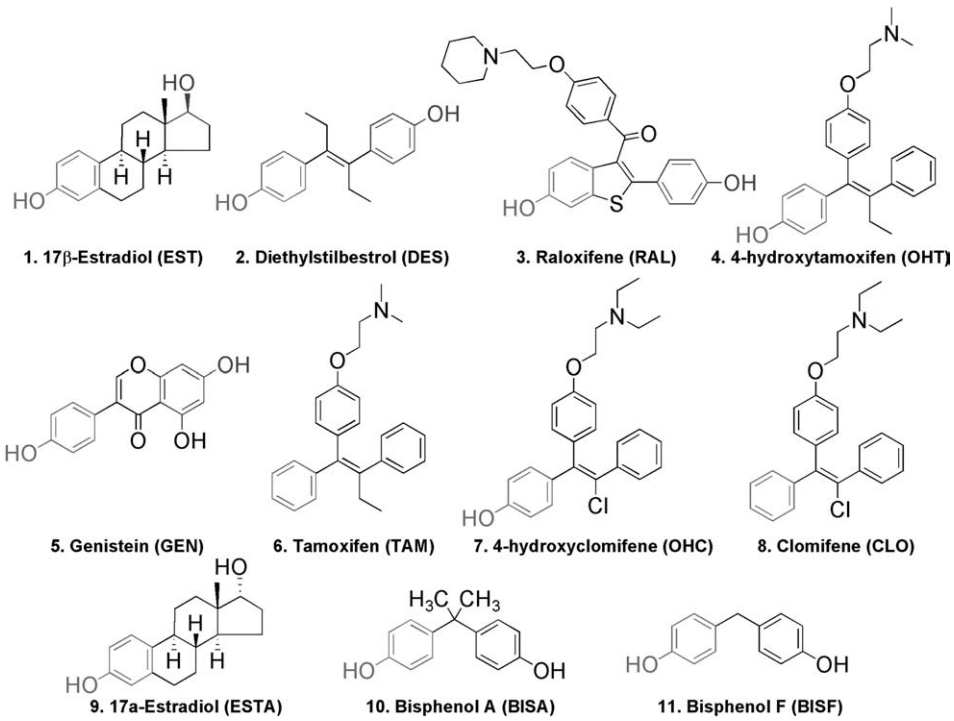


Fig. 2.2. Estrogen-like compounds used in ER-ligand binding calculations [24].

package [26]. The single-point energy calculations were carried out on MODEL 1 and MODEL 2 at the FMO-HF/STO-3G level using the ABINIT-MP program [17]. The fragmentations were done by two amino acid residues for the ER, and the ligand and the water molecule were treated as a single fragment. The computation times on 16 Dual Xeon 3.06GHz clusters (32 CPUs) were 2.2 hours for MODEL 1. For comparison, CHARMM [27] calculations were also carried out for ER-ligand binding energies.

The energy of each of the three systems, i.e., the receptor, E_{receptor}, the ligand, E_{ligand}, and the ER-ligand complex, E_{complex}, were calculated. The binding energy for a given ligand, ΔE_{ligand}, can be obtained as a difference in the energies of the complex and of its components as follows,

$$\Delta E_{\text{ligand}} = E_{\text{complex}} - (E_{\text{receptor}} + E_{\text{ligand}}) \tag{2}$$

and the binding energies of a ligand relative to that of 17β-estradiol (EST), $\Delta\Delta E_{\text{ligand}}$, are defined as

$$\Delta\Delta E_{\text{ligand}} = \Delta E_{\text{ligand}} - \Delta E_{\text{EST}} \tag{3}$$

Then, the correlation between $\Delta\Delta E_{\text{ligand}}$ and the experimental relative binding affinity (RBA) [28] was examined; RBA of each ligand is calculated as the ratio of concentrations

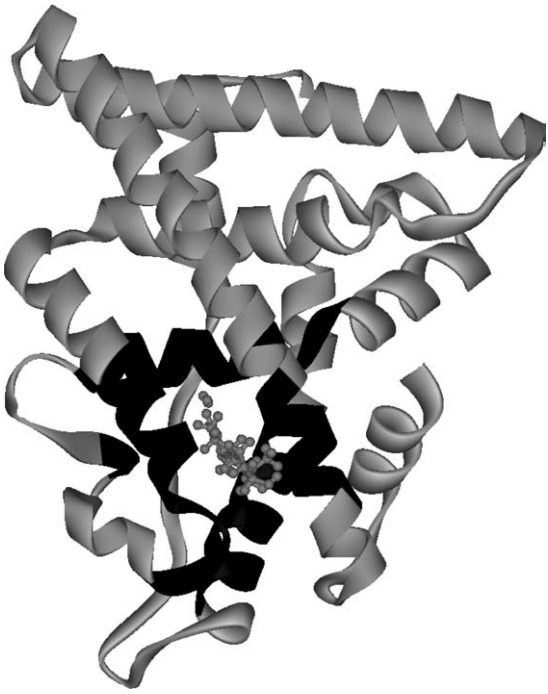


Fig. 2.3. The ERα LBD complexed with EST (1) [25]. MODEL 1 includes the whole complex (241 residues), and in the inside, residues belonging to MODEL 2 (50 residues) are displayed as a dark color surrounding the ligand and a water molecule (ball and stick model).

of EST or competitor required to reduce the specific radioligand binding by 50 percent (= ratio of IC50 values). In the CHARMM calculations for MODEL 1, the relationship between calculated relative binding energies ($\Delta\Delta E$) versus experimental RBA of eight ligands was examined, and no correlation was found between $\Delta\Delta E$ and log (RBA/100); the correlation coefficient r was 0.035. These calculations suggest that the intermolecular interactions between ER and ligands are poorly estimated by the classical force field. The FMO results for MODEL 2 are shown in Fig. 2.4. For the eight compounds, 1–6, 9 and 10, whose RBA values are known experimentally, the correlation between $\Delta\Delta E$ and log (RBA/100) was good ($r = 0.837$). Using the correlation equation obtained by the least-squares fitting procedure, the log (RBA/100) values of ligands 7, 8 and 11, whose RBA are unknown, can be predicted from the calculated $\Delta\Delta E$ values.

Using MODEL 1, the ΔE and $\Delta\Delta E$ values are also calculated for the complexes of ER with the four ligands (1–4), which are shown in Table 2.1. The differences of ΔE between MODEL 1 and MODEL 2 are less than 3 kcal/mol except for the RAL. Thus, MODEL 2 with the 50 residues is a reliable model for describing the interaction between ER and the ligands, but for a large ligand such as RAL, additional residues which make contact with the ligand should be included.

Another interesting finding is a difference in the charge distribution between complexed and individual component molecules. The total net charges of ligands are changed negatively

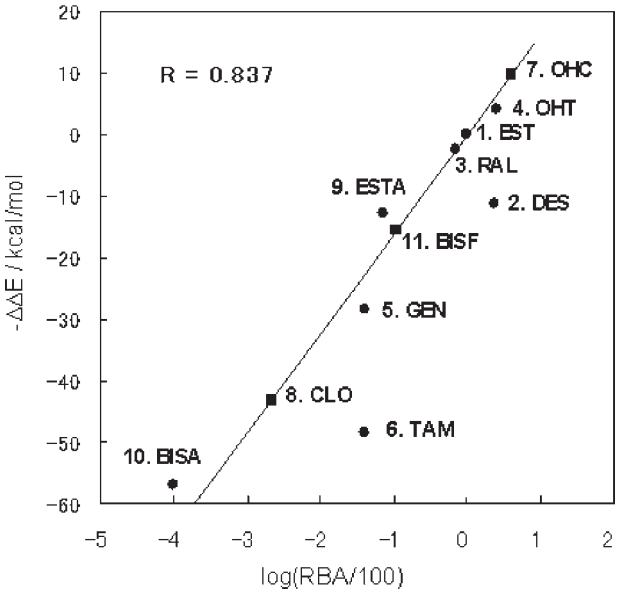


Fig. 2.4. Relationship between calculated relative binding energies ($\Delta\Delta E$) versus experimental RBA of eight ligands (circles), and the estimation of experimental values for three ligands (squares) [24].

Table 2.1 Experimental RBA, calculated binding energies (ΔE), and relative binding energies ($\Delta\Delta E$) of several ligands with ER. Energies are in kcal/mol [24]

		MODEL 2 (50 residues)		MODEL 1 (241 residues)	
Ligand	log(RBA/100)	ΔE (kcal/mol)	$\Delta\Delta E$ (kcal/mol)	ΔE (kcal/mol)	$\Delta\Delta E$ (kcal/mol)
EST (**1**)	0.00	–37.80	0.00	–37.65	0.00
DES (**2**)	0.37	–26.70	11.10	–28.33	9.32
RAL (**3**)	–0.16	–35.30	2.50	–26.13	11.52
OHT (**4**)	0.41	–41.73	–3.92	–38.19	–0.54
GEN (**5**)	–1.40	–9.42	28.38	—	—
TAM (**6**)	–1.40	10.62	48.42	—	—
OHC (**7**)	—	–47.62	–9.82	—	—
CLO (**8**)	—	5.29	43.10	—	—
ESTA (**9**)	–1.15	–25.07	12.73	—	—
BISA (**10**)	–4.00	19.22	57.02	—	—
BISF (**11**)	—	–22.23	15.57	—	—

by the values –0.001 ~ –0.181e when complexed with ER. The same order of positive charges is induced on Glu353 by +0.006 ~ +0.198e, and negative charges are slightly induced on Arg394 and His 524 by 0 ~ –0.022e. Considerable electrons are supplied from Glu353 (an electron donor) to the ligands (an electron acceptor) with the exception of the non-hydrogen bonding ones. The quantity of electrons transferred from ER to the ligands

is highly correlated with the binding energy; ΔE becomes larger or the complex becomes more stable with the increase in the negative charge of the ligand. A change of electron densities between complexed and individual component molecules is also considered for ER-EST(1) system (Fig. 2.5). The remarkable change of electron densities occurs on Glu353-EST interaction site, and some of them occur on His524-EST and Arg394-EST interaction sites. Thus, the stabilization in the ER-ligand binding mostly arises from the ligand-Glu353 interaction and slightly from ligand-His524 and ligand-Arg394 interactions. This fact suggests that the hydrogen bond between ER and the ligand plays an important role in characterizing the charge transfer and the concerted stabilization of the ER-ligand complex, which could be accounted for only through the quantum mechanical calculations.

The FMO-HF method presented herein may provide a powerful tool for assessing the affinity of putative xenoestrogens *in silico* prior to biological experiments. However, there would be a considerable contribution of many other effects to binding affinity, which were ignored in the above calculations; for example, a solvent effect, solvation free energy, hydrophobic interaction, induced fitting and dispersion energy. The calculated results

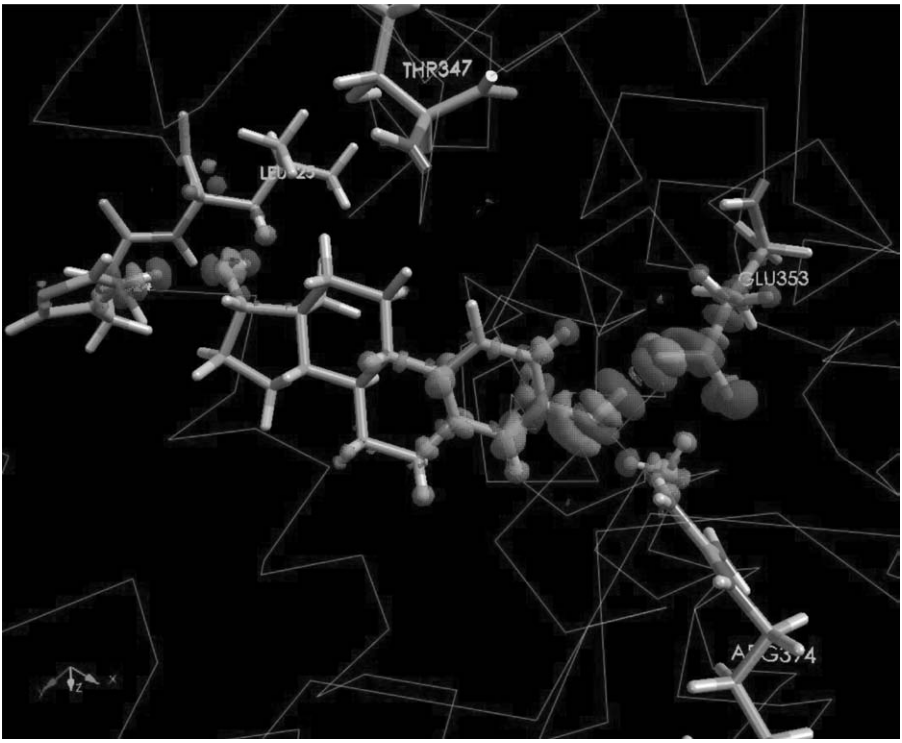


Fig. 2.5. The difference of electron densities ($\Delta\rho$) between complexed and individual component molecules; $\Delta\rho = \rho_{complex} - (\rho_{ER} + \rho_{ligand})$. The stick representation refers to EST(1), a water molecule, and residues (Glu353, Thr347, Arg394, His524, and Leu525) with significant $\Delta\rho$ values. The thin line represents the C_α backbone structure of ERα LBD except for the above residues. The blue (minus) and red (plus) represent the sign of the isosurface at ±0.002e/Bohr^3 (see Color Plate 4).

nevertheless obtained a good correlation with experimental results, and therefore the relative binding affinity is thought to be correlated with the enthalpic relative binding energies, and other effects for each ligand could be assumed to be similar to the first approximation. We will consider these effects and include electron correlation in future studies.

2.4 IMPLEMENTATION OF THE MP2 METHOD

The hydrogen-bonding interaction plays the key role in the bindings of biochemical molecules, besides the electrostatic interaction [29]. The stabilization of hydrogen-bonding would be increased if the electron correlations were to be incorporated by the use of post-HF calculations. The interaction of dispersion or van der Waals type is another important factor, where it must be evaluated at the post-HF level of theory. Among the post-HF calculations, the second-order Møller–Plesset perturbation (MP2) is the simplest and the computationally most economical treatment to include the correlation size-extensively. Although the density functional theory (DFT) calculations have become the *de facto* standard of the correlated method, the MP2 has an advantage over the DFT in treating both the hydrogen-bonding and dispersion appropriately.

The MP2 correlation energy is given by the contraction of MO integral list of (*ia*,*jb*) (Mulliken notation with *i*,*j* for doubly occupied MOs and *a*,*b* for virtual MOs) with the orbital energies of ε_i, ε_j, ε_a and ε_b [30]:

$$E_{\mathrm{MP2}} = \sum_{ijab} (ia,jb)[2(ia,jb)-(ib,ja)]/[\varepsilon_i+\varepsilon_j-\varepsilon_a-\varepsilon_b] \tag{4}$$

The (*ia*,*jb*) list is generated by the series of quarter transformations from the AO integral list, (*pq*,*rs*). We have developed an efficient code to calculate the MP2 energy in a parallelized integral-direct fashion in the ABINIT-MP system [12,13]. This code has the following features:

- The AO index of *s* is the parallelization parameter.
- No communication across the worker processes is needed.
- Multiple screening is incorporated to reduce the operation costs.
- The BLAS library is used in the innermost processing.

The introduction of correlation should induce the reorganization of electron distribution obtained by the HF calculations. We thus implemented the ability to calculate the MP2 density matrix, where the response elements, to describe the reorganization, are approximately given by

$$P_{ai} = [\sum_{jbc} A_{ij}^{bc} (jc,ab) - \sum_{jkb} A_{jk}^{ab} (jb,ik)]/ [\varepsilon_i-\varepsilon_a] \tag{5}$$

$$A_{ij}^{ab} = 2 [2(ia,jb)-(ib,ja)]/[\varepsilon_i+\varepsilon_j-\varepsilon_a-\varepsilon_b] \tag{6}$$

The additional lists of (*ia*,*bc*) and (*ia*,*jk*) should be prepared as the (*ia*,*jb*) list. The other density elements of P_{ab} and P_{ij} are similarly constructed by integral contractions. The MP2 calculation for each fragment can be independently performed as the HF

calculation. Namely, the FMO-MP2 parallelism has a dual layer of fragment indices (upper level) and the AO indices (lower level) on the distributed computational resources. The total MP2 correlation energy and density are also given by the standard formula of the two-body FMO scheme.

We here introduced a couple of benchmark FMO-MP2 calculations with the 6-31G basis set on 32 Dual Xeon (3.06 GHz and 2 GB memory) machines [13]. The 1 024 water cluster was used as the first test case, where the MP2 density was also computed. The hydrogen-bonding energies per water molecule were calculated to be −19.08 kcal/mol by HF and −23.09 kcal/mol by MP2. In the three-dimensional network of water cluster, the single water molecule has four hydrogen-bondings on average. The MP2 energy per bond (or the quarter value) was in agreement with the typical hydrogen-bonding energy range of 6–7 kcal/mol via the post-HF methods. The average values of Mulliken net charge for the oxygen atom were −0.91 by HF and −0.83 by MP2, respectively. This indicated that an excess ionicity described at the HF level is corrected by the MP2 correlation. The turnaround time of the MP2 job with 64 processors was only 343 seconds.

The second FMO-MP2 benchmark was concerned with the drug design for HIV-1. Namely, the complex between HIV-1 protease and an inhibitor molecule, lopinavir, was employed (see Fig. 2.6). The numbers of amino acid residues, atoms and basis functions were 198, 3 225 and 17 423, respectively. The single residue was treated as the fragment monomer. The binding energy of lopinavir to the HIV-1 protease was evaluated to be −61 kcal/mol by HF and −135 kcal/mol by MP2. This MP2 energy enlargement could be attributed to the

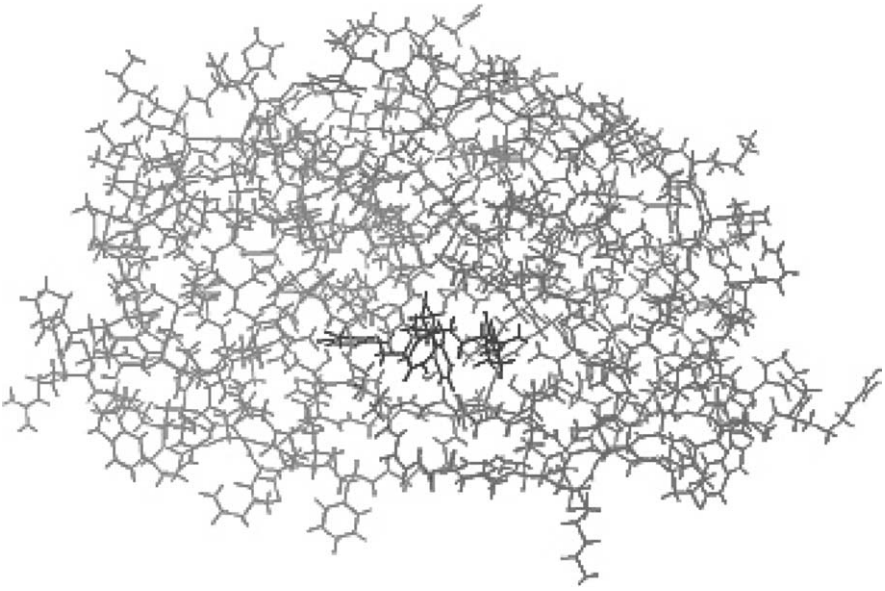


Fig. 2.6. Graphic representation of HIV-1 protease and lopinavir. Two equivalent subunits are colored in red and green, and lopinavir is indicated in blue. Original color figure can be found in the web version of Ref. [13] (see Color Plate 5).

enhanced hydrogen-bonding energy and the introduction of dispersion energy. The turnaround time was only 14.3 hours, demonstrating the practical applicability of FMO-MP2 calculations to a variety of realistic proteins. Such actual applications are described elsewhere.

2.5 CIS CALCULATIONS

A number of biochemical systems, e.g., rhodopsin in vertebrate eyes, are photoactive through electronic excitations. The central region in photoactive proteins is called the chromophore, and it consists of the photon-absorbing pigment part and some neighboring residues which should provide the electrostatic or hydrogen-bonding interactions with the pigment. Surrounding residues also put some electrostatic potentials on the chromophore system. These environmental effects cause the red shifts or blue shifts in excitation energies. For example, the opsin shift is an archetype. To treat such environmental shifts theoretically, a sort of hybrid approach of quantum mechanics (QM) and classical molecular mechanics (MM) has been frequently used. Namely, QM or molecular orbital (MO) calculations are used only for the photoactive molecule and neighboring area when needed. Surrounding parts are treated in MM ways. Although the MM technique makes the inclusion of environmental effects feasible, there are many prefixed empirical parameters. Especially, the MM description for the hydrogen-bonding, which plays the crucial role for water molecules and amino acids [29], has not yet been well enough established. This leads to an ambiguity in QM/MM calculations. A full treatment of QM would thus be desirable if possible. Recently, Kitaura *et al.* [31,32] proposed an extension of FMO toward the multilayer treatment (MLFMO), based on the GAMESS program. In Ref. [32], the multiconfiguration self-consistent field (MCSCF) was used in the 'chemically important region' (layer 2), by keeping the surrounding part (layer 1) at the HF level of theory.

The most low-lying of excited states in molecules could be semi-quantitatively described by the calculations of configuration interaction with singles (CIS), where a typical overestimation in excitation energies is 1–2 eV relative to the experimental values [33]. The overestimation in CIS energies can be corrected by introducing the correlation and relaxation via the CIS(D) approach developed by Head-Gordon *et al.* [34]. For the excited states, the time-dependent DFT (TD-DFT) calculations nowadays have become a standard quantitative method. However, the TD-DFT description could collapse for the charge-transfer (CT) excitations with erratically low energies depending on functional [35]. This situation should cause a difficulty in treating the chromophore consisting of the pigment and neighboring residues, since the spurious residue–pigment CT states could appear easily. The way of CIS and CIS(D) should thus be safer or better than that of TD-DFT for the photoactive proteins.

Recently, we have developed an integral-direct CIS code in the ABINIT-MP system [14], by utilizing the MLFMO technique [31,32]. The chromophore is just the CIS region as layer 2. The entire system consisting of layers 1 and 2 is first calculated by the usual FMO-HF procedure. Then, the CIS calculation is performed for layer 2 (L2), where the MLFMO-CIS energy would be given in an additive manner,

$$E(\text{MLFMO-CIS}) = E(\text{FMO-HF}) + \Delta E^{\text{L2}}(\text{CIS}) \tag{7}$$

Here the second term on the right-hand side of Eq. (7) is the excitation energy by CIS. In our CIS implementation [14], the so-called sigma vector set for iterative diagonalization of the Hamiltonian matrix is directly evaluated from the (*pq*,*rs*) list. The computational cost of integral processing scales as N^4 (N is the number of AOs) like the HF calculation. The parallelization is straightforward.

The benchmark test of MLFMO-CIS was performed on the photoactive yellow protein (PYP) which is a light sensor in halophilic bacteria [14]. The condition of benchmark was the same as the FMO-MP2 case [13]. The PYP contains a total of 125 amino acid residues (see Fig. 2.7). The central pigment is the *p*-coumaric acid linked by a thioester bond with Cys69. In the CIS calculation, the surrounding residues of Glu46, Tyr42 and Arg52 were also included into layer 2. The CIS dimension was 47 265, and its lowest excitation energy was calculated to be 4.28 eV. As shown in Fig. 2.8, this excitation was of typical HOMO-LUMO type. The calculated energy was considerably higher than the

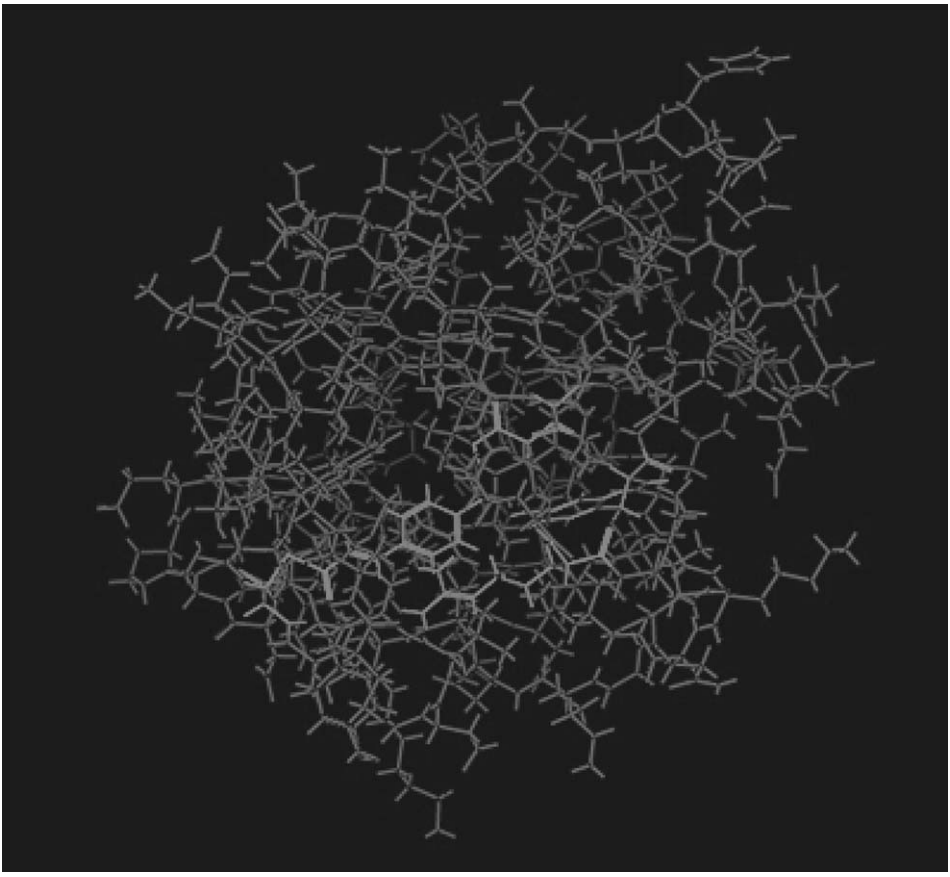


Fig. 2.7. Graphic representation of PYP. The CIS region is colored in red. Original color figure can be found in the web version of Ref. [14] (see Color Plate 6).

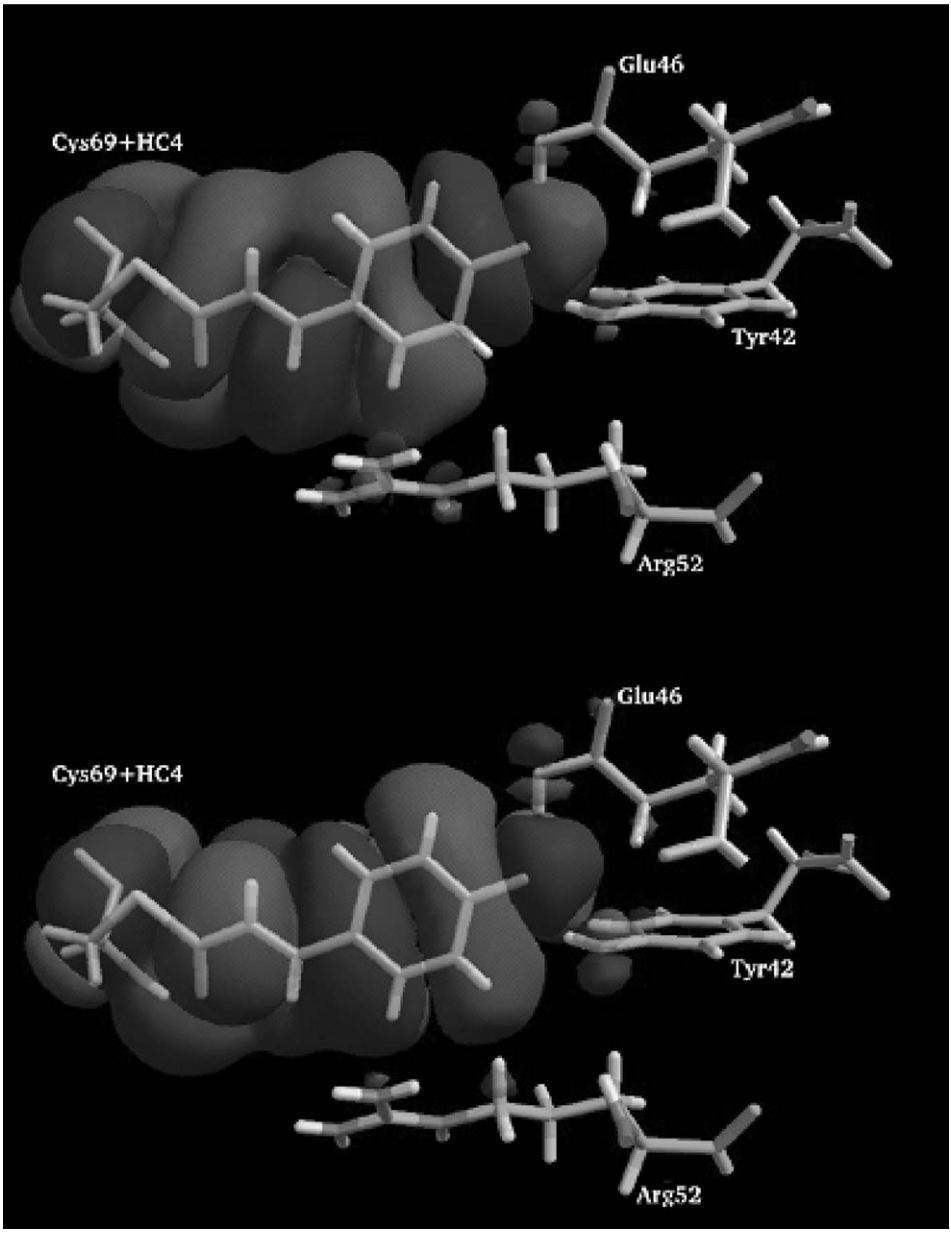


Fig. 2.8. Hole (lower graphic) and particle (upper graphic) natural orbitals of the lowest singlet excited state of PYP. Phasing is shown in red and blue. Occupation numbers are 1.06 for hole and 0.94 for particle. Original color figure can be found in the web version of Ref. [14] (see Color Plate 7).

experimental value of 2.78 eV (blue light) [36]. But, this discrepancy could be attributed to the missing correlation and relaxation at the CIS level of theory, and it would be remedied by the CIS(D) correction [34]. Works to implement the CIS(D) ability in ABINIT-MP are in progress. Finally, we should notice the timing of the MLFMO-CIS calculation on the PYP. The turnaround time for the total job was 11.2 hours, where the CIS kernel time was only 24.4 minutes. This demonstrates the applicability of the MLFMO-CIS method for realistic photoactive proteins.

2.6 SUMMARY

It is often observed that the quantum effects such as charge transfer, electronic polarization and bond breaking play a significant role in the context of molecular biology. The ligand-receptor docking is one of the examples, in which the binding affinity is governed by the interfacial charge transfer between the protein and the ligand molecules to a substantial extent. Such an observation demonstrates the limitation of the conventional simulation approaches based on the classical force field and the consequent necessity of quantum mechanical treatment of biomolecules. The ABINIT-MP software then provides a powerful and useful tool to enable the quantum chemical calculations for huge biomolecules based on the FMO method. We have been developing the improvement of the ABINIT-MP software in various ways, such as the inclusion of more accurate accounts for electron correlation effects and excited states, to make more efficient *ab initio* calculations for biomolecules feasible.

2.7 ACKNOWLEDGEMENTS

This work was supported by the 'Frontier Simulation Software for Industrial Science' project of the Ministry of Education, Culture, Sports, Science and Technology (MEXT) and by the Core Research for Evolutional Science and Technology (CREST) project of the Japan Science and Technology Agency (JST).

2.8 REFERENCES

1 K. Kitaura, T. Sawai, T. Asada, T. Nakano and M. Uebayasi, Chem. Phys. Lett., 312 (1999) 319.
2 K. Kitaura, E. Ikeo, T. Asada, T. Nakano and M. Uebayasi, Chem. Phys. Lett., 313 (1999) 701.
3 T. Nakano, T. Kaminuma, T. Sato, Y. Akiyama, M. Uebayasi and K. Kitaura, Chem. Phys. Lett., 318 (2000) 614.
4 K. Kitaura, S. Sugiki, T. Nakano, Y. Komeiji and M. Uebayasi, Chem. Phys. Lett., 336 (2001) 163.
5 T. Nakano, T. Kaminuma, T. Sato, K. Fukuzawa, Y. Akiyama, M. Uebayasi and K. Kitaura, Chem. Phys. Lett., 351 (2002) 475.
6 D.G. Fedorov R.M. Olson, K. Kitaura, M.S. Gordon and S. Koseki, J. Comput. Chem. 25 (2004) 872.
7 D.G. Fedorov and K. Kitaura, J. Chem. Phys., 120 (2004) 6832.
8 D.G. Fedorov and K. Kitaura, Chem. Phys. Lett., 389 (2004) 129.
9 D.G. Fedorov and K. Kitaura, J. Chem. Phys., 121 (2004) 2483.
10 H. Sekino, Y. Sengoku, S. Sugiki and N. Kurita, Chem. Phys. Lett., 378 (2003) 589.

11 S. Sugiki, N. Kurita, Y. Sengoku and H. Sekino, Chem. Phys. Lett., 382 (2003) 611.
12 Y. Mochizuki, T. Nakano, S. Koikegami, S. Tanimori, Y. Abe, U. Nagashima and K. Kitaura, Theor. Chem. Acc., 112 (2004) 442.
13 Y. Mochizuki, S. Koikegami, T. Nakano, S. Amari and K. Kitaura, Chem. Phys. Lett., 396 (2004) 473.
14 Y. Mochizuki, S. Koikegami, S. Amari, K. Segawa, K. Kitaura and T. Nakano, Chem. Phys. Lett., 406 (2005) 283.
15 Y. Mochizuki, Chem. Phys. Lett., 410 (2005) 165.
16 Y. Mochizuki, K. Fukuzawa, A. Kato, S. Tanaka, K. Kitaura and T. Nakano, Chem. Phys. Lett., 410 (2005) 247.
17 ABINIT-MP and BioStation Viewer are available from http://www.fsis.iis.u-tokyo.ac.jp/en/result/software/
18 Y. Komeiji, T. Nakano, K. Fukuzawa, Y. Ueno, Y. Inadomi, T. Nemoto, M. Uebayasi, D.G. Fedorov and K. Kitaura, Chem. Phys. Lett., 372 (2003) 342.
19 D.J. Mangelsdorf, C. Thummel, M. Beato, P. Herrlich, G. Schütz, K. Umesono, B. Blumerg, P. Kastner, M. Mark, P. Chambon and R.M. Evans, Cell, 83 (1995) 835.
20 M.G. Parker, N. Arbuckle, S. Dauvois, P. Danelian, and R. White, Ann. N. Y. Acad. Sci., 684 (1993) 119.
21 C. Sonnenschein and A.M. Soto, J. Steroid Biochem. Molec. Biol., 65 (1998) 143.
22 A.M. Brzozowski, A.C. Pike, Z. Dauter, R.E. Hubbard, T. Bonn, O. Engström, L. Öhman, G.L. Greene, J.-Å. Gustafsson and M. Carlquist, Nature, 389 (1997) 753.
23 A.K. Shiau, D. Barstad, P.M. Loria, L. Cheng, P.J. Kushner, D.A. Agard and G.L. Greene, Cell, 95 (1998) 927.
24 K. Fukuzawa, K. Kitaura, K. Nakata, T. Kaminuma and T. Nakano, J. Comp. Chem., 26 (2005) 1.
25 K. Fukuzawa, K. Kitaura, K. Nakata, T. Kaminuma and T. Nakano, Pure Appl. Chem., 75 (2003) 2405.
26 M.J. Frisch *et al.,* Gaussian 98, Revision A.7, Gaussian, Inc., Pittsburgh PA, 1998.
27 CHARMm: Chemistry at HARvard Macromolecular Mechanics (CHARMm) – Version 25.2 Revision: 98.0731.
28 G.G.J.M. Kuiper, J.G. Lemmen, B. Carlsson, J.C. Corton, S.H. Safe, P.T. van der Saag, B. van der Burg and J.-Å. Gustafsson, Endocrinology, 139 (1998) 4252.
29 G.A. Jeffrey, An Introduction to Hydrogen Bonding, Oxford University Press, New York, 1997.
30 A. Szabo and N.S. Ostlund, Modern Quantum Chemistry, MacMillan, New York, 1982.
31 D.G. Fedorov, T. Ishida and K. Kitaura, J. Phys. Chem. A, 109 (2005) 2638.
32 D.G. Fedorov and K. Kitaura, J. Chem. Phys., 122 (2005) 054108-1.
33 J.B. Foresman, M. Head-Gordon, J.A. Pople and M.J. Frisch, J. Phys. Chem., 96 (1992) 135.
34 M. Head-Gordon, R.J. Rico, M. Oumi and T.J. Lee, Chem. Phys. Lett., 219 (1994) 21.
35 A. Dreuw, J.L. Weisman and M. Head-Gordon, J. Chem. Phys., 119 (2003) 2943.
36 Y. Imamoto, M. Kataoka and F. Tokunaga, Biochem., 35 (1996) 14047.

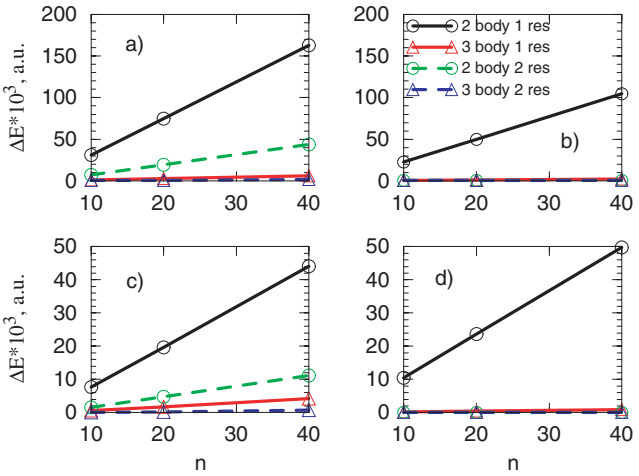


Plate 3. The error in energy (milliHartree) for FMO-DFT: (a) α-(ALA)n, (b) β-(ALA)n and for FMO-RHF: (c) α-(ALA)n, (d) β-(ALA)n, measured against the corresponding *ab initio* values. The basis set is 6–31G* throughout (illustration appears on page 25 of this volume).

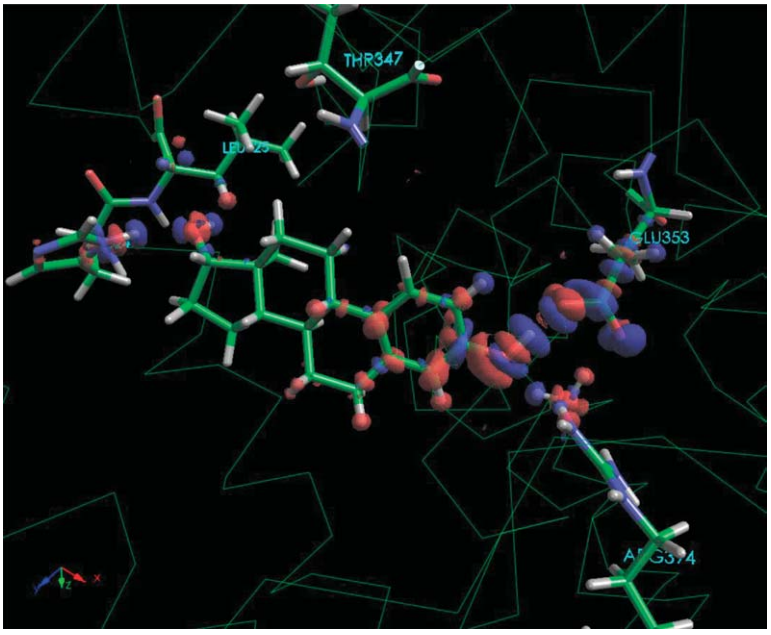


Plate 4. The difference of electron densities ($\Delta\rho$) between complexed and individual component molecules; $\Delta\rho = \rho_{complex} - (\rho_{ER} + \rho_{ligand})$. The stick representation refers to EST(1), a water molecule, and residues (Glu353, Thr347, Arg394, His524, and Leu525) with significant $\Delta\rho$ values. The thin line represents the $C\alpha$ backbone structure of EPα LBD except for the above residues. The blue (minus) and red (plus) represent the sign of the isosurface at ±0.002e/Bohr^3 drawn by BioStation Viewer [17] (illustration appears on page 45 of this volume).

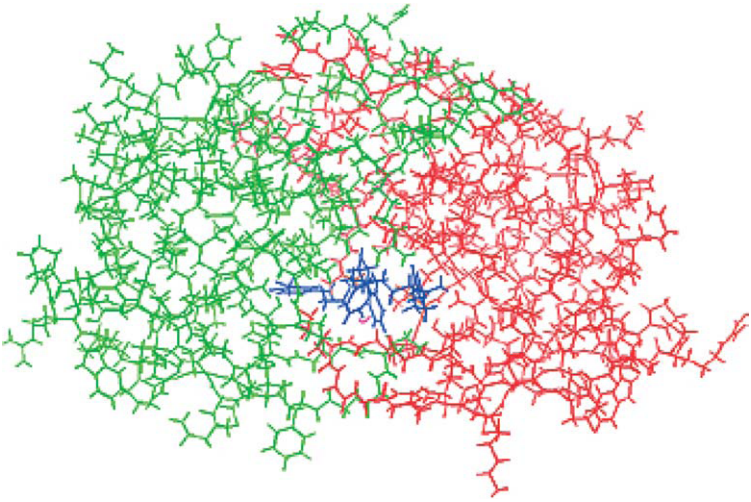


Plate 5. Graphic representation of HIV-1 protease and lopinavir. Two equivalent subunits are colored in red and green, and lopinavir is indicated in blue. Original color figure can be found in the web version of Ref. [13] (illustration appears on page 47 of this volume).

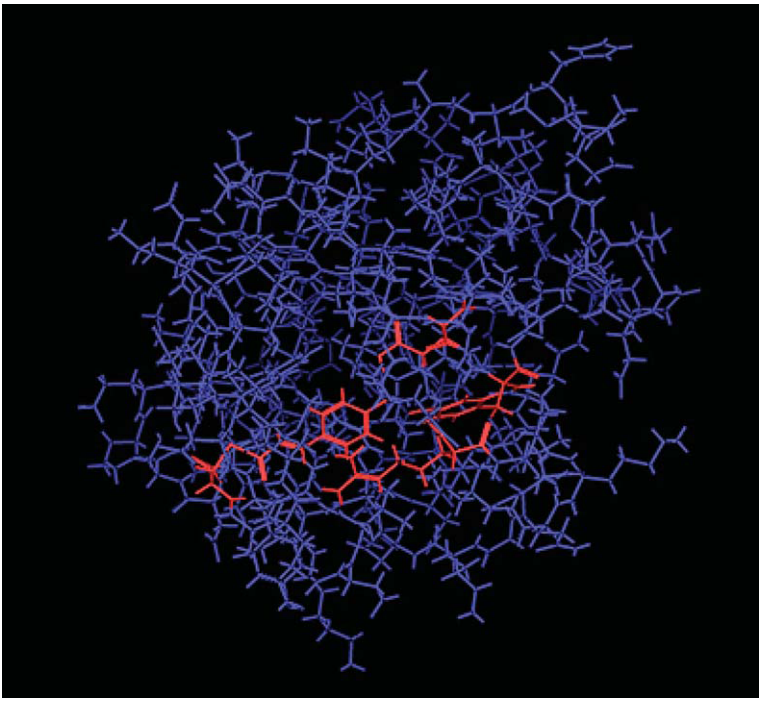


Plate 6. Graphic representation of PYP. The CIS region is colored in red. Original color figure can be found in the web version of Ref. [14] (illustration appears on page 49 of this volume).

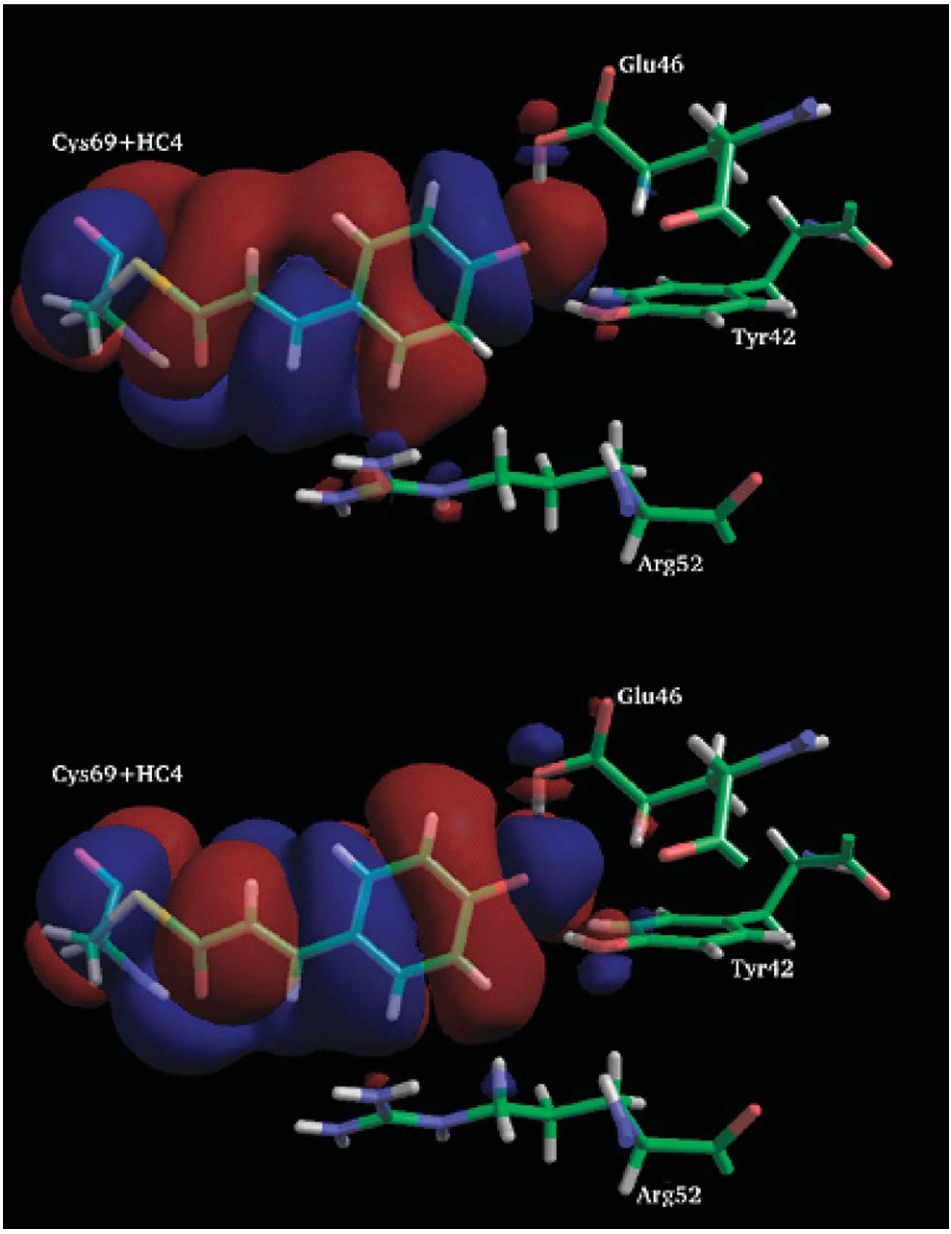


Plate 7. Hole (lower graphic) and particle (upper graphic) natural orbitals of the lowest singlet excited state of PYP. Phasing is shown in red and blue. Occupation numbers are 1.06 for hole and 0.94 for particle. Original color figure can be found in the web version of Ref. [14] (illustration appears on page 50 of this volume).

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 3

Combined DFT and electrostatic calculations of pK_as in proteins: study of cytochrome c *oxidase*

Dragan M. Popovic, Jason Quenneville and Alexei A. Stuchebrukhov

Department of Chemistry, University of California,
One Shields Avenue, Davis, CA 95616, U.S.A.

Abstract

Cytochrome *c* oxidase is a redox-driven proton pump which converts atmospheric oxygen to water and couples the oxygen reduction reaction to the creation of a membrane proton gradient. The structure of the enzyme has been solved; however, the mechanism of proton pumping is still poorly understood. Recent calculations indicate that one of the histidine ligands of the enzyme's Cu_B center, His291, may play the role of the pumping element. In this paper, we review the first principles calculations that are used to study models of the catalytic center of *CcO*, and calculate the pK_a of the His291 residue for both the reduced and oxidized states of the catalytic center of the enzyme. We also review the structure and function of *CcO*, describe the proposed mechanism of proton pumping, and identify the key problems that can be addressed computationally in this area and describe the methods for their solution. The outstanding theoretical and computational challenges of the area are also discussed.

3.1 INTRODUCTION

Cellular respiration, a process of oxidizing nutrient molecules to carbon dioxide and water, involves the transfer of electrons through a series of membrane protein complexes in the so-called respiratory electron transport chain. The initial electron donors here are the nutrient molecules, such as glucose; the final acceptor is atmospheric oxygen. This process is coupled with generation of an electrochemical proton gradient across the membrane, which drives the synthesis of ATP, and some other energy consuming processes [1,2]. The overall process is the major mechanism of biological energy transduction in cells.

Cytochrome *c* oxidase (*CcO*) is the terminal enzyme of the respiratory electron transport chain in mitochondria of eukaryotes and in aerobic bacteria; it catalyses the reduction of oxygen to water, and pumps protons across the membrane, a process which results in the membrane proton gradient [3–7]. In the course of its catalytic activity, electrons are transferred to the heme a_3 / Cu_B catalytic site, where the reduction of oxygen takes place, via two additional redox cofactors, Cu_A and heme *a* (see Fig. 3.1). In the cycle, the transfer of each of the four electrons, delivered by cytochrome *c*, is accompanied by translocation of two protons, of which one is consumed internally in the Cu_B–Fea_3 center for the reduction of oxygen and the other is pumped across the membrane [6,7]. The overall reaction can be expressed as follows:

$$O_2 + 4e^- + 8H^+(\text{in}) \rightarrow 2H_2O + 4H^+(\text{out}) \quad (1)$$

where (in) and (out) indicate two sides of the membrane, where the protons belong. The structure of *CcO* has been solved for several organisms, however, the molecular mechanism of proton pumping remains largely unknown. A detailed review of the

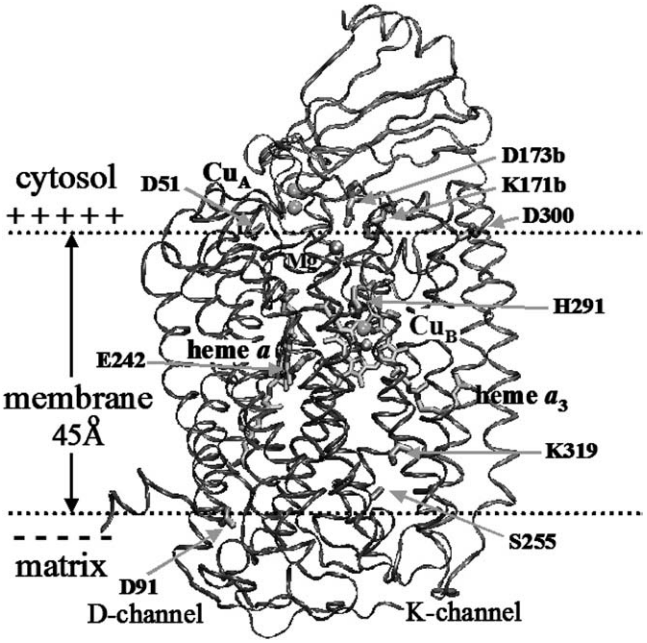


Fig. 3.1. The structure of the core catalytic subunits of cytochrome *c* oxidase from bovine heart [15] The subunit A, together with a helical extension of subunit B, is almost entirely integrated in the membrane, while the globular water-soluble domain of subunit B is exposed to solvent on the periplasmic side of the membrane. The two hemes and the Cu_B center are located in the helical part of subunit *A* that is embedded in a membrane of 45 Å thickness [23], while the Cu_A center consisting of two copper atoms is located in the subunit *B*, at the interface between the two subunits [108]. Several residues in the proton input D- and K-channels and on the output side, important for a proton translocation, are also displayed In the figure (see text for details) (see Color Plate 8).

enzyme structure together with recent experimental work on the kinetics of coupled electron and proton transfer reactions in the enzyme can be found in the literature [3–5,8−12].

For a proper function of the enzyme, a correlated motion of electrons and protons [13] coming from the opposite sides of the membrane is required. The electrostatic interaction between electrons and protons, which results in a dependence of the protonation state of the enzyme on its redox state, plays a key role here. To estimate such interactions, and thereby to get insights into possible mechanisms of the coupled reactions, computer simulations can be used.

In general, in redox metallo-enzymes the theoretical problem is multifaceted; it involves questions about pure electron transfer, with all complexities of the subject; proton transfer, no lesser problem in protein media, and finally their coupling. Electrons are transferred between metal sites via thermally assisted quantum tunneling, whereas protons, much heavier particles, need water intermediates to move between protonatable groups in the protein. Both electron tunneling and internal water in the proteins are rich subjects where computer simulation plays a key role. The overall kinetics of the coupled reactions in proteins can be rather complicated and involves multiple possibilities, a subject that only recently came to the focus of detailed experimental and theoretical studies.

One of the most basic questions one needs to address in any discussion of this rich subject is the nature and free energy of the initial and final states of the system. Once the initial and final states are well defined, and the free energy of these states is known, one can further address the question of kinetics, i.e., how the system actually goes from one state to another during the catalytic cycle of the enzyme.

In this review we will focus exclusively on the energetics problem. An interested reader is referred to reviews of other subjects from this area published elsewhere [13,14].

More precisely, the problem that we will be interested in is formulated as follows: for a given redox state of the metal centers of *CcO* (there are four of them: Cu_A, heme *a*, heme a_3 and Cu_B) find the equilibrium protonation state of the enzyme, and estimate energy of possible non-equilibrium states of interest that might be involved in the cycle of the enzyme. The equilibrium protonation states can be described in terms of pK_as of the protonatable groups in the protein, which are redox-dependent in our problem. Since the details of the mechanism of function are not known, the essence of this type of calculation is to estimate feasibility of the proposed hypothesis, and to make specific predictions that could be tested later experimentally. One of the problems of understanding the proton transport in *CcO* is that the key residues involved in redox-coupled proton activity still remain unidentified. Here we will show how computations allow one to provide crucial information about the enzyme.

The plan of the paper is as follows. We first add necessary details of the structure and function of *CcO*, describe the proposed mechanism of proton pumping, and identify the key problems that can be addressed computationally; next we describe the methods for the solution of the formulated problems, and then results of calculations will be presented together with a discussion of the assumptions of the computational models, factors that are important in such calculations, and theoretical and computational challenges of the area.

3.2 STRUCTURE AND FUNCTION OF CcO

Figure 3.1 shows the structure of *CcO*, location of all its four redox centers, and some important residues in proton entrance and exit channels.

The calculations reported here are based on the crystal structure of mitochondrial cytochrome *c* oxidase from bovine heart (PDB code, 2OCC) obtained at a resolution of 2.3Å, solved by Yoshikawa *et al*. [15]. The enzyme possesses 13 subunits and seven metal centers: heme *a*, heme a_3, Cu_A, Cu_B, Mg, Na and Zn. Subunit *A* consists of 12 transmembrane helices (514 residues) that are arranged in quasi-threefold rotational symmetry forming the three internal channels filled with water molecules. Subunit *A* contains heme *a*, heme a_3 and Cu_B, as cofactors. Histidines, His61 and His378, are axial ligated to heme *a*, while heme a_3 together with copper atom (Cu_B) and coordinating ligands (His376 to iron and His240, His290, His291 to copper) form the binuclear center. One of the histidines (His240) bound to Cu_B is cross-linked to a tyrosine (Tyr244) by post-translational modification [15–17].

Subunit *B* possesses 227 residues and Cu_A center. It consists of a globular domain located on the periplasmic side of the membrane and two transmembrane helices, which interact with helices of subunit *A*. The Cu_A center consists of two copper atoms bridged with two cysteins (Cys196B, Cys200B) and one histidine ligand each (His161B or His204B). In addition they are connected to Met207B (sidechain Sδ atom) or Glu198B (backbone O), such that ligands of both copper atoms form distorted tetrahedrons.

Two other redox inactive metal ions, Mg^{2+} and Na^+, are located in the subunit *A*. The Mg center is located at the interface between subunits *A* and *B* and very close to the periplasmic-membrane surface. It is ligated by His368, Asp369, Glu198*B* and three water molecules. The whole complex is neutral since two negatively charged acidic groups compensate for the charge of the Mg^{2+}-ion. Also, Na^+ is neutralized by ligation of Glu40, Gly45 (backbone oxygen), Ser441 and one water molecule. This binding site is located in proximity of propionate *A* of heme *a* (PRAa).

Previous studies have shown that subunits *A* and *B* from different species exhibit remarkable similarity [16,18,19] and the two-subunit enzyme is active in proton pumping [20,21]. Additional subunits may have a supporting role [18,19], maintaining the structural integrity [22] and making an additional shield of protection around the catalytic core-subunit. Therefore, all our calculations were performed only on chains *A* and *B* including a membrane model, as an infinite low dielectric slab of 45Å thickness [23], that shields the membrane-exposed protein surface from the solvent, Fig. 3.1. By these simplifications, the essence of the system is not lost, yet its size is significantly reduced.

To summarize the discussion of this and the previous section, the function of the enzyme can be described as follows. Electrons are delivered, one at a time, to Cu_A center, see Fig. 3.1, from where they are transferred via heme *a* to the binuclear catalytic center consisting of heme a_3 and Cu_B. In the catalytic center a dioxygen molecule is bound (and split) to both Fe and Cu ions. Upon each electron transfer, one proton is transferred to the catalytic center, and one proton is pumped from the negative (lower) side of the membrane, to the positive (upper) side, as shown in Fig. 3.1. Both chemical and pumped protons originate from the negative (lower) side of the membrane.

The fundamental question concerns the molecular mechanism of proton pumping, i.e., how exactly the free energy of oxygen reduction in *CcO* is transformed to the driving force that pushes pumped protons against the membrane electrochemical gradient.

3.3 THE PROTON PUMPING MECHANISM

The proposed proton pumping mechanism [24] is based on the following key structural features and observations made in previous studies, see Fig. 3.2. There are two chains of water molecules connecting Glu242 to propionate *D* of heme a_3 ($PRDa_3$) and to the catalytic site of the enzyme (i.e., a hydroxide ion bound to either Fe or Cu metal centers in BNC) [25,26]. $PRDa_3$ has a strong salt bridge with a nearby Arg438. No proton transfer to $PRDa_3$ from Glu242 is possible, unless Arg438 is deprotonated. There is a water

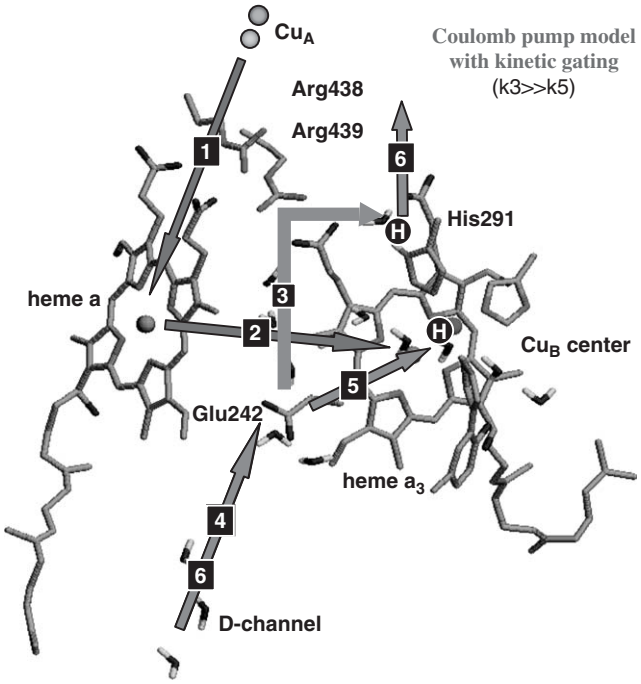


Fig. 3.2. The key structural elements of the proposed pumping mechanism of cytochrome *c* oxidase [24], and the sequence of transitions during one pumping cycle. The relative position of the four metal redox-active centers (Cu_A, heme *a*, heme a_3 and Cu_B complex) and the three protonatable sites (Glu242, His291 and OH^- ligand in the BNC) are shown. According to the model, a stepwise electron transfer that brings an electron from Cu_A to the binuclear center (consisting of heme a_3 and the Cu_B complex) is coupled to proton translocation from Glu242 to the His291 group, which serves here as a proton-loading site of the pumping mechanism in *CcO*. Two protonation sites, PLS and one in the BNC, are shown as H-circles. The PT and ET steps are shown by arrows. The key assumption of the model is that the proton transfer rate upon ET between the two hemes (step 2) is much higher along the pumping channel 3 than along the chemical channel 5 ($k_3 >> k_5$), in other words step 3 occurs before step 5 (see Color Plate 9).

molecule, Wa3, both predicted in Ref. [25] and visible in crystal structure [16,27], which is located between two propionates of heme a_3, and hydrogen bonded to δ1 nitrogen of His291–the 'top' ligand of Cu_B center. With this molecule there is proton connectivity between Glu242 and Nδ1 of His291 via the Glu242−PRDa_3−Arg438–Wa3–His291 chain. According to electrostatic calculations [28], the Nδ1 position of His291 is protonated only when both Fe and Cu_B centers are in formally reduced states (i.e. when the total charge on the metal ion plus the ligand is 2+ for Fe-a_3, and 1+ for Cu_B). When one of the metal centers of BNC is oxidized (or, equivalently, when one of the ligands in BNC accepts a proton), His291 becomes deprotonated.

The following mechanism of proton pumping is proposed (see Fig. 3.2). During the cycle, the stable state of the catalytic center, before any additional electrons are supplied to the system, is that in which one of the metal centers is formally oxidized, e.g. $Fe^{3+}-H_2O$, or $Cu^{2+}-H_2O$. This state is established in a previous step of the cycle, when a chemical proton is accepted by one of the hydroxy ligands of the binuclear center. In this state, His291 (i.e., Nδ1 position) is deprotonated, and Glu242 is protonated.

(1) and (2) An electron is supplied to the system via cyt *c* to Cu_A, which is transferred to heme *a* and then to heme a_3–Cu_B binuclear center. One of the metal ions is reduced, and the overall charge of the BNC becomes one charge unit more negative.

(3) In response to the increased negative charge of the binuclear center, the proton from Glu242 has now a driving force to move closer to the binuclear center. There are two pathways leading from Glu242 to two possible sites: one is leading to the binuclear center (BNC) itself, the second is leading to the His291 Nδ1 position–the proton loading site (PLS). The assumption (kinetic gating) is made that the rate of proton transfer to His291 is much higher than that of transfer to the binuclear center (the basis for this key assumption is discussed later this chapter). Therefore, the protonation of His291 occurs *before* that of the binuclear center. The fast proton transfer from Glu242 to His291 occurs by Grotthuss mechanism via Arg438 and PRDa_3, details of the transfer are discussed below. With this transition, His291 becomes protonated, as equilibrium electrostatic calculations predict for a redox state of the BNC in which now both metal centers are in a formally reduced state.

(4) In this step Glu242 is reprotonated.

(5) Now the second chemical proton can be transferred to the binuclear center, using the second path connecting Glu242 and the catalytic center of *CcO*. A significant driving force, despite the presence of the proton on a nearby His291, is due to formation of a water molecule in BNC– which is the main source of energy in the process. This transition occurs after the first proton has moved to the His291 proton loading site. The formed state of BNC is such that one of the metal ions is formally oxidized (one positive charge now added). In this redox state, at equilibrium, His291 has to be deprotonated.

(6) The previously formed state has two additional protons present in two closely located sites: PLS (His291) and BNC, but only one additional electron residing on one of the metal ions of BNC. This state is metastable because of significant proton repulsion. The state is stabilized, therefore, by the expulsion of the proton from the His291 PLS site. The expulsion of the proton from His291 from a state in which one

of the metal centers is formally oxidized is predicted by electrostatic calculations, which show that in this redox state, His291 has to be deprotonated.

(7) Glu242 is reprotonated again, and one turnover of the cycle is complete. The formed state is stable until the next electron is passed through the system. This last step of reprotonation of Glu242 may be correlated with the expulsion from His291, and/or subsequent re-reduction of heme *a* in the next turnover in the cycle.

Each time a proton is pumped, in a single turnover of the pump, a stable state is formed until the next electron is injected into the system. Thus, for each electron passing through the chain, there is one pumped proton. For each oxygen molecule, four electrons are required to form two water molecules; therefore a maximum of four protons can be pumped.

The proposed model is based on the simulation of water inside of the protein [25,29] and the combined DFT and electrostatic calculations of energies of different protonation states of the enzyme cycle. Below we review these calculations and demonstrate that the results indeed provide support for the key assumptions of the model. We will focus our discussion on the calculation of the redox-dependent value of His291 – the key element in the proposed pump. Other results related to the above model can be found in the original papers from this group [28,30−32].

The assumptions of the model are of two types: energetic, such as that His291 changes its protonation state depending on the redox state of the metals of the catalytic center, and kinetic, such as upon electron reduction of the catalytic center, the proton from Glu242 it transferred to His291 and *not* to the OH^- group in the catalytic center. To verify the latter, the rates of the proton transfer reactions need to be calculated, the subject that is outside the scope of this review. We will focus here only on the energetics questions.

3.4 COMPUTATIONAL METHODS FOR pK_a CALCULATIONS

The combination of the DFT and electrostatic calculations has been used successfully in the past to compute the pK_a values and redox potentials in various enzymes and proteins [33−37]. This approach takes advantage of both methods by combining the accuracy of quantum chemical calculations with a detailed description of electrostatic interactions of the whole protein. Density functional theory is applied to a relatively small quantum–mechanical system (QM) of interest to optimize its geometry, to evaluate electronic energies, and to obtain the atomic partial charges for different redox and protonation states of the complex. Electrostatic calculations, on the other hand, are used to evaluate the solvation energy of the complex in the protein environment and the Coulomb interactions of the complex with protein charges. Both the polarization of the protein medium and the protein charges affect the electron distribution of the complex, therefore the electrostatic and DFT calculations are carried in a self-consistent way. This method was shown to produce pK_a values in proteins that are generally in good agreement with experimental data [34,37–40].

3.4.1 Overview

Here we briefly summarize the DFT/electrostatic method of pK_a calculations, and introduce different energy terms, which are discussed later in the text.

In the calculation we divide the whole system into an active site complex (QM system) and the surrounding medium–protein, membrane, and the external aqueous phase. The main idea is to treat the QM system with an accurate DFT method–in this case we are interested in proton affinity of a complex–and to include the dielectric effects of the medium with classical electrostatic methods. The protein medium also includes partial atomic charges, charges of redox centers, and charges of the titratable groups, which directly affect the electronic energies of the complex.

The dielectric medium involves protein itself, the membrane, and the surrounding solvent. The protein by itself is an inhomogeneous dielectric, because it has many internal cavities which contain water molecules. These regions should have a higher dielectric constant than the regions of 'dry' protein. For regions of dry protein a specific value of dielectric constant ε_p is assumed, and for the cavities a higher value ε_c is used. Neither of these values is well defined, although most researches would agree that ε_p should be close to 4, while the cavities can have dielectric constant ε_c anywhere from 4 to 80. The upper value would correspond to a dielectric response equivalent to that of the bulk water.

We begin with a review of pK_a calculations in water. The pK_a value is related to the free energy of deprotonation in aqueous solution (ΔG_{aq}^{deprot}), which is a sum of two contributions: the free energy of deprotonation in vacuum (ΔG_{vac}^{deprot}), and the solvation energy difference between the deprotonated and protonated forms of the protonatable group ($\Delta\Delta G_{solv}^{deprot}$), see also [31,37]:

$$pK_a = \frac{1}{kT \ln 10}(\Delta G_{aq}^{deprot}) = \frac{1}{kT \ln 10}(\Delta G_{vac}^{deprot}) + \Delta\Delta G_{solv}^{deprot} \quad (2)$$

The solvation energy difference $\Delta\Delta G_{solv}^{deprot}$ obtained from Eq. (2):

$$\Delta\Delta G_{solv}^{deprot} = \Delta G_{solv}^{A^-} + \Delta G_{solv}^{H^+} - \Delta G_{solv}^{HA} \quad (3)$$

The solvation energy of deprotonated ($\Delta G_{solv}^{A^-}$) or protonated (ΔG_{solv}^{HA}) species is calculated by solving the Poisson–Boltzmann equation as described in more detail below. The solvation energy of a proton, $\Delta G_{solv}^{H^+}$, is set to –260.5 kcal/mol, which is estimated from the experimentally measured potential of the standard hydrogen electrode [41].

The first term in Eq. (1), the free energy of deprotonation in vacuum, can be formally expressed as:

$$\Delta G_{vac}^{deprot} = \Delta E_{elec} + \Delta E_{vib} + \Delta E_{trans}^{H^+} + \Delta(pV) - T(S^{H^+}) \quad (4)$$

From left to right, the terms in Eq. (3) denote the electronic and the zero point vibrational energy differences between the deprotonated and protonated species, the energy of proton ejection, the energy due to the gas phase volume change, and finally the entropy contribution.

As an example of the above calculations the absolute pK_a of MeIm was obtained. The vibrational energy term (ΔE_{vib}), whose main contribution is the zero-point energy of the N–H bond that is being lost upon deprotonation, was estimated from a normal-mode analysis of MeIm to be –9.0 kcal/mol. The main assumption here is that the frequency change of the N–H bond upon coordination to copper ion is very small, therefore the zero-point energy is transferable and needs to be determined only once for the proper

model compound, i.e., MeIm. The translational energy of a proton ($\Delta E_{\text{trans}}^{H^+}$) is estimated to be $3/2k_BT$, while the energy due to the gas-phase volume change is estimated from ideal gas approximations to be k_BT. Finally, the entropic term of the gas-phase free energy of a proton was set to 7.8 kcal/mol as derived from the Sackur–Tetrode equation [42]. We obtain values of 7.7 and 7.8 for the δ– and ε–pK_a values [31], respectively (compared with established experimental values of 6.6 and 7.0 for histidine) [43].

To probe the accuracy of calculations of the electronic energies, the experimentally measured enthalpy of proton affinity can be used. The proton affinity is defined as $-\Delta H_r$ for the reaction $B + H^+ \rightarrow BH^+$ and is therefore written [44]

$$PA = \Delta H_{\text{vac}}^{\text{deprot}} = \Delta E_{\text{elec}} + \Delta E_{\text{vib}} + 5/2\ kT \tag{5}$$

where the change in vibrational energy during heating from 0 to 298 K has been neglected. The proton affinity of the nitrogen at the 3-position in 4-methylimidazole that we calculate (227.0 kcal/mol) compares well with the most recent experimental value (226.2 kcal/mol) [45].

In a similar way, the *absolute* values of pK_a in a protein can be obtained. Alternatively, one can compute the pK_a *shift* of the group of interest relative to a suitable model compound, and by using the experimentally determined pK_a value of the model compound, to evaluate the pK_a of the group as

$$pK_a^{\text{site}} = pK_a^{\text{model}} + \Delta pK_a \tag{6}$$

The relative shift of the group's pK_a with respect to the model compound can be expressed as

$$\Delta pK_a = \frac{1}{kT \ln 10}\,\Delta\Delta G_{\text{shift}} = \frac{1}{kT \ln 10}\,(\Delta\Delta E_{\text{elec}} + \Delta\Delta G_{\text{solv}}) \tag{7}$$

where $\Delta E_{\text{elec}} = \Delta E_{\text{elec}}^{\text{depr}} - \Delta E_{\text{elec}}^{\text{prot}}$ is a quantum-mechanically (QM) calculated electronic energy difference in the gas phase between the deprotonated and protonated forms, and $\Delta\Delta E_{\text{elec}}$ is its shift relative to the model compound; $\Delta G_{\text{solv}} = \Delta G_{\text{solv}}^{\text{depr}} - \Delta G_{\text{solv}}^{\text{Prot}}$ is the difference in solvation energy between the deprotonated and protonated forms, and $\Delta\Delta G_{\text{solv}}$ is the shift of the solvation energy relative to the model compound. (In Eq. (6), when the *difference* of pK_as is computed between two compounds, the free energy of a solvated proton as well as some entropy contributions cancel out; thus, e.g., only electronic energy, i.e., the enthalpic part, of the vacuum term remains.)

The total solvation energy (x = *deprotonated or protonated* form) is divided into several components:

$$G_{\text{solv}}^{x} = G_{\text{Born}}^{x} + G_{\text{strain}}^{x} + G_q^{x} \tag{8}$$

The Born solvation energy (G_{Born}^{x}), and the energy of interaction with protein charges (G_q^{x}), also known as the reaction and protein field terms, respectively, are two main contributions to the free energy of solvation of the active site complex in the protein environment. The third term (G_{strain}^{x}) is the so-called 'strain' energy; this term reflects the energy cost associated with the reorganization of electronic density distribution of the

active site complex induced by the reaction field of the polarized medium. (The origin of this term is as follows. The charges of the QM system polarize the surrounding medium; in turn, the induced polarization of the medium slightly changes the charge distribution of the QM system, compared with the initial charge distribution in the vacuum.)

The pK_a of the active site complex can finally be calculated as

$$pK_a^{\text{site}} = pK_a^{\text{model}} + \frac{1}{kT \ln 10}(\Delta\Delta E_{\text{elec}} + \Delta\Delta G_{\text{strain}} + \Delta\Delta G_{\text{Born}} + \Delta\Delta G_q). \qquad (9)$$

We have confirmed that both methods – the absolute and relative pK_as – give practically identical results. In the following we will be using the above expression and the relative method. For calculation of pK_a of His291, the model compound 4-methylimidazole will be used.

3.4.2 Density functional calculations

The starting structure of the active site was taken from the X-ray crystal structure of bovine heart cytochrome *c* oxidase obtained by Yoshikawa *et al*. at 2.3 Å resolution (PBD code, 2OCC) [15]. The QM-model (Fig. 3.3) used to calculate the pK_a of a Cu_B ligand consists of the Cu atom, methylimidazole molecules representing coordinated histidines 240, 290 and 291, a methyl group representing tyrosine 244 (which is cross-linked to H240), and an H_2O Cu_B-ligand. The methyl groups attached to the imidazoles represent the β-methylene groups of histidine.

Energies of each coupled protonation/redox state of the complex were calculated using density functional theory (DFT) [46,47] and the Jaguar 5.5 quantum chemistry package [48]. All calculations used the hybrid density functional B3LYP [49], and open-shell electronic configurations (for oxidized Cu_B) were calculated with restricted open-shell DFT. Reported energies used the LACV3P+* basis set, while geometry optimizations used LACVP+*, as abbreviated in the Jaguar program [48,50]. These basis sets include non-relativistic electron core potential for the Cu atom. The basis sets for the metal valence electrons (and all

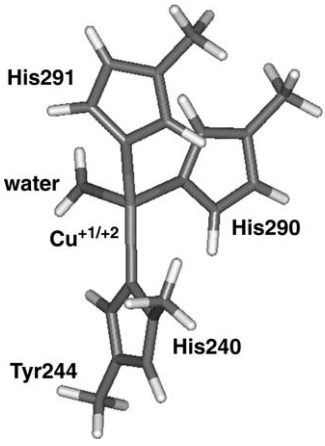


Fig. 3.3. The model of the Cu_B center used in the pK_a calculation of His291 (see Color Plate 10).

non-metal electrons) are of split valence quality (6-31+G*) [51–56] for LACVP+* and triple split valence quality (6-311+G*) [57–60] for LACV3P+*. They include polarization and diffuse functions for all heavy atoms.

Geometry optimizations were conducted in Cartesian space, in vacuum. In the procedure, the Cu atom, the ligating ε– nitrogen and methyl group of methylimidazoles were kept fixed. This scheme allows relaxation of the remaining atoms while preserving the integrity of the crystal stucture with respect to both the unique Cu-ligand geometry as well as the attachment location of the histidines to the protein backbone. Four optimizations were performed for the protonated and deprotonated forms of His291, and each of these for oxidized and reduced CuB. The geometry of the Cu-ligated water molecule was set using molecular mechanics. This water molecule is known to be very weakly bound and quite mobile [61,62].

To quantify the effect of the electronic reorganization of the solute and to obtain the corresponding set of the ESP-fitted charges for the QM considered system, the complex was surrounded by a continuum dielectric ($\varepsilon = 4$ or 80) and the self-consistent reaction field (SCRF) method was applied as implemented in Jaguar 5.5 [63,64]. The probe radius of the surrounding dielectric was set to 1.4 Å, and the standard Jaguar set of the van der Waals radii for atoms was used. The ESP atomic charges were generated by using a modified version of the CHELPG procedure of Breneman and Wiberg [65].

To correctly describe the effects of the protein charges on the pK_a of His291, the equilibrium proton distribution in *CcO* for the corresponding redox state of metal centers is required. To this end, we performed the standard continuum electrostatic calculations (see next section) to determine the equilibrium protonation state of the titratable residues in *CcO*, embedded in membrane, for different redox states of the enzyme. In different redox states the equilibrium proton distribution is slightly different due to the proton uptake from the solution (aqueous phase), and very different from the standard protonation state.

3.4.3 Continuum electrostatic calculations

The standard electrostatic calculations have been recently applied in studies of bacterial cytochrome *c* oxidase [66] and several other membrane proteins that take part in energy transduction in cells, such as bacterial photosynthetic reaction center [67–69], bacteriorhodopsin [70,71] and cytochrome bc_1 complex [72], have been recently investigated by electrostatic methods.

The protonation states of the titratible residues in *CcO* for different redox states of the enzyme were determined as described in Ref. [28], by numerically solving the linearized Poisson–Boltzmann equation on a grid [72–81] (with a three-step focusing procedure) for the inhomogeneous dielectric of the protein–membrane–water system with the program MEAD [82,83]. The protonation states sampling was subsequently done by employing a Monte-Carlo titration technique as implemented in the program KARLSBERG [84]. The electrostatic calculations were performed on chains *A* and *B*, which belong to monomer I from the crystal structure of mitochondrial *CcO* from bovine heart (PDB code, 2OCC) [15]. We considered here 146 sites as protonatable, including all residues such as: Asp, Glu, Arg, Lys, Tyr, Cys, heme propionates, N- and C-termini, δ– and ε–His sites, with the exception of the residues coordinated to the metal centers. In addition, Tyr244 and His291, belonging to Cu_B center, are considered as titratable as well. Of three His ligands at the

Cu_B center (240, 290, 291), only His291 can possibly be a titratable site, because His240 is cross-linked to Tyr244, whereas His290, according to the crystal structure, is hydrogen bonded to nearby Thr309, which in turn is hydrogen bonded to the carbonyl oxygen of Phe305 in all redox states [85] and therefore is always protonated at the Nδ1 position. To several explicit water molecules included in the calculation were assigned TIP3P charges [86].

In our calculations, the membrane is modeled as a low dielectric slab of 45Å that covers the central part of the enzyme. Taking the thermodynamic average over different protonation states of the titratable groups one can evaluate the equilibrium charge distribution in *CcO* for the given temperature, pH and redox state of metal centers. The obtained charge distribution of the protein is used in electrostatic calculations of the solvation energy term G_q^x, Eq. (4). We found that using the standard protonation state of the protein leads to completely erroneous results.

3.4.4 Solvation calculations

Once the charges of the active site complex have been calculated, its geometry has been optimized, and an equilibrium charge distribution of the protein has been obtained, the DFT optimized active site cluster needs to be inserted back into the protein in order to set up the electrostatic solvation calculations. The MEAD program [83] has been used to calculate the solvation energy of the quantum-chemically treated cluster in the aqueous phase, a continuum low dielectric medium or in the protein (see below all different models). Solvation of the active site cluster in the protein consists of two main contributions – the Born-solvation energy and the energy of the interaction with the protein charges (reaction and protein field energies).

The Poisson equation (or PBE) is solved numerically for the system divided into three dielectric regions, represented by dielectric constants of 1, 4 and 80, for the quantum-chemically considered active site, a protein, and solvent region, respectively. Such choice of the values for the dielectric regions is physically reasonable and very well supported by numerous studies [87–90]. Using $\varepsilon_{protein} = 4$ approximately compensates for the effects of electronic and nuclear polarization, allowing some mobility of the protein dipoles and reorientational relaxation of the protein [35], which otherwise are not explicitly taken into account. However, in the QM considered system, electronic and nuclear degrees of freedom are treated explicitly, no additional shielding is required and therefore $\varepsilon_{QM\text{-}system} = 1$. $\varepsilon_{water} = 80$ is traditionally used in continuum solvent electrostatic calculations to describe the screening properties of the aqueous phase. In addition to these typical values for the dielectric regions, we also examined the dependence of the results on the various values of the dielectric constant associated with water-filled cavities inside the protein.

In these calculations, the ESP-fitted charges are used for the QM system, while partial charges of the protein atoms are taken from the CHARMM22 parameter set [91], modified so as to reflect the equilibrium protonation state of the enzyme [28]. The backbone atoms of the Cu_B ligands are also included in electrostatic calculations. They affect the energetics through the influence of the background charges (so they are not part of QM system). The protein charges [72] and radii [91] that we use in this application

have been used successfully before to calculate the electrostatic energies, protonation probabilities of titratable groups and redox potentials of cofactors in different proteins [28,72,81,92,93].

The electrostatic potential for the system was solved by the Poisson equation (PE) with a three-step grid-focusing procedure. The lattices with highest resolution were centered at the active site cluster. The probe radius of the surrounding dielectric was set to 1.4 Å. Ionic strength was set to 0, since we found earlier that the mobile ions do not make any significant contribution to the calculated potentials in the interior of the *CcO*, where the active site complex is located [28]. Similar weak dependence on the ionic strength has been reported for other trans-membrane protein systems [66,68,71,72,94,95].

The pK_as of His291 and Glu242 reported here were calculated using the equilibrium protonation states of the enzyme, which were determined as in our previous work [28] for each redox state involved. It should be noted that the pK_as of all groups in the enzyme should be calculated in a self-consistent manner, and in general, the protonation state of the rest of the protein should depend on the protonation state of the selected group. Here the self-consistency is easily achieved, because the found protonation state of His291 and Glu242 is either essentially 1 or 0, depending on the redox state of the enzyme.

To examine the strength of the coupling of the protonation state of His291 and other protonatable groups of the protein, two sets of charges for the protein were obtained – for protonated and for deprotonated states of His291. The pK_a values of His291 determined for these two sets of charges were found to differ by less than 0.3 pK_a units.

3.5 RESULTS AND DISCUSSION

3.5.1 Computational models

In order to get a sense of how different factors affect the apparent pK_a values of His291, we considered several computational models shown in Fig. 3.4.

Model (a) represents the quantum-chemically (QM) treated system, Fig. 3.3, solvated in a continuum dielectric of $\varepsilon = 80$. Using this model, we obtain the pK_a values of the groups in water.

Model (b) is the same as (a), only for dielectric $\varepsilon = 4$. The pK_as obtained from this model should illustrate the magnitude of the effect of a low dielectric environment of the protein.

Model (c) adds protein charges to the uniform dielectric of $\varepsilon = 4$. This is still a very rough description of the protein system because the complexity of the protein-solvent surface, inner water-filled cavities, and solvent itself is missing.

Model (d) introduces the inhomogeneity of the dielectric medium represented by the inner water cavities with $\varepsilon = 80$, shape of the protein, membrane, and the outside aqueous phase. No protein charges are present here.

Finally, model (e) includes all components of the protein system – inhomogeneity of the dielectic environment, as in (d), and protein charges as in (c).

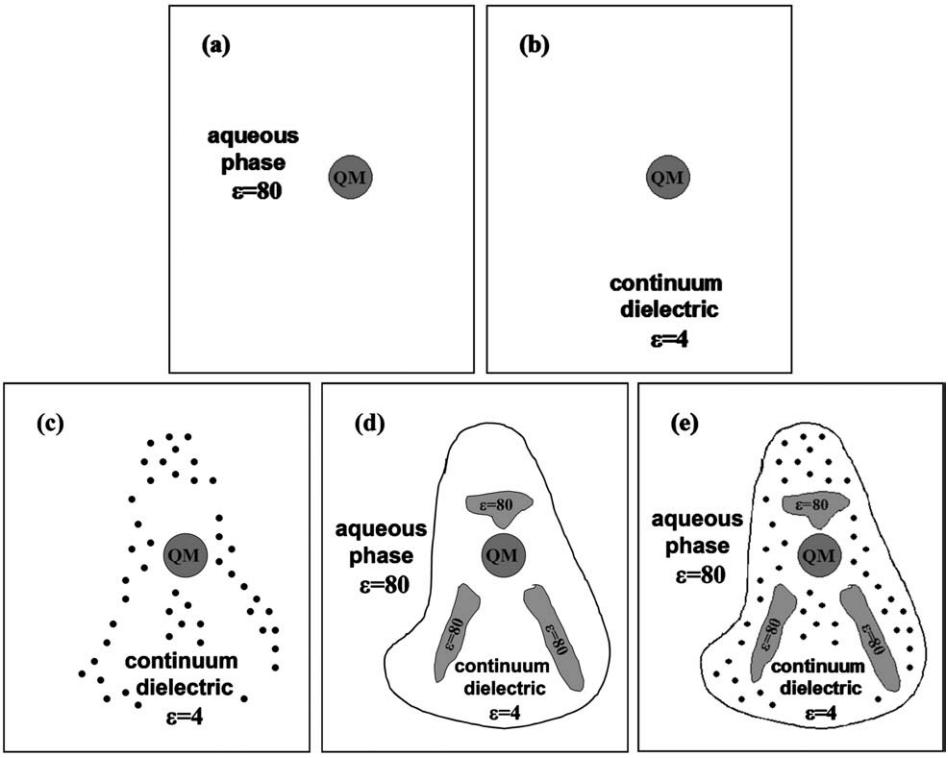


Fig. 3.4. Schematics of different models used in this study. In a model (a) the QM system that consists of a Cu_B center and coordinated ligands (with dielectric constant $\varepsilon = 1$ inside) is solvated in the aqueous phase. (b) The QM system is solvated in a low dielectric of $\varepsilon = 4$. (c) The QM system is embedded in a continuum dielectric of $\varepsilon = 4$ and surrounded with protein charges. (d) The QM system is embedded in a inhomogeneous dielectric medium represented by $\varepsilon = 4$ within the borders of the protein, and $\varepsilon = 80$ inside the inner water filled cavities and outside the protein, while protein charges are neglected. A model (e) includes a full protein-solvent description including dielectric inhomogeneities and protein charges, whereas the QM system is treated by the DFT method. Crystalographically, more than 150 water molecules have been found in the inner protein cavities [16,27,96]. Only the most important water filled cavities are schematically shown in the figure. They represent the proton input D- and K-channels leading to the active site and the large water filled cavity on the proton output side of the protein (see Color Plate 11).

3.5.2 The calculation of pK_a in water

As explained in the Methods section, in order to calculate pK_as in a protein, we first calculate the pK_a values of the corresponding groups in water. For a calculation of the pK_a values of δ-nitrogen of His291 ligand of Cu_B complex, we use MeIm as a reference compound, for which the experimental pK_a is known. The experimental value of pK_a of δ-nitrogen of MeIm is 6.6 [43]. Using this value and the difference in electronic and solvation energies between the MeIm and Cu_B model complex, we can evaluate the pK_a value of the latter in the aqueous phase. The results are presented in Tables 3.1 and 3.2. (Here, and later in the text, by protonated and deprotonated forms of MeIm we mean protonated and deprotonated δ-nitrogen of the compound, i.e., cationic and neutral forms of MeIm, see

Table 3.1 Solvation energy difference (ΔG_{solv}, in kcal/mol) between the deprotonated and protonated forms of 4-methylimidazole and of the Cu_B complex (oxidized or reduced) in aqueous solution. Here and elsewhere, the protonated and deprotonated forms are the corresponding states of δ-nitrogen of MeIm and the His291 ligand of the Cu_B compound. (i.e., a protonated form of MeIm is a cation, while the deprotonated MeIm is a neutral form where δ-proton gets deprotonated.) The "strain" and the Born contributions are given as a difference in energy between the deprotonated and protonated forms

Compound	$E_{reorganization}$		'Strain' energy	Born energy	ΔG_{solv} in water ($\varepsilon = 80$)
	Deprotonated form	Protonated form			
4-methylimidazole	+4.0	+1.5	+2.5	+52.2	+54.7
$Cu_B^{2+}(H_2O)$ complex	+10.1	+3.4	+6.7	+86.7	+93.4
$Cu_B^{1+}(H_2O)$ complex	+9.5	+3.3	+6.2	+8.5	+14.7

Table 3.2 The calculated pK_a values of the His291 site in the oxidized and reduced Cu_B complex in the aqueous phase relative to the pK_a of the model compound (methylimidazole). The corresponding electronic and solvation contributions to the pK_a calculations, and their differences relative to the reference compound, are also shown in the table. Energies are reported in kcal/mol.

Compound	ΔE_{elec}	$\Delta\Delta E_{elec}$	ΔG_{solv}	$\Delta\Delta G_{solv}$	$\Delta\Delta G_{shift}$	ΔpK_a	pK_a
4-methylimidazole	+234.6	0	+54.7	0	0	0.0	6.6[a]
$Cu_B^{2+}(H_2O)$ complex	+198.7	-35.9	+93.4	+38.7	+2.8	+2.0	8.6[b]
$Cu_B^{1+}(H_2O)$ complex	+283.7	+49.1	+14.7	−40.0	+9.1	+6.6	13.2[b]

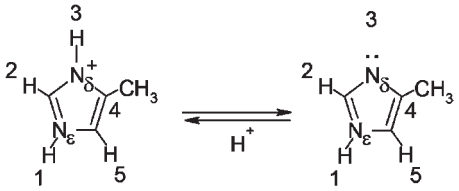
[a]Experimental value
[b]Calculated value

Scheme 3.1. Similarly, protonated and deprotonated forms of Cu_B center mean protonated and deprotonated δ-nitrogen of His291 ligand.)

Table 3.1 lists the calculated values of the solvation energy changes of the MeIm and Cu_B complex upon protonation of these compounds. There are two contributions – the electronic reorganization energy contribution (due to solvent back reaction), which is called the 'strain' energy, and the Born solvation energy. The total solvation energy is the sum of these two contributions.

Table 3.2 lists the calculated pK_a values of His291 of the Cu_B complex together with individual contributions of electronic and solvation energy, and the corresponding data for the reference compound MeIm. Here, the electronic energy difference in gas phase between the deprotonated and protonated forms (ΔE_{elec}), a shift of the electronic gas phase energy relative to the MeIm model compound ($\Delta\Delta E_{elec}$), the difference in aqueous phase solvation energy between the two forms (ΔG_{solv}), a shift of the solvation energy relative to the model compound ($\Delta\Delta G_{solv}$) and total shift in kcal/mol and pK_a units are given separately to show their individual contributions to pK_a values.

The sidechain of an isolated histidine has two macroscopic pK_a values of 7.0 and 14.0, for the first and the second deprotonation. Compared with the methylimidazole, whose microscopic pK_a in water has been measured to be 6.6 (for the Nδ-site proton) [43], the equivalent histidine ligated to Cu_B metal center possesses pK_as of 8.6 in oxidized and 13.2 in reduced form, according to our present calculations.



Scheme 3.1. Deprotonation of the 4-Methylimidazole Molecule

3.5.3 The pK_a of His291 in *CcO*

Table 3.3 presents a summary of the results of the electrostatic solvation calculations for different dielectric models ((a), (b), (d)) applied in our study. The solvation energy of the Cu_B complex is calculated in a continuum low dielectric medium of $\varepsilon = 4$ (shown in standard letters), inhomogeneous dielectric of protein with the dielectric boundaries which reflects the real shape of the protein (shown in bold), and in continuum solvent of $\varepsilon = 80$ (aqueous phase, shown in italics). Here we consider only the effects of the dielectric medium on solvation energies, while the effects of the surrounding protein charges are not included yet.

The solvation in an inhomogeneous dielectric of the protein region is somewhat larger than in a uniform dielectric of $\varepsilon = 4$ but much smaller than in the aqueous phase. Also the 'strain' energy in the aqueous phase is larger than in a low dielectric medium because the high dielectric medium polarizes the charges of the QM system much stronger than the low dielectrics can. Since the reorganization of the charge densities due to the induced polarization from the medium produces an energetic cost on the electronic gas phase energies, the strain contributions always have a positive sign.

Table 3.4 gives results of electrostatic calculations for model (c) and (e). The solvation energies are compared in a uniform low dielectric of $\varepsilon = 4$ and in the 'real' protein. The surrounding protein charges are included here also. It is interesting to examine the differences in the reaction and protein field energies between the two models.

As expected, considering the Born energy, all four species are better solvated in an inhomogeneous protein dielectric, which includes high-ε water cavities, than in a continuum dielectric of low ε. Also, the protonated forms of each redox pair are better solvated than deprotonated ones. This is especially true for inhomogeneous protein environment, where, for instance, comparing their Born energies the oxidized protonated Cu_B complex (total charge $+2$) is much better solvated in the dielectric environment of *CcO* than the oxidized deprotonated form (total charge $+1$). One may also expect that the Born energy of the oxidized protonated Cu_B complex (total charge $+2$) should be the largest, the one of the reduced-deprotonated form (total charge 0) should be the smallest, while the oxidized deprotonated and reduced protonated forms (with total charges of $+1$) should have similar values. All discrepancies from this are due to the specific charge distributions in different species.

The effect of the interaction with the protein charges is much smaller in the 'real' protein than in a uniform ε of 4. A continuum low dielectric medium significantly destabilizes the deprotonated form of His291 in both redox states. In the 'real' protein there is a large bulb filled with water molecules around the position of the His291 site, which significantly screens the influence of the surrounding charges and therefore the deprotonated form is less destabilized

Table 3.3 Comparison of the solvation energy (kcal/mol) of the Cu_B complex, in oxidized or reduced state with His291 either in deprotonated or protonated form, between continuous aqueous phase ($\varepsilon = 80$), uniform low dielectic medium ($\varepsilon = 4$) and inhomogeneous dielectric of protein (represented by $\varepsilon = 4$ inside the protein and $\varepsilon = 80$ in water-filled cavities). Only the effect of the dielectric medium on solvation is considered, while the effect of the surrounding protein charges is neglected. For further comparison, the solvation energy contributions from deprotonated and protonated methylimidazole in aqueous solution are also given

Protonation form	Born energy	'Strain' energy	Solvation energy	Dielectric medium
$Cu_B^{2+}(H_2O)$ oxidized complex:				
$His291^-$	−55.0	+3.4	−51.6	
	−63.4	**+3.4**	**−60.0**	
	−78.3	*+10.1*	*−68.2*	
His291H	−121.1	+1.5	−119.6	
	−137.9	**+1.5**	**−136.4**	
	−165.0	*+3.4*	*−161.6*	
Cu_B^{2+} complex	+66.1	+1.9	+68.0	continuum ($\varepsilon = 4$)
Deprotonated–Protonated	**+74.5**	**+1.9**	**+76.4**	**inhomogeneity ($\varepsilon_1 = 4, \varepsilon_2 = 80$)**
	+86.7	*+6.7*	*+93.4*	*aqueous phase ($\varepsilon = 80$)*
$Cu_B^{1+}(H_2O)$ oxidized complex:				
$His291^-$	−33.3	+3.6	−29.7	
	−40.3	**+3.6**	**−36.7**	
	−50.7	*+9.5*	*−41.2*	
His291H	−41.1	+1.4	−39.7	
	−49.0	**+1.4**	**−47.6**	
	−59.2	*+3.3*	*−55.9*	
Cu_B^{1+} complex	+7.8	+2.2	+10.0	continuum ($\varepsilon = 4$)
Deprotonated–Protonated	**+8.7**	**+2.2**	**+10.9**	**inhomogeneity ($\varepsilon_1 = 4, \varepsilon_2 = 80$)**
	+8.5	*+6.2*	*+14.7*	*aqueous phase ($\varepsilon = 80$)*
4-methylimidazole:				
Deprotonated	*−12.2*	*+4.0*	*−8.2*	
Protonated	*−64.4*	*+1.5*	*−62.9*	
Deprotonated–Protonated	*+52.2*	*+2.5*	*+54.7*	*aqueous phase ($\varepsilon = 80$)*

by the interaction with protein charges (~+35 versus. +9 kcal/mol). This destabilization is mostly due to the adjacent negatively charged propionates of heme a_3 and since in both redox states, OORO and OORR, they are deprotonated, the effect of the protein charges, within the same model, becomes very similar (+34.1 versus. +35.8 or +8.0 versus. +9.6).

Table 3.5 is an overview showing the difference in electronic gas phase energies and the difference in solvation energies between protonated/deprotonated forms, and calculated pK_a values for the oxidized and reduced Cu_B complex obtained in different computational models applied in our study.

Table 3.4. Total Solvation (kcal/mol) of the Cu_B complex in a low dielectric continuum of $\varepsilon = 4$ and in the protein including the protein charges. The total solvation consists of three contributions: the Born-solvation energy, the interaction with protein charges (the so-called reaction and protein field), and 'strain' energy. The protonation state of the rest of the protein is the equilibrium protonation state for the corresponding redox state of enzyme, OORO or OORR

Compound	Born energy	Interaction with protein charges	'Strain' energy	Total solvation	Redox state [A-a-a_3-B][a] Dielectric medium
$Cu_B^{2+}(H_2O)$(His291$^-$)	*−55.0*	*−1.4*	*+3.4*	*−53.0*	**OORO**
	63.4	+18.6	+3.4	−41.4	
$Cu_B^{2+}(H_2O)$(His291H)	*−121.1*	*−35.5*	*+1.5*	*−155.1*	
	−137.9	+10.6	+1.5	−125.8	
Cu_B^{2+} complex	*+66.1*	*+34.1*	*+1.9*	*+102.1*	*continuum* $\varepsilon = 4$
Deprotonated–Protonated	+74.5	+8.0	+1.9	+84.4	protein
$Cu_B^{1+}(H_2O)$(His291$^-$)	*−33.3*	*+27.4*	*+3.6*	*−2.3*	**OORR**
	−40.3	+24.7	+3.6	−12.0	
$Cu_B^{1+}(H_2O)$(His291H)	*−41.1*	*−8.4*	*+1.4*	*−48.1*	
	−49.0	+15.1	+1.4	−32.5	
Cu_B^{1+} complex	*+7.8*	*+35.8*	*+2.2*	*+45.8*	*continuum* $\varepsilon = 4$
Deprotonated–Protonated	+8.7	+9.6	+2.2	+20.5	protein

[a]Redox state [A-a-a3-B] refers to the state of the four redox-active metal centers: Cu_A, heme *a*, heme a_3 and Cu_B complex, respectively.

The calculated aqueous phase pK_a values of His291 ligand of the Cu_B center are 8.6 and 13.2; therefore in aqueous phase at pH 7 this complex would be protonated in both redox states. Already in a low continuum dielectric without protein charges, the pK_a value of the His site becomes redox-dependent. The His-ligand would be distinctly deprotonated with pK_a = –9.9 in oxidized and distinctly protonated in reduced state (pK_a = +9.8). Introducing the dielectric inhomogeneities but without the protein charges, the pK_a value of oxidized Cu_B complex rises by about 6 pK_a units, however, the δ-His position remarkably is still deprotonated since its pK_a is equal to –3.7. There is not much change here in the pK_a of the reduced complex, such that the pK_a redox-dependence remains.

For a full description of the protein-solvent system, the equilibrium protein charge distribution for the given redox state of all four metal centers is also required. Assuming the reduced state of heme a_3, the pK_a of the His291 site in the oxidized Cu_B center of bovine *CcO* is between 2.1–4.0 depending on the redox state of other metal centers. When the binuclear center is in the two-electron reduced state, the His291 site has a pK_a above 17. In the fully reduced enzyme (FR = R_4) His291 possesses the highest pK_a of 19.7, while in the mixed-valence enzyme (MV = R_2) the pK_a of His291 is just slightly lower, i.e., 17.4. Thus, the addition of the protein charges to dielectric inhomogeneities further increases the apparent pK_a values of the His291-ligand by an additional 5.8 to 7.7 pK_a units

Table 3.5 The pK_a values of His291 residue as a part of oxidized or reduced Cu_B center in cytochrome *c* oxidase. The pK_a of His291 is calculated relative to the pK_a of the model compound−methylimidazole in aqueous phase. It is shown how different dielectric medium and interaction with protein charges affect the calculated pK_a of the His291 site. Energies are reported in kcal/mol

Redox state	ΔE_{elec}	$\Delta\Delta E_{elec}$	ΔG_{solv}	$\Delta\Delta G_{solv}$	$\Delta\Delta G_{shift}$	ΔpK_a	pK_a
In aqueous phase:							
4-methylimidazole	+234.6	0	+54.7	0	0	0.0	6.6
$Cu_B^{2+}(H_2O)$ oxidized	+198.7	−35.9	+93.4	+38.7	+2.8	+2.0	8.6
$Cu_B^{1+}(H_2O)$ reduced	+283.7	+49.1	+14.7	−40.0	+9.1	+6.6	13.2
In continuum dielectric of $\varepsilon = 4$ (no protein charges):							
$Cu_B^{2+}(H_2O)$ oxidized	+198.7	−35.9	+68.0	+13.3	−22.6	-16.5	-9.9
$Cu_B^{1+}(H_2O)$ reduced	+283.7	+49.1	+10.0	−44.7	+4.4	+3.2	9.8
In inhomogeneous dielectric of protein, $\varepsilon = 4$ and 80 (no protein charges):							
$Cu_B^{2+}(H_2O)$ oxidized	+198.7	−35.9	+76.4	+21.7	−14.2	-10.3	-3.7
$Cu_B^{1+}(H_2O)$ reduced	+283.7	+49.1	+10.9	−43.8	+5.3	+3.8	10.4
In continuum $\varepsilon = 4$ (including protein charges):							
OORO	+198.7	−35.9	+102.1	+47.4	+11.5	+8.4	15.0
OORR	+283.7	+49.1	+45.8	−8.9	+40.2	+29.3	35.9
In protein (including dielectric inhomogeneity and protein charges):							
OORO	+198.7	−35.9	+84.4	+29.7	−6.2	−4.5	2.1
RORO	+198.7	−35.9	+85.7	+31.0	−4.9	−3.5	3.1
ORRO	+198.7	−35.9	+86.8	+32.1	−3.8	−2.7	3.9
RRRO	+198.7	−35.9	+87.0	+32.3	−3.6	−2.6	4.0
OORR	+283.7	+49.1	+20.5	−34.2	+14.9	+10.8	17.4
RORR	+283.7	+49.1	+21.8	−32.9	+16.2	+11.8	18.4
ORRR	+283.7	+49.1	+22.7	−32.0	+17.1	+12.5	19.1
RRRR	+283.7	+49.1	+23.6	−31.1	+18.0	+13.1	19.7

for the oxidized Cu_B center and about 7 to 9.3 pK_a units in the corresponding reduced states of the enzyme.

It should be noted that by using model (c), which is the Cu_B complex embedded in a uniform low dielectric region that also includes the appropriate protein charges, the pK_a of the His291 site becomes much larger, 15 and 36 for the OORO and OORR redox states, respectively. Thus, adding the protein charges to the continuum low dielectric environment will produce a huge increase of the apparent pK_a values, by about 25 pK_a units. Compared with the more sophisticated model (e), one can conclude that the difference is caused by having the inhomogeneous system with the well-defined dielectric boundaries, which accounts for the shape of the protein molecule, including external solvent and introducing the water-filled cavities inside a protein where they are in reality located. It came out that the cavities inside the cytochrome *c* oxidase appear to play a very important role. Although the effect of the cavities is not so large for the Born solvation energies, the crucial role of the water filled cavities appears to be in the screening effect on the protein charges, which can significantly change the energetics of the protonation reactions in the protein.

Table 3.6 The pK_a values of Glu242 residue. The pK_a of Glu242 is calculated relative to the pK_a of the model compound – propionic acid in aqueous phase. It is shown how a different dielectric medium and interaction with protein charges affect the calculated pK_a of the Glu242 site. Energies are reported in kcal/mol

Redox state	ΔE_{elec}	$\Delta\Delta E_{elec}$	ΔG_{solv}	$\Delta\Delta G_{solv}$	$\Delta\Delta G_{shift}$	ΔpK_a	pK_a
In aqueous phase:							
propionic acid	+357.0[a]	0	−60.0	0	0	0.0	4.4
In continuum dielectric of $\varepsilon = 4$ (no protein charges):							
Glu242	+357.0	0	−46.1	+13.9	+13.9	+10.1	14.5
In inhomogeneous dielectric of protein, $\varepsilon = 4$ and 80 (no protein charges):							
Glu242	+357.0	0	−49.5	+10.5	+10.5	+7.6	12.0
In continuum $\varepsilon = 4$ (including protein charges):							
Glu242 in OORO	+357.0	0	−33.0	+27.0	+27.0	+19.6	24.0
Glu242 in OORR	+357.0	0	−32.9	+27.1	+27.1	+19.7	24.1
In protein (including dielectric inhomogeneity and protein charges):							
Glu242 in ORRO	+357.0	0	−49.9	+10.1	+10.1	+7.4	11.8
Glu242 in OORR	+357.0	0	−53.0	+7.0	+7.0	+5.1	9.5
Glu242 in OORO	+357.0	0	−53.1	+6.9	+6.9	+5.0	9.4
Glu242 in RRRO	+357.0	0	−50.1	+9.9	+9.9	+7.2	11.6
Glu242 in ORRR	+357.0	0	−49.9	+10.1	+10.1	+7.4	11.8
Glu242 in RRRR	+357.0	0	−49.5	+10.5	+10.5	+7.6	12.0
Glu242 in RORR	+357.0	0	−52.4	+7.6	+7.6	+5.5	9.9
Glu242 in RORO	+357.0	0	−52.4	+7.6	+7.6	+5.5	10.0
Glu242 in OOOR	+357.0	0	−53.0	+7.0	+7.0	+5.1	9.5

[a] The electronic energy difference was obtained with the B3LYP functional and the 6-311G** basis set which gives better agreement with experimentally measured proton affinities.

3.5.4 The dielectric constant of protein cavities

In the described calculations, the cavities of the interior of the protein were assumed to be filled with water, and were assigned a dielectric value of 80, as was done, e.g., in Refs. [35-37,40]. Indeed, a lot of internal water is observed in the crystal structure of *CcO*, [16,27,96]. There is a significant degree of ambiguity, however, in the dielectric constant of these water-filled cavities. Moreover, the protein dynamics, with water molecules moving in and out of the cavities, introduces an additional complication to the problem. A discussion of the appropriate value of the dielectric of the cavities is beyond the scope of this paper; we therefore explore here only the dependence of the main results and qualitative conclusions on the possible values of this important parameter of the model.

In order to clarify the importance of water-filled channels and cavities we examine how the pK_as of Glu242 and His291 depend on the dielectric constant associated with the protein cavities. Fig. 3.5 shows a dependence of the His291 pK_a on the dielectric constant of cavities. Here, the dependence is shown only for the OORO redox state of enzyme, which is the most critical for the whole model of proton pumping via the His291 as a proton-loading site. More detailed information on this dependence is presented in Table 3.7, where results are shown for the various ε values of the cavities.

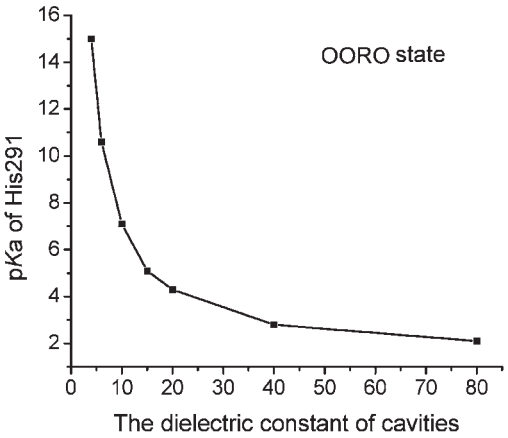


Fig. 3.5. The pK_a dependence of the His291 residue on the dielectric constant of water-filled cavities in the OORO redox state. The obtained dependence is affected by changes of the ΔG_{Born} and ΔG_{q} energy terms. In addition to the expected decrease of the ΔG_q term with an increase of the $\varepsilon_{\text{cavity}}$, also the Born energies of deprotonated and protonated species change in such a way that an increase of the ε cavity causes a larger stabilization of the protonated form (+2) compared with the deprotonated form (+1).

There is experimental evidence that water molecules inside *CcO* are quite mobile [19, 97, 98]. First of all, the enzyme produces the water molecules in the binuclear center and on a time scale of 10 ms they leave the active site and their appearance is noted near the position of Mg center (in EPR measurements using isotopic $^{17}O_2$) [99].Thus, the Cu_B center is located at the beginning of the so-called water exit channel(s). Second, it is believed that water molecules actively participate and facilitate a proton transfer through the enzyme [25, 26, 100], what is associated with the significant reorganization and reorientation of water molecule dipoles. Such large mobility of inner water molecules most likely is related to the higher value of dielectric constant.

Our general conclusion is that the pumping scheme of *CcO* proposed in this and in previous papers, is valid for a surprisingly wide range of possible values of the dielectric for the cavities, namely, for values greater than roughly 10, the general pumping scheme emerging from this work remains the same. This is an encouraging result, because the exact value of this parameter would be difficult to control by the enzyme.

3.6 CONCLUSIONS

We have calculated the pK_a values of His291 – the key element of the proposed pumping model of *CcO*. In a similar way the pK_a of another key group Glu242 has been estimated [30]. In our most recent calculations [101], using the same computational methods, we have also calculated the pK_a of the OH^- group in the binuclear center of the enzyme, and have shown that energetically it is a favorable process to protonate OH^-, i.e., to form water, at the expense of the deprotonation of His291. This is the key energy-generating step of the pump. Thus, the calculations provide the basis for, indeed helped to discover, the proposed pumping model of cytochrome oxidase. The work is summarized as follows.

The active site Cu_B complex was treated by density functional theory, while the effects of the protein charges and solvation were incorporated using the continuum electrostatic calculations. Our model does not account for any protein dynamic effects and deals only with a fixed crystal structure. The dielectric constants are chosen in a usual in such calculations empirical way, and therefore the obtained results should be considered in keeping with the limitations of the model in mind. The choice of the dielectric constants for a protein and the protein cavities that may contain internal water has been a topic of active debate for a very long time (see Refs [90,102-106], a discussion of a new procedure with optimized dielectric constants [107], and references cited therein). We chose the dielectric constants in the same way as it was done in recent similar DFT/electrostatic studies [35-37,40]. The good agreement between the calculated and measured pK_a of Glu242 residue, for instance, also gave us reason to believe that, although not perfect, our model still manages to account for most of the electronic and electrostatic energies of this complex system in a reasonably accurate way.

The pK_a values were calculated for several redox states of the enzyme, and the influence of different factors on the pK_a values was analyzed in detail. We were particularly interested in determining factors that would make it possible for this complex system to work as a proton pump. To accomplish this goal we considered several computational models and results of these calculations are briefly summarized here.

The dielectric inhomogeneities and water-filled cavities, model (d), increased the pK_as of His291 in both redox states (-3.7 and $+10.4$) and slightly decreased the pK_a of Glu242 to 12.0. However, qualitative conclusions here are the same – there is a redox-dependent acidity of the His291 site, but still Glu possesses a pK_a two units higher than His, and a hypothetical transfer of a pump proton from Glu242 to His291 would be energetically unfavorable.

Inclusion of the protein charges, which are found in a self-consistent manner, model (e), finely tunes the pK_a values to become suitable for proton pumping. Namely, the pK_as of His291 increase further by six to nine pK_a units, at the same time, the pK_a of Glu242 slightly decreases or remains the same depending on the redox state of the heme *a*, see Fig. 3.6. Table 3.7 compares the pK_as of these two residues for the different redox states of the enzyme. We find that for a given redox state of the binuclear complex, additional electrons on two other redox centers (Cu_A and heme *a*) only slightly modify the Glu242 and His291 pK_a values.

Based on the calculated pK_a values, the proton transfer from Glu242 to His291 can only occur after electron transfer to the binuclear center of the enzyme. For example, the reduction of the heme *a* increases the pK_a of Glu242 by about two units, making it more difficult to transfer a proton to His291 before an electron is transferred further to the catalytic center of the enzyme.

Comparing the effects of the addition of the protein charges to continuum and inhomogeneous dielectric environments, we concluded that the internal protein cavities, presumably filled with water, appear to play a very important role and have a huge influence on the pK_a values of the buried residues, with which they are in (direct) contact. The reason, obviously, is the high dielectric constant associated with the cavities, and their screening effect.

The effects of different factors on the pK_a values are summarized in Fig. 3.6. As shown in this figure, going from the aqueous phase to a low dielectric continuum, the pK_as

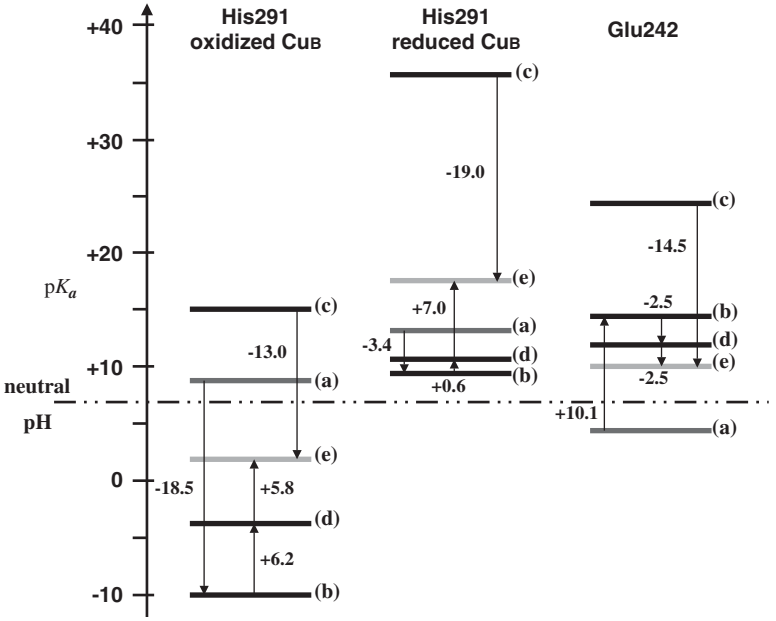


Fig. 3.6. A schematic of the pK_a values of His291 in the oxidized and reduced Cu_B complex and the pK_as of Glu242, for different models ((a)–(e)) applied in our study. The (a) designated bars refer to the pK_a values in the aqueous phase; the (e) designated bars represent the calculated pK_as in the protein (cytochrome oxidase). The arrows with the numbers represent the magnitude of the pK_a shifts caused by the different structural effects. See the text for the explanation (see Color Plate 12).

Table 3.7 The pK_as of Glu242 and the His291 site in the different redox states of cytochrome *c* oxidase. R stands for the reduced and O for the oxidized states of redox cofactors of *CcO* shown in Fig. 3.1. The pK_as are compared for the different dielectric constants of water filled cavities (ε = 15, 20 or 80). In all calculations, the dielectric constants of QM system and protein part are set to 1 and 4, respectively

Redox state	Glu242			His291		
ε_{cavity}	15	20	80	15	20	80
OORR	12.9	11.9	9.5	22.6	21.1	17.4
OORO	13.0	12.0	9.4	5.1	4.3	2.1
RORO	13.8	12.6	10.0	6.5	5.4	3.1
ORRO	15.9	14.7	11.8	7.4	6.3	3.9
RRRO	15.8	14.6	11.6	7.7	6.5	4.0
RORR	13.7	12.6	9.9	23.8	22.3	18.4
ORRR	15.9	14.7	11.8	24.6	23.1	19.1
RRRR	16.3	15.1	12.0	25.5	23.9	19.7

change by –18.8, –3.4 and +10.1, for His-oxidized-Cu_B, His-reduced-Cu_B and Glu-site, respectively. Introducing dielectric inhomogeneities, the Born solvation energy changes so as to shift the pK_a by +6.2, +0.6 and –2.5 units, respectively. Addition of the protein charges causes the shifts of +5.8, +7.0 and –2.5 pK_a units, respectively. Finally, the

screening effects of the protein inner water cavities and external high dielectric solvent are estimated to be –13.0, –19.0 and –14.5 units, respectively.

It has been shown that pK_a of His 291 depends very strongly on the redox state of the Cu_B center, while a much weaker dependence is observed for Glu242, and its pK_a values remain relatively high for all redox states of the enzyme (see Table 3.7). This is not surprising, given that His291 is much closer located to a redox center than Glu242. Also, since in the reduced state of Cu_B His291 is protonated, while in the oxidized state His291 is deprotonated, the total charge of Cu_B *together* with its ligand His291 does not change, and hence the redox state of the Cu_B center has much less effect on the pK_a of a distant Glu242 residue. We find rather that the pK_a of Glu242 is coupled to the redox state of the adjacent heme *a*. This dependence is less than three pK_a units, yet can be important. Each time Glu242 donates a proton, it needs to get reprotonated, hence the obtained high pK_a values of Glu242 totally make sense.

Summarizing, the obtained results support our proposal, Ref. [24], that the His291 ligand of a Cu_B center works as a proton pump element in *CcO*. This prediction is currently the subject of several experimental investigations which aim to probe the role of His291 in *CcO*.

3.7 ACKNOWLEDGEMENTS

We would like to thank Xuehe Zheng for his helpful input. This work has been supported by the NSF grant and a research grant from the NIH (GM54052).

3.8 REFERENCES

1 P. Mitchell, Nature, 191 (1961) 144.
2 P.J. Mitchell, Theor. Biol., 62 (1976) 327.
3 M. Wikström, Curr. Op. Struct. Biol., 8 (1998) 480.
4 R.B. Gennis, Proc. Natl. Acad. Sci. USA, 95 (1998) 12747.
5 H. Michel, J. Behr, A. Harrenga and A. Kannt, Annu. Rev. Biopys. Biomol. Struct., 27 (1998) 329.
6 G.T. Babcock and M. Wikström, Nature, 356 (1992) 301.
7 S. Ferguson-Miller and G.T. Babcock, Chem. Rev., 7 (1996) 2889.
8 H. Michel, Biochemistry 38 (1999) 15129.
9 M. Wikström, Biochemistry, 39 (2000) 3515.
10 D. Zaslavsky and R.B. Gennis, Biochim. Biophys. Acta, 1458 (2000) 164.
11 M. Wikström, A. Jasaitis, C. Backgren, A. Puustinen and M.I. Verkhovsky, Biochim. Biophys. Acta, 1459 (2000) 514.
12 S. Han, S. Takahashi and D.L Rousseau, J. Biol. Chem., 275 (2000) 1910.
13 A.A. Stuchebrukhov, J. Theor. Comp. Chem., 2 (2003) 91.
14 A.A. Stuchebrukhov, Theor. Chem. Acc., 110 (2003) 291.
15 S. Yoshikawa, K. Shinzawa-Itoh, R. Nakashima, R. Yaono, E. Yamashita, N. Inoue, M. Yao, M.J. Fei, C.P. Libeu, T. Mizushima, H. Yamaguchi, T. Tomizaki and T. Tsukihara, Science, 280 (1998) 1723.
16 C. Ostermeier, A. Harrenga, U. Ermler and H. Michel, Proc. Natl. Acad. Sci. USA, 94 (1997) 10547.
17 G. Buse, T. Soulimane, M. Dewor, H.E. Meyer and M. Bluggel, Protein Sci., 8 (1999) 985.
18 S. Iwata, C. Ostermeier, B. Ludwig and H. Michel, Nature, 376 (1995) 660.
19 T. Tsukihara, H. Aoyama, E. Yamashita, T. Tomizaki, H. Yamaguchi, K. Shinzawa-Itoh, R. Nakashima, R. Yaono and S. Yoshikawa, Science, 272 (1996) 1136.
20 T. Haltia, M. Saraste and M. Wiksrom, EMBO J., 10 (1989) 2015.

21 R.W. Hendler, K. Pardhasaradhi, B. Reynafarje and B. Ludwig, Biophys. J., 60 (1991) 415.
22 M.R. Bratton, M.A. Pressler and J.P. Hosler, Biochemistry, 38 (1999) 16236.
23 R.B. Gennis, Science, 280 (1998) 1712.
24 D.M. Popovic and A.A. Stuchebrukhov, FEBS Letters, 566 (2004) 126.
25 X. Zheng, D.M. Medvedev, J. Swanson and A.A. Stuchebrukhov, Biochim. Biophys. Acta, 1557 (2003) 99.
26 M. Wikström, M.I. Verkhovsky and G. Hummer, Biochim. Biophys. Acta, 1604 (2003) 61.
27 M. Svensson-Ek, J. Abramson, G. Larsson, S. Tornroth, P. Brzezinski and S. Iwata, J. Mol. Biol., 321 (2002) 329.
28 D.M. Popovic and A.A. Stuchebrukhov, J. Am. Chem. Soc., 126 (2004) 1858.
29 M. Tashiro and A.A. Stuchebrukhov, J. Phys. Chem. B, 109 (2005) 1015.
30 D.M. Popovic, J. Quenneville and A.A. Stuchebrukhov, J. Phys. Chem. B, 109 (2005) 3616.
31 J. Quenneville, D.M. Popovic and A.A. Stuchebrukhov, J. Phys. Chem. B, 108 (2004) 18383.
32 D.M. Popovic and A.A. Stuchebrukhov, J. Phys. Chem. B, 109 (2005) 1999.
33 J.M. Mouesca, J.L. Chen, L. Noodleman, D. Bashford and D. A. Case, J. Am. Chem. Soc., 116 (1994) 11898.
34 J. Li, C.L. Fischer, J.L. Chen, D. Bashford and L. Noodleman, J. Phys. Chem., 96 (1996) 2855.
35 J. Li, M.R. Nelson, C.Y. Peng, D. Bashford and L. Noodleman, J. Phys. Chem. A, 102 (1998) 6311.
36 J. Li, C.L. Fisher, R. Konecny, D. Bashford and L. Noodleman, Inorg. Chem., 38 (1999) 929.
37 G.M. Ullmann, L. Noodleman and D.A. Case, J. Biol. Inorg. Chem., 7 (2002) 632.
38 J.L. Chen, L. Noodleman, D.A. Case and D.J. Bashford, Phys. Chem., 98 (1994) 11059.
39 B. Kallies and R. Mitzner, J. Phys. Chem. B, 101 (1997) 2959.
40 W.H. Richardson, C. Peng, D. Bashford, L. Noodleman and D.A. Case, Int. J. Quant. Chem., 61 (1997) 207.
41 H. Reiss and A. Heller, J. Phys. Chem., 89 (1985) 4207–4213.
42 T.L. Hill, An introduction to statistical thermodynamics, Dover, New York, 1860.
43 M. Tanokura, Biochem. Biophys. Acta, 742 (1983) 576.
44 J.E.D. Bene, H.D. Metter, M.J. Frisch, B.J. Luke and J.A. Pople, J. Phys. Chem., 87 (1983) 3279.
45 J. Catalan, J.L.G. d. Paz, M. Yanez, R.M. Claramunt, C. Lopez, J. Elguero, F. Anvia, J.H. Quian, M. Taagepera and R.W. Taft, J. Am. Chem. Soc., 112 (1990) 1303.
46 P. Hohenberg and W. Kohn, Physical Review, 136 (1964) B864.
47 K. Kohn and L.J. Sham, Physical Review, 140 (1965) A1133.
48 Jaguar 5.5, S., L.L.C., Portland, OR, 1991-2003. ed.
49 A.D. Becke, J. Chem. Phys., 98 (1993) 5648.
50 P.J. Hay and W.R. Wadt, J. Chem. Phys., 82 (1985) 299.
51 R. Ditchfield, W.J. Hehre and J.A. Pople, J. Chem. Phys., 54 (1971) 724.
52 W.J. Hehre and J.A. Pople, J. Chem. Phys., 56 (1972) 4233.
53 W.J. Hehre, R. Ditchfield and J.A. Pople, J. Chem. Phys., 56 (1972) 2257.
54 J.S. Binkley and J.A. Pople, J. Chem. Phys., 66 (1977) 879.
55 P.C. Hariharan and J.A. Pople, Theor. Chim. Acta, 28 (1973) 213.
56 M.M. Francl, W.J. Pietro, W.J. Hehre, J.S. Binkley, M.S. Gordon, D.J. DeFrees and J.A. Pople, J. Chem. Phys., 77 (1982) 3654.
57 T. Clark, J. Chandrasekhar, G.W. Spitznagel and P.v.R. Schleyer, J. Comput. Chem., 4 (1983) 294.
58 M.J. Frisch, J.A. Pople and J.S. Binkley, J. Chem. Phys., 80 (1984) 3265.
59 R. Krishnan, J.S. Binkley, R. Seeger and J.A. Pople, J. Chem. Phys., 72 (1980) 650.
60 A.D. McLean and S.G. Chandler, J. Chem. Phys., 72 (1980) 5639.
61 M.R.A. Blomberg, P.E.M. Siegbahn, G.T. Babcock and M. Wikström, J. Am. Chem. Soc., 122 (2000) 12848.
62 P.E.M. Siegbahn, M.R.A. Blomberg and M.L. Blomberg, J. Phys. Chem. B, 107 (2003) 10946.
63 D.J. Tannor, B. Marten, R. Marphy, R.A. Friesner, D. Sitkoff, A. Nicholls, M.G.W.A.I. Ringnalda and B. Honig, J. Am. Chem. Soc., 116 (1994) 11875.
64 B. Marten, K. Kim, C. Cortis, R.A. Friesner, R.B. Murphy, M.N. Ringnalda, D. Sitkoff and B. Honig, J. Phys. Chem., 100 (1996) 11775.
65 C.M. Breneman and K.B. Wiberg, J. Comput. Chem., 11 (1990) 361.
66 A. Kannt, R.D. Lancaster and H. Michel, Biophys. J., 74 (1998) 708.
67 M.R. Gunner, A. Nicholls and B. Honig, J. Phys. Chem., 100 (1996) 4277.
68 C.R. Lancaster, H. Michel, B. Honig and M.R. Gunner, Biophys. J., 70 (1996) 2469.

69 B. Rabenstein, G.M. Ullmann and E.W. Knapp, Biochemistry, 37 (1998) 2488.
70 R.V. Sampogna and B. Honig, Biophys. J., 66 (1994) 1341.
71 V.Z. Spassov, H. Luecke, K. Gerwert and D. Bashford, J. Mol. Biol., 312(2001) 203.
72 D.M. Popovic, S.D. Zaric, B. Rabenstein and E.W. Knapp, J. Am. Chem. Soc., 123 (2001) 6040.
73 D. Bashford and M. Karplus. Biochemistry, 29 (1990) 10219.
74 A. Nicholls and B. Honig, J. Comp. Chem., 12 (1991) 435.
75 D. Bashford and M. Karplus, J. Phys. Chem., 95 (1991) 9557.
76 P. Beroza, D.R. Fredkin, M.Y. Okamura and G. Feher, Proc. Natl. Acad. Sci. USA, 88 (1991) 5804.
77 M.R. Gunner and B. Honig, Proc. Natl. Acad. Sci. USA, 88 (1991) 9151.
78 A.-S. Yang, M.R. Gunner, R. Sompogna and B. Honig, Prot. Struct. Funct. Genet., 15 (1993) 252.
79 P. Beroza and D.A. Case, J. Phys. Chem., 100 (1996) 20156.
80 A.M. Baptista, P.J. Martel and C.M. Soares, Biophys. J., 76 (1999) 2978.
81 D.M. Popovic, A. Zmiric, S.D. Zaric and E.W. Knapp, J. Am. Chem. Soc., 124 (2002) 3775.
82 D. Bashford and K. Gerwert, J. Mol. Biol., 224 (1992) 473.
83 D. Bashford, An object-oriented programming suite for electrostatic effects in biological molecules, in; Y. Ishikawa, R.R. Oldehoeft, J.V.W. Reynders and M. Tholburn, (Eds.) scientific computing in object-oriented parallel environments, Springer, Berlin, 1997, Vol. 1343, p. 233.
84 B. Rabenstein, Karlsberg online manual, http://lie.chemie.fu-berlin/karlsberg/ 1999.
85 S. Yoshikawa, personal comunication.
86 W.L. Jorgenson, J. Chandrasekhar and J.D. Madura, J. Chem. Phys., 79 (1983) 926.
87 K. Sharp and B. Honig, Ann. Rev. Biophys. Biophys. Chem., 19 (1990) 301.
88 T. Simonson and D. Perahia, Proc. Natl. Acad. Sci. USA, 92 (1995) 1082.
89 T. Simonson and C.L. Brooks, J. Am. Chem. Soc., 118 (1996) 8452.
90 A. Warshel and A. Papazyan, Curr. Op. Struct. Biol., 8 (1998) 211.
91 J.A.D. MacKerell, D. Bashford, M. Bellot, J. R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodholm, I.W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus, J. Phys. Chem., 102 (1998) 3586.
92 P. Vagedes, B. Rabenstein, J. Åqvist, J. Marelius and E.W. Knapp, J. Am. Chem. Soc., 122 (2000) 12254.
93 B. Rabenstein and E.W. Knapp, Biophys. J., 80 (2001) 1141.
94 I. Klapper, R. Fine, K.A. Sharp and B.H. Honig, Prot. Struct. Funct. Genet., 1 (1986) 47.
95 A. Onufriev, A. Smondyrev and D. Bashford, J. Mol. Biol., 332 (2003) 1183.
96 T. Tsukihara, K. Shimokata, Y. Katayama, H. Shimada, K. Muramoto, H. Aoyama, M. Mochizuki, K. Shinzawa-Itoh, E. Yamashita, M. Yao, Y. Ishimura and S. Yoshikawa, Proc. Natl. Acad. Sci. USA, 100 (2003) 15304.
97 A.E. Garcia and G. Hummer, Proteins, 38 (2000) 261–272.
98 C. Backgren, G. Hummer, M. Wikström and A. Puustinen, Biochemistry, 39 (2000) 7863–7867.
99 B. Schmidt, J. McCracken and S. Ferguson-Miller, Proc. Natl. Acad. Sci. USA, 100 (2003) 15539–15542.
100 E. Olkhova, M.C. Hutter, M.A. Lill, V. Helms and H. Michel, Biophys. J., 86 (2004) 1873–1889.
101 D.M. Popovic, J. Quenneville and A.A. Stuchebrukhov, Biochim. Biophys. Acta Bioenergetics, (2006) in press.
102 A. Warshel and S.T. Russel, Quart. Rev. Biophys., 17 (1984) 283.
103 M.K. Gilson and B. Honig, Biopolymers, 25 (1986) 2097.
104 S.C. Harvey, Proteins, 5 (1989) 78.
105 A. Warshel and J. Aqvist, Ann. Rev. Biophys. Biophys. Chem., 20 (1991) 267.
106 M.K. Gilson, Curr. Opin. Struct. Biol., 5 (1995) 216.
107 M. Wojciechowski, T. Grycuk, J.M. Antosiewicz and B. Lesyng, Biophys. J., 84 (2003) 750.
108 P.R. Rich, I.C. West and P. Mitchell, FEBS Lett., 233 (1988) 25.

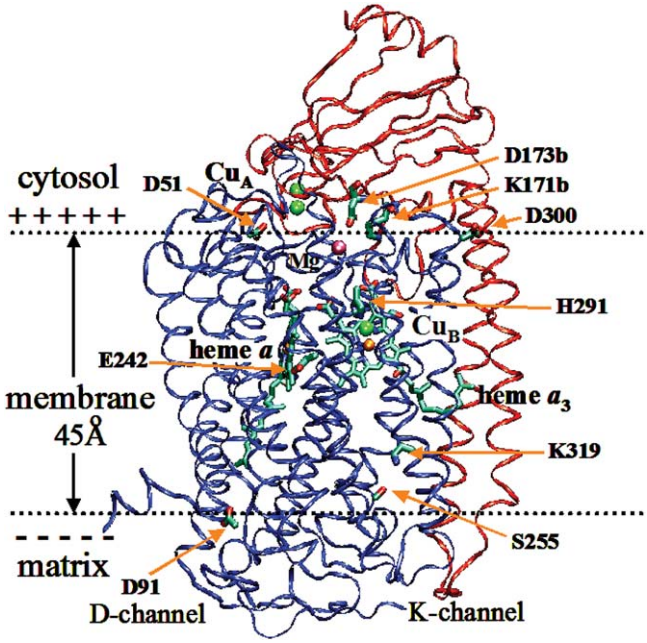


Plate 8. The structure of the core catalytic subunits of cytochrome *c* oxidase from bovine heart [15] The subunit A (in blue color), together with a helical extension of subunit B (in red), is almost entirely integrated in the membrane, while the globular water-soluble domain of subunit B is exposed to solvent on the periplasmic side of the membrane. The two hemes and the Cu_B center (in green) are located in the helical part of subunit *A* that is embedded in a membrane of 45 Å thickness [23], while the Cu_A center consisting of two copper atoms is located in the subunit *B*, at the interface between the two subunits [108]. Several residues in the proton input D- and K-channels and on the output side, important for a proton translocation, are also displayed In the figure (see text for details) (illustration appears on page 54 of this volume).

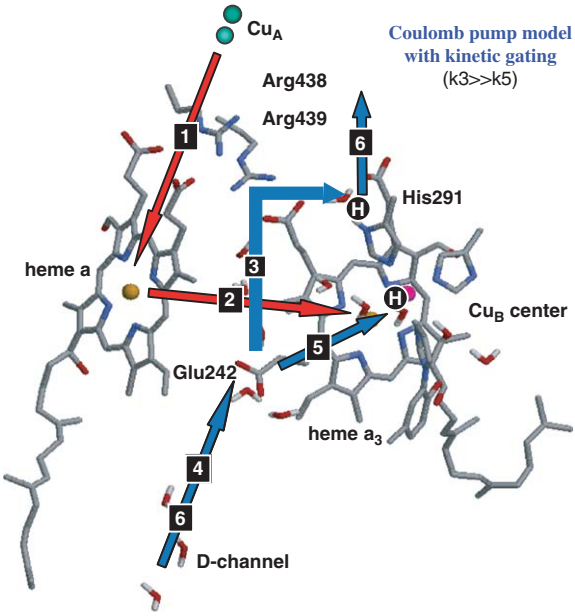


Plate 9. The key structural elements of the proposed pumping mechanism of cytochrome *c* oxidase [24], and the sequence of transitions during one pumping cycle. The relative position of the four metal redox-active centers (Cu_A, heme *a*, heme a_3 and Cu_B complex) and the three protonatable sites (Glu242, His291 and OH^- ligand in the BNC) are shown. According to the model, a stepwise electron transfer that brings an electron from Cu_A to the binuclear center (consisting of heme a_3 and the Cu_B complex) is coupled to proton translocation from Glu242 to the His291 group, which serves here as a proton-loading site of the pumping mechanism in *CcO*. Two protonation sites, PLS and one in the BNC, are shown as H-circles. The PT and ET steps are shown by blue and red arrows, respectively. The key assumption of the model is that the proton transfer rate upon ET between the two hemes (step 2) is much higher along the pumping channel 3 than along the chemical channel 5 ($k_3 >> k_5$), in other words step 3 occurs before step 5 (illustration appears on page 57 of this volume).

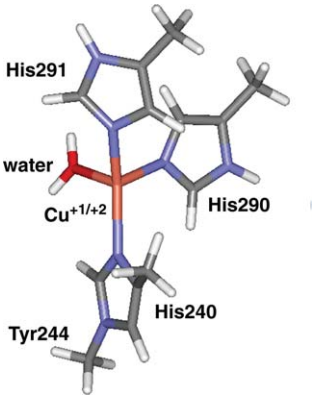


Plate 10. The model of the CuB center used in the pKa calculation of His291 (illustration appears on page 62 of this volume).

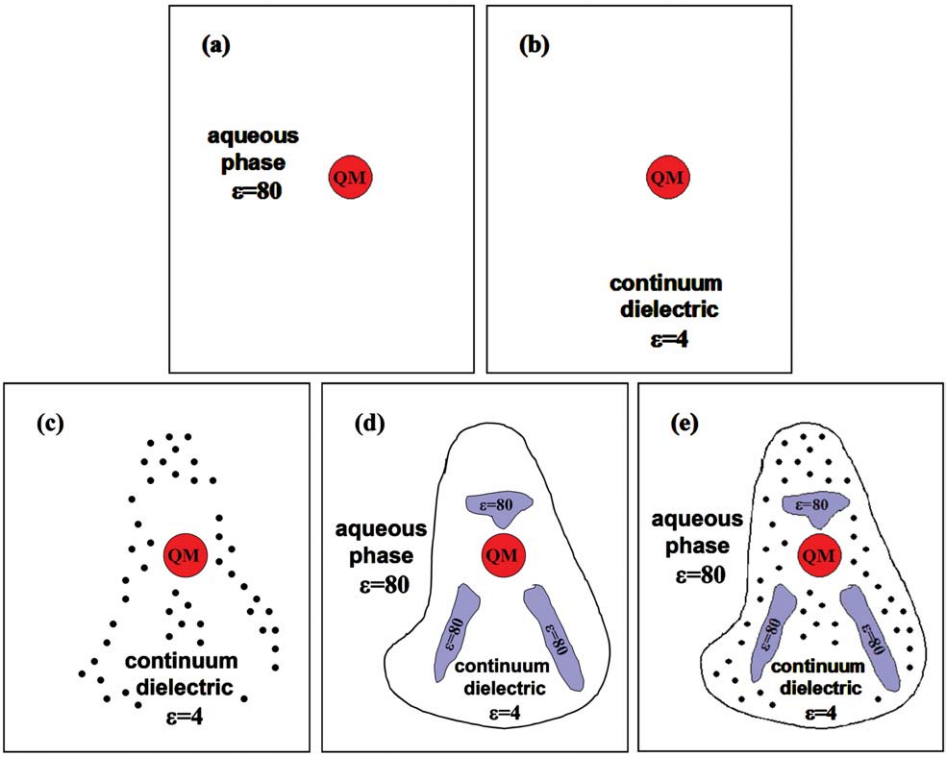


Plate 11. Schematics of different models used in this study. In a model (a) the QM system that consists of a CuB center and coordinated ligands (with dielectric constant e 5 1 inside) is solvated in the aqueous phase. (b) The QM system is solvated in a low dielectric of e 5 4. (c) The QM system is embedded in a continuum dielectric of e 5 4 and surrounded with protein charges. (d) The QM system is embedded in a inhomogeneous dielectric medium represented by e 5 4 within the borders of the protein, and e 5 80 inside the inner water filled cavities and outside the protein, while protein charges are neglected. A model (e) includes a full protein-solvent description including dielectric inhomogeneities and protein charges, whereas the QM system is treated by the DFT method. Crystalographically, more than 150 water molecules have been found in the inner protein cavities [16,27,96]. Only the most important water filled cavities are schematically shown in the figure. They represent the proton input D- and K-channels leading to the active site and the large water filled cavity on the proton output side of the protein (illustration appears on page 66 of this volume).

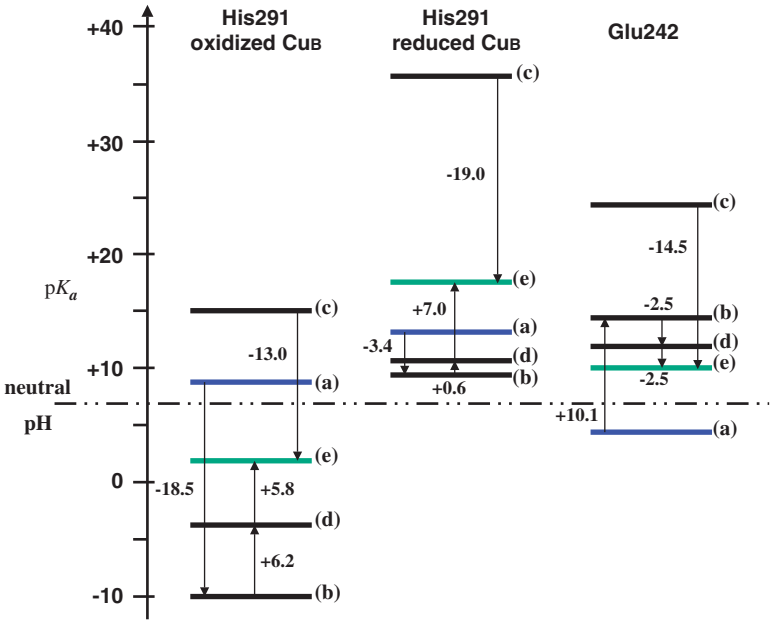


Plate 12. A schematic of the p*K*a values of His291 in the oxidized and reduced Cu_B complex and the pK_as of Glu242, for different models ((a)–(e)) applied in our study. Blue bars are related to the pK_a values in the aqueous phase and green bars represent the calculated pK_as in the protein (cytochrome oxidase). The arrows with the numbers represent the magnitude of the pK_a shifts caused by the different structural effects. See the text for the explanation (illustration appears on page 75 of this volume).

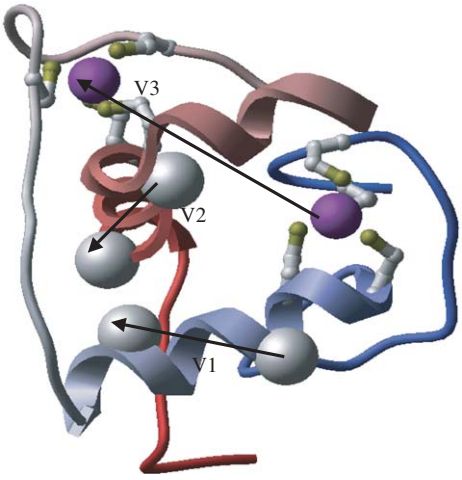


Plate 13. Schematic representation of the crystal GR DBD second monomer (pdb accession code 1 glu). The thiolate ligands to the zinc atoms are shown (hydrogens are omitted), and the arrows display the three unit vectors (*V*1, *V*2 and *V*3) used for calculating the re-orientational times. *V*1 and *V*2 are defined through the C_α atoms (light spheres) of Val 462 to Ala 467 (*V*1), Arg 496 to Leu 501 (*V*2) and *V*3 through the two zinc ions (dark spheres) (illustration appears on page 127 of this volume).

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 4

Watson–Crick hydrogen bonds: nature and role in DNA replication

Célia Fonseca Guerra and F. Matthias Bickelhaupt

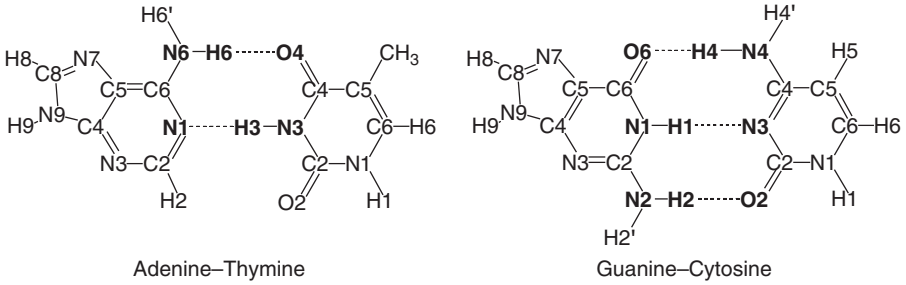
Afdeling Theoretische Chemie, Scheikundig Laboratorium der Vrije Universiteit, De Boelelaan 1083, NL-1081 HV Amsterdam, The Netherlands

Abstract

The hydrogen bonds in DNA Watson–Crick base pairs have long been considered predominantly electrostatic phenomena. In this chapter, we show with state-of-the-art calculations that this is not true and that electrostatic interactions and covalent contributions in these hydrogen bonds are in fact of the same order of magnitude. We discuss the steric and electronic effects of the backbone and the influence of the molecular environment on hydrogen bonding in Watson–Crick pairs. Furthermore, we address the question of how far these hydrogen bonds can be considered resonance-assisted hydrogen bonds (RAHB). Finally, the implications of our findings for our understanding of the highly accurate process of DNA replication are discussed.

4.1 INTRODUCTION

The genetic information of living organisms resides in deoxyribonucleic acid (DNA), a macromolecule that consists of two helical strands of nucleotides that are twisted around each other thus rendering the well-known double helix [1]. The genetic information is encoded by means of the order in which the four nucleic acids adenine (A), thymine (T), guanine (G) and cytosine (C) occur in a DNA strand. The two strands of the DNA double helix are held together by the hydrogen bonds of the Watson–Crick base pairs that form selectively between A and T, and between G and C. The strength of these hydrogen bonds, according to gas-phase experiments, is –12.1 and –21.0 kcal/mol for AT and GC, respectively (bond enthalpies at 298 K) [2]. This is an important contribution to the intrinsic stability of the DNA molecule. Other factors that play a role are, for example, the solvent effects that occur under the physiological conditions in cells and the aromatic stacking interactions between the bases within a strand (Scheme 4.1).



Scheme 4.1

In addition to rendering stability to the carrier of genetic information, the cohesion of the two DNA strands through Watson–Crick pairs also has another important consequence: it causes the two strands to be complementary copies of each other in terms of the order of the composing nucleotides. This principle is the basis of enzyme-catalyzed DNA replication and transcription, two chemical reactions that play a key role in the highly complex biological processes of cell growth and gene expression. In these reactions, the strands of the double helix dissociate, and the genetic information is copied by using one of them as a template for making a new complementary strand. In the standard model, the immensely high accuracy with which this process occurs is ascribed to the hydrogen bonds between the complementary bases [1]. Recently, Kool *et al*. [3] found experimental evidence for shape complementarity contributing to the accuracy of the replication. Also, other factors [4] such as solvation, hydrophobicity and stacking can contribute to the process of nucleotide insertion.

In this chapter, the role of hydrogen bonding and sterical fit in the event of molecular recognition between the template and the incoming DNA base in the active site of polymerase is evaluated with state-of-the-art density functional theory (DFT) studies that were carried out with the Amsterdam Density Functional (ADF) program [5]. First, we validate the accuracy of DFT for hydrogen bonds by comparing the performance of various density functionals (and basis sets) for the water dimer as well as the Watson–Crick pairs AT and GC with *ab initio* Hartree–Fock, MP2 and CCSD(T) calculations and available experimental data. Here, we anticipate that BP86/TZ2P evolves as an excellent DFT approach for studying biological molecules involving hydrogen bonding. Then, the role of charge transfer or donor–acceptor interactions in the hydrogen bonds of the Watson–Crick base pairs is investigated through an analysis of the electronic structure, the electronic density and a quantitative decomposition of the bond energy into the electrostatic interaction, the repulsive orbital interactions (Pauli repulsion between occupied orbitals) and the bonding orbital interactions (charge transfer and polarization), all within the conceptual framework provided by Kohn–Sham molecular (KS-MO) theory [6]. The influence of solvent molecules, counter ions and the backbone of the structure of these hydrogen bonds is examined and it is shown how this leads to the solution of a, for a long time, unresolved discrepancy between experiment and theory [7,8]. Furthermore, we investigate how and to what extent the π-electron system contributes to the strength of the hydrogen bonds in Watson–Crick pairs, in order to provide a more quantitative basis for their classification as resonance assisted hydrogen bonds as proposed by Gilli *et al.* [9].

Finally, based on analyses involving 2,4-difluorotoluene and toluene as mimics of thymine, we propose a synthesis of the standard model (that ascribes the high accuracy of the replication process to hydrogen bonding) [1] and Kool's steric model [3].

4.2 ACCURACY OF DFT FOR HYDROGEN BONDS: WATER DIMER

Before results on DNA base pairs are presented, we need a validation of DFT for hydrogen bonds. We have compared for the water dimer H_2O•HOH our BP86/TZ2P approach with various high-level *ab initio* benchmark computations. Our BSSE-corrected hydrogen bond energy of – 4.3 kcal/mol agrees within a few tenths of a kcal/mol with values obtained at MP2/aug-cc-pVQZ (– 4.9 kcal/mol) [10] and CCSD(T)/aug-cc-pVQZ (–5.1 kcal/mol; with counterpoise correction: – 4.9 kcal/mol) [11]. More importantly, our value of 2.90 Å for the **O–O** distance matches those from MP2/aug-cc-pVQZ (2.90 Å; with counterpoise correction: 2.92 Å) [10] and CCSD(T) computations (2.90 Å [11]; with counterpoise correction [12]: 2.93 Å) either exactly or within three hundredths of an Å. In Section 4.3, it is demonstrated that the same accuracy can be obtained for the Watson–Crick base pairs AT and GC.

4.3 STRUCTURE OF HYDROGEN BONDS OF DNA

4.3.1 Watson–Crick base pairs

After having established that DFT performs adequately for simple hydrogen bonds, like the one that occurs in the water dimer, we acquired similar results for the bond energy of the Watson-Crick base pairs. First of all, our computed bond enthalpies –13.0 kcal/mol for AT and –26.1 kcal/mol for GC agree excellently with the experimental values of –11.8 and –23.8 kcal/mol for AT and GC, respectively, but also with theoretical values of others. For example, the corresponding values obtained by Bertran *et al*. [13] at B3LYP are –10.9 and –24.0 kcal/mol and those obtained by Sponer *et al*. [14] at MP2 are –9.5 and –20.8 kcal/mol.

However, no such beautiful agreement with experimental values was achieved when we computed the hydrogen bond distances for the Watson–Crick pairs. We arrived at striking discrepancies with experimental structures. At BP86/TZ2P, we found N6**–**O4 and N1**–**N3 hydrogen bond distances in AT of 2.85Å and 2.81Å, whereas the experimental values are 2.95Å and 2.82Å [15]. Even more eye-catching, as can be seen in Fig. 4.1, was the situation for the three hydrogen bonds in GC, i.e., O6**–**N4, N1**–**N3 and N2**–**O2, for which we found a bond length pattern that is short–long–long, i.e., 2.73, 2.88 and 2.87Å at BP86/TZ2P, at significant variance with the experimental values which are long–long–short (2.91, 2.95 and 2.86Å) [16]. The disagreement between theoretical and experimental hydrogen bond distances in the Watson–Crick pairs was not new. It had been encountered before in several DFT and *ab initio* studies [13,14]. Therefore we started an investigation of the effect of the backbone and the environment on the hydrogen bonds in AT and GC.

It is important to realize that the experimental structures stem from X-ray diffraction measurements on crystals of sodium adenylyl-3′,5′-uridine hexahydrate (1) [15] for AT

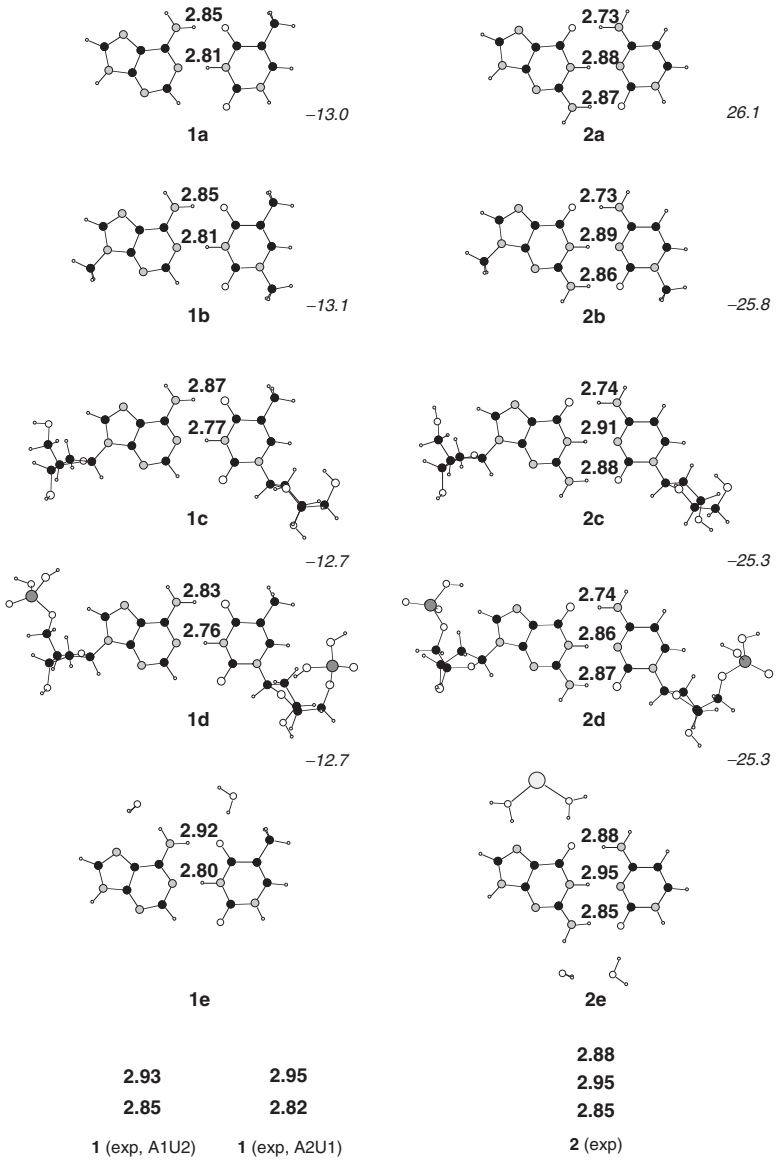


Fig. 4.1. Bond Energies (in kcal/mol) and N6–O4 and N1–N3 distances (in Å) in AT (**1a**), methylated AT (**1b**), AT with deoxyribose residues (**1c**), AT with deoxyribose 5′-phosphate residues (**1d**) and a AT crystal model system (**1e**). Bond Energies and O6–N4, N1–N3 and N2–O2 distances in GC (**2a**), methylated GC (**2b**), GC with deoxyribose residues (**2c**), GC with deoxyribose 5′-phosphate residues (**2d**) and a GC crystal model systems (**2e**). Geometries from BP86/TZ2P (**1a**–**1c**, **1e**, **2a**–**2c**, **2e**) and BP86/DZP (**1d**, **2d**) computations. Geometries of **1c**–**1d** and **2c**–**2d** were optimized without any symmetry constraint whereas for the other systems (**1a**, **1b**, **1e**, **2a**, **2b** and **2e**) C_s symmetry has been used. All energies have been calculated at the BP86/TZ2P level. N6–O4 and N1–N3 distances from the X-ray crystal structure of sodium adenylyl-3′,5′-uridine hexahydrate (**1**, see also Scheme 4.2) [11] and from the X-ray crystal structure of sodium guanylyl-3′,5′-cytidine nonahydrate (**2**) [12]. Note that there are two experimental values for both N6–O4 and N1–N3 because the two AU pairs in the crystal of **1** experience different environments.

(or AU) and sodium guanylyl-3′,5′-cytidine nonahydrate (2) [16] for GC. The base pairs in these crystals differ from the theoretical model systems studied so far, in two important fashions: (i) they are part of a small double helix consisting of two base pairs in which bases along a strand are connected via a ribose–phosphate–ribose backbone, and (ii) they experience interactions with the environment in the crystal, in particular water molecules, ribose OH groups and counter ions. In view of the very shallow potential energy surfaces that we found for Watson–Crick base pairing, it seems plausible that the effects of the backbone and the molecular environment in the crystal could cause the discrepancy with more simplistic AT and GC models. This has led us to study the effect of the backbone and the molecular environment at the BP86/TZ2P level, the level of theory which yields excellent results for the water dimer (see Section 4.2 and Ref. [8]).

4.3.2 Effect of the backbone

The effect of the backbone on Watson–Crick base pairing appears to be marginal. It was studied by stepwise going from Watson–Crick pairs of plain nucleic bases via nucleotides, which enabled us to study the electronic effect alone, to strands consisting of two nucleotides. These strands also cover the sterical strain, which the hydrogen bonds experience from the backbone in DNA. Figure 4.1 gives the geometries and energies (**1a**–**1d**, **1f**, **2a**–**2d**). In the first place, comparison of plain AT (**1a**) and GC (**2a**) with methylated AT (**1b**) and GC (**2b**) shows that methylation at the positions where the glucosidic N**–**C bond occurs in nucleosides (N9 in A and G, N1 in T and C), has essentially no influence on Watson–Crick pairing, i.e., hydrogen-bond distances and energies ΔE differ by only 0.01 Å and 0.3 kcal/mol, respectively. Similarly, only very small effects occur on substituting hydrogen with 2′-deoxyribose (**1c** and **2c**) or neutral 2′–deoxyribose-5′-phosphate residues (**1d** and **2d**) at the N9 and N1 atoms of the purine and pyrimidine bases, respectively. For AT, the hydrogen-bond energy decreases by only 0.3 kcal/mol on going from the Watson–Crick pair of plain nucleic bases (**1a**, –13.0 kcal/mol) to those of either nucleosides or nucleotides (1c and 1d, both –12.7 kcal/mol). At the same time, the N6**–**O4 and N1**–**N3 hydrogen-bond distances go from 2.85 and 2.81 Å (**1a**) to 2.87 and 2.77 Å in 1c and to 2.83 and 2.76 Å in 1d (Fig. 4.1). This does not resolve the discrepancy with the experimental values of 2.95 and 2.82 Å (**1**). Likewise, for GC, the hydrogen-bond energy changes only slightly as we go from the Watson–Crick pair of plain nucleic bases (2a, –26.1 kcal/mol) to those of either nucleosides or nucleotides (**2c** and **2d**, both –25.3 kcal/mol). The O6**–**N4, N1**–**N3 and N2**–**O2 hydrogen-bond distances change only by up to 0.03 Å along **2a**, **2c** and **2d** (Fig. 4.1), leaving us with the erroneous bond lengths and bond-length pattern of short–long–long versus the experimental order of long–long–short (**2**).

We went one step further by also introducing, in addition to electronic effects, the effect associated with the sterical restrictions or strain exerted on the hydrogen bonds by the backbone in DNA oligomers. This was done using a Watson–Crick complex between two strands that each consist of two nucleotides, namely, deoxyadenylyl-3′,5′-deoxyuridine, i.e., $(dApdU)_2$ shown schematically in Scheme 4.2 and designated **1f** in Fig. 4.2. This model closely resembles the corresponding adenylyl-3′,5′-uridine complex $(ApU)_2$ in the crystal (**1**) studied by Seeman *et al*. [15] (we have only removed the 2′-OH groups of ribose to reduce somewhat the immense computational cost).

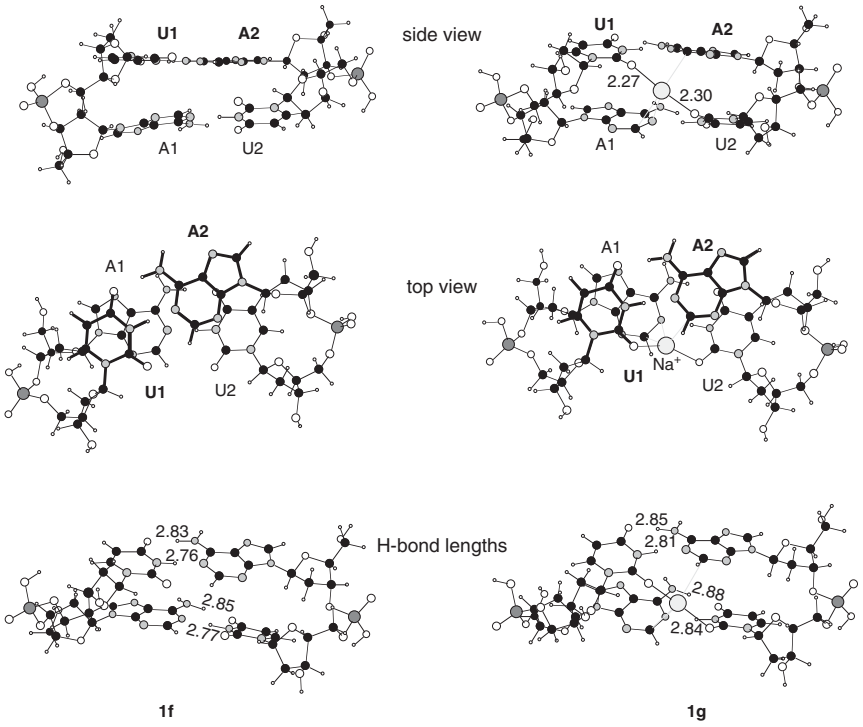
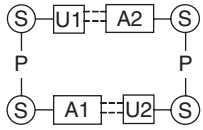


Fig. 4.2. Different perspectives of the BP86/DZP structure of the Watson–Crick-type dimer of deoxyadenylyl-3′,5′-deoxyuridine (see also Scheme 4.2), $(dApdU)_2$ without (**1f**) and with Na^+ ion (**1g**), both optimized in C1 symmetry without any symmetry constraint. The illustration shows N6–O4 and N1–N3 distances (Å) in AU pairs (lower panel) and the distances between the O2 atoms of each uracil base and the Na^+ ion (upper right).



Scheme 4.2

The BP86/DZP geometry of 1f is shown from different perspectives in Fig. 4.2, left. The AU hydrogen-bond distances in **1f** (Fig. 4.2) differ only slightly, at most by 0.03 Å, from those of plain AT (1a) also obtained at BP86/DZP (2.84 and 2.79Å, see Ref. [8]). The Watson–Crickpairing energy ΔE of **1f** amounts to –20.9 kcal/mol at BP86/TZ2P//BP86/DZP. Note that, although **1f** involves two AU pairs, this is significantly *less* than twice the pairing energy ΔE of AT (1a). This can be ascribed to the strain in the backbone, which shows up in the much higher preparation energy ΔE_{prep} of 9.2 kcal/mol, and not to the actual interaction energy ΔE_{int} of –30.2 kcal/mol between the strands which, in fact, *is twice as strong* as that of a single base pair (see Table 4.1).

Table 4.1 Analysis of the base-pairing interaction of adenine (A) with thymine (T), 2,4-difluorotoluene (F) and toluene (B) and between guanine and cytosine[a]

	AT	AF	AF*[b]	AB*[b]		GC
Hydrogen bond distances (in Å)						
N6–X4[c]	2.85	3.23	2.85	2.85	O6–N4	2.73
N1–Y3[d]	2.81	3.39	2.81	2.81	N1–Y3	2.88
					N2–O2	2.87
Bond energy decomposition (in kcal/mol)						
ΔE_{Pauli}	39.2	7.8	41.5	36.7		52.1
ΔV_{elstat}	–32.1	–7.0	–22.6	–15.5		–48.6
$\Delta E_{\text{Pauli}} + \Delta V_{\text{elstat}}$	7.1	0.8	18.9	21.2		3.5
ΔE_{σ}	–20.7	–3.8	–15.0	–11.4		–29.3
ΔE_{π}	–1.7	–0.2	–1.0	–0.6		–4.8
ΔE_{oi}	–22.4	–4.0	–16.0	–12.0		–34.1
ΔE_{int}	–15.3	–3.2	2.9	9.2		–30.6
ΔE_{prep}	2.3	0.2				4.1
ΔE	–13.0	–3.0				–26.5
ΔE (C_1)	–13.0	–3.0				–26.1

[a] Computed at BP86/TZ2P. Data for AT and GC from Ref. [18]. Data for AF, AF* and AB* computed with slightly different numerical integration settings from Ref. [19].
[b] In AF* and AB*, hydrogen bonds have been compressed to the equilibrium distances of AT.
[c] X4 = O4, F4 and H4 in the bases T, F and B.
[d] Y3 = N3, C3 and C3 in the bases T, F and B.

In conclusion, the backbone has only a marginal influence on hydrogen bonds in Watson–Crick pairs and is thus not the source of the disagreement between theoretical and experimental structures mentioned above. Yet, this finding was useful for speeding up our further efforts to find out what is the origin of this discrepancy because now it was justified to leave out the backbone and thus significantly reduce the computational cost when we tackled the next question in our quest: what is the effect of the molecular environment that the bases experience in crystals **1** and **2** that were studied experimentally?

4.3.3 Effect of the crystal environment

Reconciliation of theory and experiment regarding AT and GC structures is achieved if one incorporates the effects of the molecular environment on the Watson–Crick pairs in the crystals of sodium adenylyl-3′,5′-uridine hexahydrate (**1**) and sodium guanylyl-3′,5′-cytidine nonahydrate (**2**) into the theoretical model systems. We begin with AT or AU. In **1** [15], the amino–group of adenine and the O4 atom of uridine interact with two water molecules (A1U2) or with two 3′-ribose-OH groups of another $(ApU)_2$ complex (A2U1) (see illustration above). This was incorporated into our model by introducing two water molecules at the corresponding positions in AT (compare **1a** and **1e** in Fig. 4.1). The N1**–**N3 bond is not much affected but the N6**–**O4 bond expands significantly for two water molecules (**1e**). This leads to hydrogen bond lengths in close agreement with experiment (1e: N6**–**O4 and N1**–**N3 are 2.92 and 2.80 Å).

In the crystal (**1**), one of the sodium ions bridges the O2 atoms of the two uracil bases (see Fig. 4.2). The bridging Na^+ ion in the real crystal (**1**) binds simultaneously to two uridine bases that are part of opposite strands of the double-helical segment and involved in different AU pairs (Fig. 4.2). We have studied this at BP86/DZP by adding an Na^+ ion to $(dApdU)_2$ (**1g**, see Fig. 4.2). Going from **1f** to **1g**, the hydrogen–bond elongates 0.02–0.07Å. Eventually, the hydrogen bond lengths in **1g** still deviate significantly from the experimental values for **1** (compare Figs. 4.1 and 4.2). Apparently, introducing only this Na^+ ion does not reconcile experiment and theory. We note that at variance with the situation in our model **1g**, the sodium ion in **1** (crystal) does not enter into the space between the layers of the two AU pairs. Instead, it remains in the minor groove where it can also bind to water molecules.

Next, we consider the environment effects on the structure of GC pairs in **2** [16]. Here, the N7 and O6 and N2 positions of guanine are involved in hydrogen bonds with water molecules that, in the case of N7 and O6, coordinate to a Na^+ ion. The O2 position of cytosine also forms a hydrogen bond with a water molecule whereas N4 hydrogen binds to a 3′-ribose-OH group of a neighbouring $(GpC)_2$ complex. We have modelled the interactions of GC with its environment by introducing four water molecules and one sodium cation. The correct hydrogen-bond length pattern (long–long–short) is found with O6**–**N4, N1**–**N3 and N2**–**O2 distances that agree excellently (i.e., within 0.01–0.03 Å) with the X-ray data (2: 2.91, 2.95 and 2.86 Å).

We conclude that our DFT approach yields results (i.e., hydrogen bond lengths and strengths) that agree excellently with experiment if the most important hydrogen bonding interactions of a base pair with its environment in the crystal are incorporated into the model system. This justifies tackling this type of computationally extremely demanding problem (i.e., biological molecules involving hydrogen bonding) with DFT as an efficient alternative to traditional *ab initio* methods.

4.4 THE NATURE OF HYDROGEN BONDS IN DNA

4.4.1 Electrostatic interactions

Next, we wish to understand the nature of the forces that hold together the two bases and establish the importance of the different components in the interaction [7,18,19,22–25]. To this end, the bond energy ΔE is separated into the preparation energy ΔE_{prep} associated with deforming the individual DNA bases from their equilibrium structure into the geometry they adopt in the Watson–Crick pair and the actual interaction energy ΔE_{int} between these deformed bases in the base pair (Eq. 1) [6].

$$\Delta E = \Delta E_{prep} + \Delta E_{int} \tag{1}$$

As the DNA bases undergo only minor geometrical changes in the Watson–Crick pairs, the corresponding preparation energies are small, 4 kcal/mol or less (see Table 4.1) and will not be further discussed. The interaction energy ΔE_{int} is however important and we have analysed it in the framework of the Kohn–Sham MO model by quantitatively

decomposing it into (i) the electrostatic attraction ΔV_{elstat} between the unperturbed charge distributions of the bases, (ii) Pauli repulsive orbital interactions ΔE_{Pauli} between occupied orbitals, and (iii) bonding orbital interactions ΔE_{oi} that account for charge transfer (i.e., donor–acceptor interactions between occupied orbitals on one moiety with unoccupied orbitals of the other, including the HOMO–LUMO interactions) and polarization (Eq. 2) [6].

$$\Delta E_{\text{int}} = \Delta V_{\text{elstat}} + \Delta E_{\text{Pauli}} + \Delta E_{\text{oi}} \tag{2}$$

In addition, we analyse the charge distribution and orbital electronic structure of the bases as well as the charge redistribution and orbital mixing that corresponds to the Pauli repulsive and bonding orbital interactions.

We begin with examining the charge distribution at the front atoms of DNA bases, i.e., the atoms that participate in the hydrogen bonds that conect the bases in the Watson–Crick pairs. Figure 4.3 shows the atomic charges of all four DNA bases [18] computed with the VDD method [17]. Not unexpectedly, the DNA bases possess the right charge distribution for achieving a favourable electrostatic interaction in the Watson–Crick base pairs: all proton-acceptor atoms have a negative charge whereas the corresponding protons they face are positively charged.

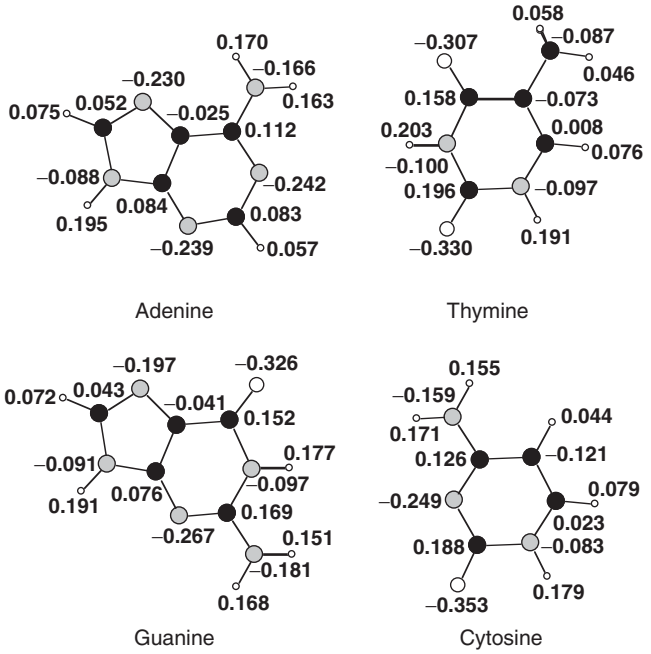


Fig. 4.3. VDD atomic charges (in a.u.) of the *isolated* bases adenine, thymine, guanine and cytosine obtained at BP86/TZ2P.

4.4.2 Orbital interactions

Next, we consider the orbital electronic structure of the DNA bases, in particular the σ-electron system in which the charge-transfer or donor-acceptor orbital interactions may occur. Our Kohn–Sham molecular orbital (MO) analysis shows that the electronic structures of the DNA bases indeed posses the basic features that are required for achieving favourable donor–acceptor orbital interactions, namely: (i) a lone pair on the proton-acceptor atom (nitrogen or oxygen) of one base pointing toward, and donating charge into, (ii) an unoccupied σ^* orbital of an N–H group of the other base. This leads to the formation of a weak bond. Figure 4.4 shows the relevant frontier-orbital interactions for AT as they emerge from our quantitative Kohn–Sham MO analysis. We see that the lone pair on the N1 atom of adenine interacts with the σ^* orbitals on H3–N3 of thymine and that the lone pair on the O4 atom of thymine is involved in donor–acceptor interactions with the σ^* orbitals on N6–H6 of adenine. A similar picture emerges for the Watson–Crick pair GC [18].

4.4.3 Charge redistribution

The VDD charge decomposition analysis [17–19] makes it possible to compute directly the *changes* in atomic charge ΔQ_A due to the formation of the base pair from separate bases, and to decompose them into contributions stemming from the charge redistributions in the σ- and π-electron systems, respectively (strictly speaking, A′ and A″ in the C_s symmetric base pairs):

$$\Delta Q_A = \Delta Q_A^{\sigma} + \Delta Q_A^{\pi} \tag{3}$$

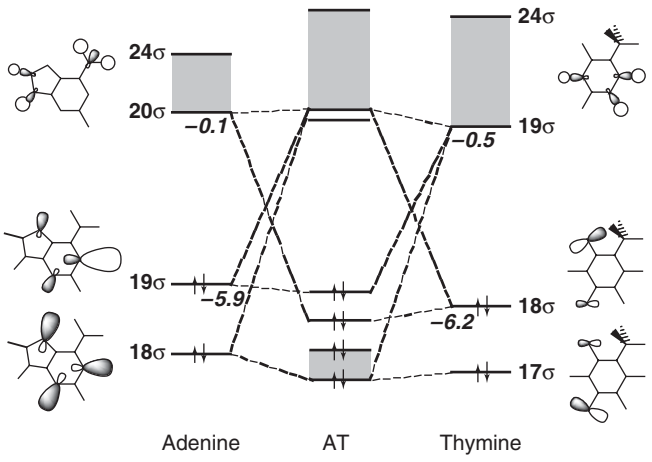


Fig. 4.4. Frontier orbital interactions (in the σ-electron system) between adenine (A) and thymine (T) in AT from Kohn–Sham DFT analyses at BP86/TZ2P, with σ_{HOMO} and σ_{LUMO} energies in eV. The group of lowest unoccupied orbitals involved is represented as a grey block. Selected orbitals of adenine (18σ to 20σ) and thymine (17σ to 19σ) are schematically represented.

Each of these components can be further decomposed into the charge rearrangements caused by Pauli repulsive orbital interactions and bonding orbital interactions:

$$\Delta Q_{\mathrm{A}} = \Delta Q^{\sigma}_{\mathrm{A,\,Pauli}} + \Delta Q^{\pi}_{\mathrm{A,\,oi}}$$

$$\Delta Q_{\mathrm{A}} = \Delta Q^{\sigma}_{\mathrm{A,\,Pauli}} + \Delta Q^{\pi}_{\mathrm{A,\,oi}} \tag{4}$$

The effect of the Pauli repulsion for the formation of AT from the bases A and T (see Fig. 4.5) is a depletion of charge density away from the central region of overlap and toward the periphery of the N6–H6•••O4 and N1•••H3–N3 hydrogen bonds. This causes a build-up of positive charge on the central hydrogen atom and of negative charge on the terminal atoms on either side of the hydrogen bond (see $\Delta Q^{\sigma}_{\mathrm{Pauli}}$ in Fig. 4.5). The changes in the atomic charges caused by the bonding orbital interaction in the σ-system (see $\Delta Q^{\sigma}_{0\mathrm{i}}$ in Fig. 4.5) shows that the electron-donor atoms of the N6–H6•••O4 and N1•••H3–N3 hydrogen bonds lose 24 and 46 mili-a.u. while the N–H bonds gain up to 54 mili-a.u. The charge redistribution in the π-electron system polarizes in such a way that the build-up of positive or negative atomic charges in the σ-electron system is cancelled or even overcompensated (see $\Delta Q^{\pi}_{0\mathrm{i}}$ in Fig. 4.5). Again, for the GC base pair, a similar picture emerges [17–19].

4.4.4 Bond energy decomposition

Now that we know that the DNA bases have suitable charge distributions for electrostatically attracting each other, and after having established the occurrence of σ charge transfer and π polarization, we want to quantitatively assess the magnitude and relative

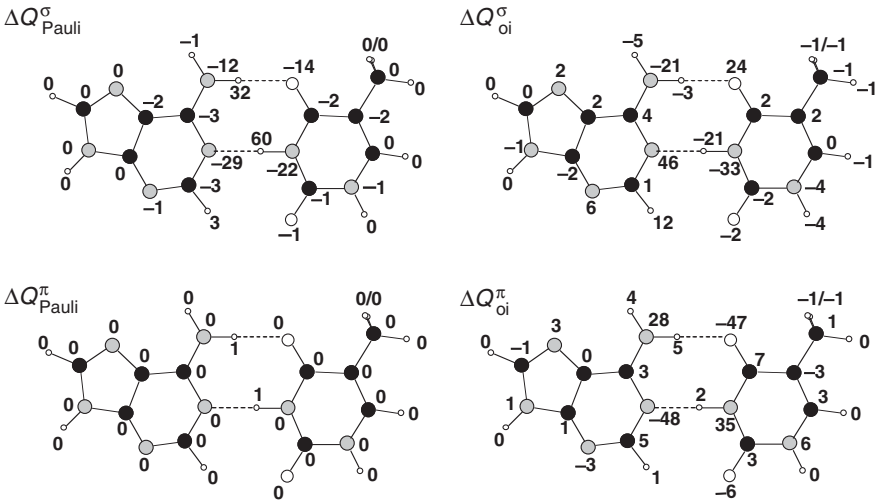


Fig. 4.5. Decomposition of VDD atomic charges (ΔQ, in mili-a.u.) associated with the formation of the AT pair from A and T into contributions caused by Pauli repulsion ($\Delta Q_{\mathrm{Pauli}}$) and bonding orbital interactions (ΔQ_{oi}) and into contributions stemming from the σ (ΔQ^{σ}) and π electrons (ΔQ^{π}) computed at BP86/TZ2P.

importance of the various components of the A–T and G–C base-pairing energy using the quantitative bond energy decomposition scheme mentioned in Section 4.4.1. The orbital interaction term ΔE_{oi} is further divided into a σ-component and a π-component. ΔE_{σ} consists mainly of the electron donor–acceptor interactions mentioned above. The π-component accounts basically for the polarization in the π-system, which partly compensates for the local build-up of charge caused by the charge-transfer interactions in the σ-system .

The striking result of our analysis is that charge-transfer or donor–acceptor orbital interactions are not at all a negligible or minor component in the hydrogen-bond energy of Watson–Crick base pairs (see Table 4.1). On the contrary, charge transfer is of the same order of magnitude as the electrostatic interaction! For AT, ΔE_{oi} is –22.4 kcal/mol and ΔV_{elstat} is –32.1 kcal/mol, and for GC, ΔE_{oi} is –34.1 kcal/mol and ΔV_{elstat} is –48.6 kcal/mol. Interestingly, the electrostatic interaction alone is not capable of providing a net bonding interaction at the equilibrium geometry at which it can only partly compensate for the Pauli-repulsive orbital interactions ΔE_{Pauli}. Without the bonding orbital interactions, the net interaction energies of AT and GC at their equilibrium structures would be repulsive by 7.1 and 3.5 kcal/mol, respectively (Table 4.1). This can be understood such that the orbital interactions cause the DNA bases to approach each other significantly more closely than if there were only electrostatic attraction. This results in hydrogen bond lengths at which the 'electrostatic-attraction-only' energy curve (i.e., $\Delta E_{int} - \Delta E_{oi} = \Delta E_{Pauli} + \Delta V_{elstat}$) is already repulsive. Our results for AT and GC agree well with and consolidate the finding of Reed and Weinhold [20] that the water dimer at equilibrium distance would be repulsive without the charge-transfer interactions.

Thus, our analyses disprove the established conception that hydrogen bonding in DNA base pairs is a predominantly electrostatic phenomenon. In this context, it is important to note that almost all arguments we found in the literature in favour of the electrostatic model were eventually based on the work of Umeyama and Morokuma [21] on the hydrogen bond in water dimers and other neutral hydrogen-bound complexes. But in fact, the analyses of Umeyama and Morokuma do reveal a significant charge-transfer component. They found that for the water dimer, for example, the total attractive interaction is provided for 72 percent by electrostatic interaction, for 21 percent by charge transfer and for 6 percent by polarization. We feel that the conclusions of Umeyama and Morokuma are not well represented if this charge-transfer component they found is completely ignored.

For both Watson–Crick pairs, i.e., AT and GC in their equilibrium geometry, we find that ΔE_{oi} provides even 41 percent of all attractive interactions, while electrostatic forces contribute 59 percent (Table 4.1). The ΔE_{oi} can be further split into 38 percent ΔE_{σ} and 3 percent ΔE_{π} for AT, and 35 percent ΔE_{σ} and 6 percent ΔE_{π} for GC. Also for mismatched base pairs [22], halogen-substituted base pairs [23] (at the H8 position for the purine and the H6 position for the pyrimidine bases) and for AT and GC mimics [24,25] in which N–H is substituted by C–H and oxygen by fluorine (see also Section 4.5) the same ratio between electrostatic and orbital interactions were obtained. We conclude that, at variance with current belief, charge-transfer or donor–acceptor orbital interactions play an essential role in the hydrogen bonds of DNA base pairs. They may also play an important role in other hydrogen-bonded complexes.

4.4.5 Cooperativity between hydrogen bonds and resonance assistance

After having established the occurrence of charge transfer in the σ system and polarization in the π system, it is of course interesting to know if the hydrogen bonds in AT and GC reinforce each other by donating charge in opposite directions, which reduces the net build-up of charge on each base, and if there is a cooperative effect or resonance assistance by the π-electrons system as suggested by Gilli *et al*. [9]. The answers for these questions were found with calculations in which individual types of orbital interactions were considered while others were switched off by removing the appropriate σ and π virtuals from the respective bases (see Table 4.2) [18].

The synergism within the σ system between charge transfer from one base to the other through one hydrogen bond and back through the other bond was obtained by comparing the situation A(σ,–)T(σ,–), where donor–acceptor interactions of both hydrogen bonds were switched on simultaneously, with the sum of A(σ,–)T(–,–) and A(–,–)T(σ,–), where the donor–acceptor interactions of each hydrogen bond work separately at a time (that is in the absence of the other). Note that π-polarization is switched off completely. The difference between the situation of simultaneously ocurring hydrogen bonds and that of separately occurring hydrogen bonds, both providing an interaction of ca. 21 kcal/mol, amounts to merely half a kcal/mol (see Table 4.2) [17]. In other words, there is only a little synergism.

The resonance assistance that the π-electron system contributes in addition to the bonding provided by the σ-electron system consists primarily of the additional ΔE_π term of 1.3 kcal/mol, which accounts for the polarization of π electrons. The synergism between charge transfer in the σ-electron system and polarization in the π-electron system can be computed as a difference between A(σ,π)T(σ,π), where both σ charge transfer and π polarization occur simultaneously, and the sum of A(σ,–)T(σ,–) and A(–,π)T(–,π), where both σ charge transfer and π polarization occur independently. The synergetic effect between σ and π electrons working simultaneously yields an additional stabilization of 0.7 kcal/mol (which stems from the increase of ΔE_π from –1.3 to –1.7 kcal/mol and of ΔE_σ from –20.4 to –20.7 kcal/mol, see Table 4.2). This is however only half the size of

Table 4.2 Analysis of the synergy between σ and π-orbital interactions in A–T and G–C[a]

	Virtuals available[b]	ΔE_σ	ΔE_π
I	A(σ,π)T(σ,π)	–20.7	–1.7
IIa	A(–,–)T(σ,–)	–12.9	
IIb	A(σ,–)T(–,–)	–8.3	
IIIa	A(σ,–)T(σ,–)	–20.4	
IIIb	A(–,π)T(–,π)		–1.3

[a] Computed at BP86/TZ2P. Data from Ref. [18].

[b] Our notation: A(σ,π)T(σ,π) corresponds to a regular computation on AT in which all σ and π virtuals are included; A(σ,–)T(σ,–), for example, indicates that all σ virtuals are available on A and T whereas the π virtuals have been removed from both bases.

the ΔE_{π} term of –1.3 kcal/mol associated with A(–,π)T(–,π). For the base pair GC similar results were found [18]. Although the contribution of the π-electron system is small compared with the bonding provided by the σ-electron system, it is able to shorten the hydrogen bonds by ca. 0.1 Å because of the shallow potential energy surface associated with these hydrogen bonds. In this respect it is justified to consider these hydrogen bonds as resonance-assisted hydrogen bonds (RAHB).

4.5 DNA REPLICATION: STERIC FACTORS VERSUS HYDROGEN BONDING

In the standard model, the immensely high accuracy with which DNA replication occurs is ascribed to the specificity of the hydrogen bonds in the Watson–Crick pairs adenine–thymine (AT) and guanine–cytosine (GC) [1]. This view has recently been challenged by experiments with artificial nucleotides [3] or with alterations of the binding pocket of polymerase [4a], which show that geometric constraints are important for the replication fidelity. Experiments [4b] with hydrophobic base pairs suggest that hydrophobicity may also be a sufficient driving force for selective replication. Here, we focus on the results by Kool and co-workers [3] who proposed that *not* hydrogen bonding in Watson–Crick pairs but steric effects, that is, the shape of DNA bases is mainly responsible for the high fidelity of DNA replication. This idea evolved from experiments in which 2,4-difluorotoluene (F), an isoster of thymine (T), encodes in a template strand the DNA polymerase-catalysed insertion of deoxyadenosine triphosphate (dATP) and that adenine (A) encodes the insertion of the deoxynucleoside triphosphate of difluorotoluene (dFTP), in spite of the supposed apolarity and absence of hydrogen bonding ability of F [3].

Although the complexation energy of AF (–3.2 kcal/mol) is indeed much weaker than that of AT (–13.0 kcal/mol) and the two hydrogen bonds in AF are 0.38 and 0.58 Å, respectively, longer than those in AT (see Table 4.1), the complexation of AF has qualitatively the same bonding mechanism. F has also qualitatively the same charge distribution as T: hydrogen bond-acceptor atoms are negatively charged while the corresponding hydrogen atoms carry a positive charge. The magnitude of these atomic charges is however much smaller in F than in T and as a consequence, the electrostatic attraction between A and F is smaller than between A and T (see Table 4.1). The orbital electronic structure reveals that, compared with T, F is a poor electron donor and a poor electron acceptor, as can be seen by comparing Figs 4.4 and 4.6. The occupied orbital with lone pair-character on the F4 atom in F is much lower in energy than the corresponding orbital with lone pair-character on the O4 atom in T. Also, the unoccupied orbitals with sizeable C3–H3 σ^* character in F are somewhat higher in energy than the corresponding N3–H3 σ^* orbitals in T.

Our analyses do confirm the idea that F is a true isoster of T, i.e. it has the same steric shape: when the AF hydrogen bonds N6–F4 and N1–C3 are compressed to the corresponding values 2.85 and 2.81Å in the AT Watson–Crick geometry (we refer to this compressed complex as AF*), the steric repulsion, owing to ΔE_{Pauli}, is only 2.8 kcal/mol higher for AF* (41.6 kcal/mol) than for AT (38.7 kcal/mol).

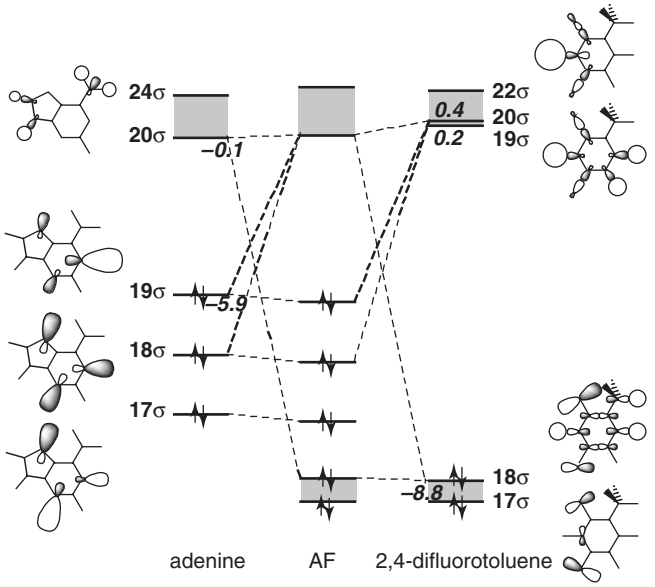


Fig. 4.6. Frontier orbital interactions (in the σ-electron system) between adenine (A) and 2,4-difluorotoluene (F) in AF from Kohn–Sham DFT analyses at BP86/TZ2P, with σ_{HOMO} and σ_{LUMO} energies of the bases (in eV). The group of lowest unoccupied orbitals involved is represented as a grey block. Selected orbitals of adenine (17σ to 20σ) and difluorotoluene (17σ, 18σ, 20σ, 22σ) are schematically represented.

The analysis of AF* furthermore highlights the importance of hydrogen bonding in enzyme-catalysed DNA replication. It has been convincingly shown that this reaction involves a relatively tight active site, which requires the new base pair between template base and incoming base to conform to a geometry very close to the Watson–Crick geometry. If an incoming nucleotide is too large to fit into the active site of DNA polymerase, it causes steric repulsion and thus a high overall activation energy for the insertion of the nucleotide. This is schematically shown below in 3 (bold lines = active site; hatched area = spatial overlap of too large bases):

3 **4**

Base pairs of the right shape do fit into the active site without much repulsion (see 4 above). Such a situation is required for achieving a low overall activation barrier of the insertion. But the transition state is also stabilized by any favourable interaction that may occur between the bases. This is indicated by dots in **5** below:

5

In the case of the natural Watson–Crick pairs AT and GC, this favourable interaction is the hydrogen bond energies ΔE of –13.0 and –26.1 kcal/mol, respectively [7,8,18]. We have shown previously that this intrinsic base-pairing interaction is affected neither by microsolvation of the major and minor grooves nor by the interaction between two stacked base pairs. Thus, the barrier for nucleotide insertion increases substantially if the stabilizing hydrogen bonds cannot be formed between the bases in the active site of the enzyme.

This is obvious for the relatively firmly bound natural Watson–Crick pairs AT and GC. But a similar argument holds true also for the weakly bound mimic AF. This is *not* obvious, at first sight, especially if one considers that the base pair in the geometrically confined active site of DNA polymerase, for which the compressed AF* complex is a model, is not even bound but instead slightly repulsive with a net hydrogen bond interaction ΔE_{int} of ca 3 kcal/mol (Table 4.1). Further analyses show, however, that without hydrogen bonding, the barrier for the process of nucleotide insertion would become much higher, in fact, restrictively high. This is clear if one realizes that the net repulsive interaction would increase much more strongly in AF* in the absence of the underlying electrostatic attraction (–22.6 kcal/mol) and covalent orbital interactions (–16.0 kcal/mol, Table 4.1). The charge-transfer character of the latter is supported by the VDD charge decomposition analysis in Fig. 4.7. Loss of the orbital interactions ΔE_{oi}, for example, would raise the overall barrier for the formation of a new base pair by 16.0 kcal/mol.

In practice, it is of course difficult to simply switch off the orbital interactions completely. They can be further reduced, however, if 2,4-difluorotoluene is replaced by a less polar mimic with even poorer electron donor and acceptor capabilities, for example, toluene (B). We find that the planar, C_S symmetric AB pair is practically unbound at BP86/TZ2P. To simulate again a tight DNA-polymerase active site as proposed by Kool (*vide supra*) [3], we have compressed the AB pair to a geometry AB* in which the N6–H6•••H4 and N1•••H3–C3 hydrogen bond distances adopt the values of the corresponding bonds in the equilibrium structure of AT (2.85 and 2.81 Å). And indeed, the net base-pairing interaction ΔE_{int} in AB* (9.2 kcal mol^{-1}) turns out substantially more repulsive than in AF* (2.9 kcal mol^{-1}). The reason is a significant decrease of both the orbital interactions ΔE_{oi} and electrostatic attraction ΔV_{elstat}, by a factor 2 (!) if compared with AT (see Table 4.1). Note that the steric repulsion term ΔE_{Pauli} does not vary so much along AT, AF* and AB*.

The above results support aspects of both and lead to a synthesis of Kool's steric model of DNA replication [3] and the standard model that is based on hydrogen bonding [1]. Starting from the situation with only steric factors, the overall barrier for insertion of a nucleotide is represented schematically in Fig. 4.8 with a dashed line. Electrostatic and orbital interactions between DNA bases can contribute to reducing the overall barrier for insertion of a nucleotide if the net hydrogen bond strength is weak and even if it is moderately

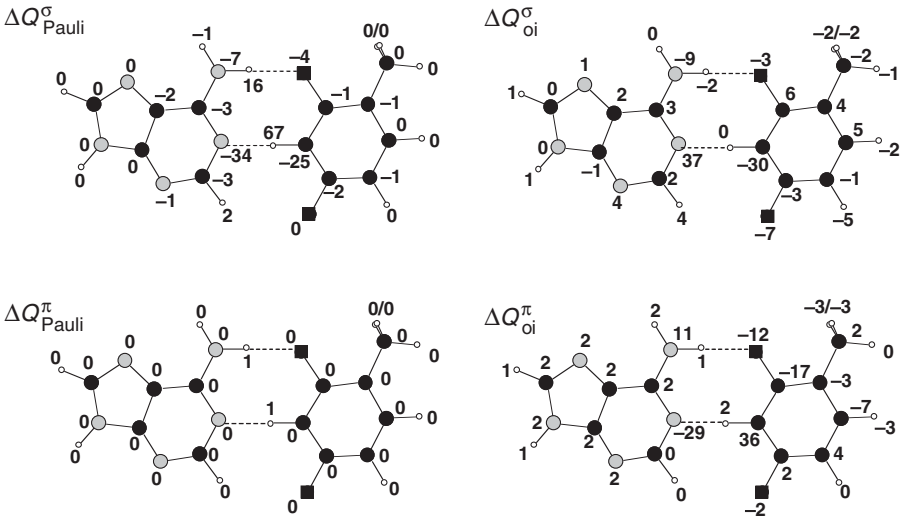


Fig. 4.7. Decomposition of VDD atomic charges (ΔQ, in mili-a.u.) associated with the formation of the AF* pair (AF pair compressed to AT Watson-Crick geometry) from A and F into contributions caused by Pauli repulsion (ΔQ_{Pauli}) and bonding orbital interactions (ΔQ_{oi}) as well as into contributions stemming from the σ (ΔQ^{σ}) and π electrons (ΔQ^{π}) computed at BP86/TZ2P.

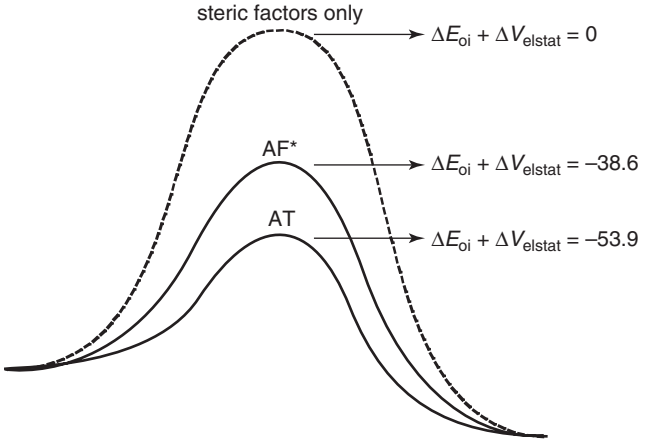


Fig. 4.8. Schematic representation of the barrier for insertion of a nucleotide with only steric factors (dashed line) contributing to the barrier height and with electrostatic and orbital interaction contributing to the barrier height (AF* and AT). The real barrier is actually a multi-step process.

repulsive. Particularly, in the case of the natural DNA base pairs, the barrier in Fig. 4.8 is lowered substantially because of the large electrostatic and orbital interaction components. Note, however, that we have not computed the barrier height as such. DNA replication is actually a multi-step process with more than one barrier, which has not yet been fully

elucidated. Our results do not rule out the possibility of the overall barrier being low for other reasons than favourable steric factors and intrinsic hydrogen bonding, e.g., favourable solvent effects or stacking interactions. It is conceivable, even likely, that one or more of these factors are active in concert for achieving efficient and selective replication.

4.6 CONCLUSIONS

Density functional theory (DFT) is an adequate and efficient tool for accurately describing systems involving (multiple) hydrogen bonding. In particular, a long-standing discrepancy between theory (*ab initio* and hybrid DFT) and experiment (X-ray crystal structures) could be traced to effects of the molecular environment in the crystal that were missing in the theoretical model systems. Inclusion of these effects in DFT computations leads to reconciliation of theory and experiment. An important implication is that DFT can be applied as an efficient alternative to conventional *ab initio* methods for tackling large biomolecular systems in which hydrogen bonds play a prominent role.

Proceeding from the above result, we were able to tackle a number of computationally demanding problems associated with the nature and behaviour of DNA base pairs as well as issues concerning the mechanism of DNA replication. In the first place, we have established the occurrence of a significant covalent or orbital-interaction component in the hydrogen bonds of Watson–Crick base pairs that, in fact, is of the same order of magnitude as the electrostatic attraction. Furthermore, the π-electron system assists the hydrogen bonds through delocalization of the π electrons. There is also a small synergetic effect of the π and σ electrons working simultaneously. In this respect, it is justified to consider the hydrogen bonds in Watson–Crick pairs as resonance–assisted hydrogen bonds (RAHB). Finally, we have also shown that, proceeding from the idea of a tight, sterically confined active site in DNA polymerase, not only the steric shape of DNA bases plays a role in the process of biomolecular recognition but also the orbital interactions and electrostatic attraction of the hydrogen bonds. This is a synthesis of the standard model of DNA replication, which is based on hydrogen bonding alone, and the more recent steric model.

4.7 ACKNOWLEDGEMENTS

We thank the National Research School Combination–Catalysis (NRSC-C) for a postdoctoral fellowship for C.F.G. and the Netherlands organization for Scientific Research (NWO) for financial support.

4.8 REFERENCES

1 (a) L. Stryer, Biochemistry, W.H. Freeman and Company, New York, 1988; (b) A. Kornberg and T.A. Baker, DNA Replication, W.H. Freeman and Company, New York, 1992.

2 I.K. Yanson, A.B. Teplitsky and L.F. Sukhodub, Biopolymers, 18 (1979) 1149. (With corrections by K. Brameld, S. Dasgupta, W.A. Goddard III, J. Phys. Chem. B, 101 (1997), 4851.)

3 E.T. Kool, J.C. Morales and K.M. Guckian, Angew. Chem., 112 (2000) 1046; Angew. Chem. Int. Ed., 39 (2000) 990 and references cited therein.
4 (a) D. Summerer and A. Marx, Angew. Chem., 113 (2001) 3806; Angew. Chem. Int. Ed., 40 (2001) 3693; (b) M. Berger, A.K. Ogawa, D.L. McMinn, Y. Wu, P.G. Schultz and F.E. Romesberg, Angew. Chem., 112, (2000), 3069; Angew. Chem. Int. Ed., 39 (2000) 2940.
5 G. te Velde, F.M. Bickelhaupt, S.J.A. van Gisbergen, C. Fonseca Guerra, E.J. Baerends, J.G. Snijders and T. Ziegler, J. Comput. Chem., 22 (2001) 931
6 F.M. Bickelhaupt and E.J. Baerends, in K.B. Lipkowitz and Rev. Comput. Chem., Kohn-Sham Density Functional Theory: Predicting and Understanding Chemistry, Vol 15, D.B. Boyd (Eds.), Wiley-VCH, New York, 2000, pp. 1–86.
7 C. Fonseca Guerra and F.M. Bickelhaupt, Angew. Chem., 111 (1999) 3120; Angew. Chem. Int. Ed., 38, (1999), 2942.
8 C.Fonseca Guerra, F.M. Bickelhaupt, J.G. Snijders and E.J. Baerends, J. Am. Chem. Soc., 122 (2000) 4117.
9 G. Gilli, F. Bellucci, V. Ferretti and V. Bertolasi, J. Am. Chem. Soc., 111 (1989) 1023.
10 P. Hobza, O. Bludsky and S. Suhai, Phys. Chem. Chem. Phys., 1 (1999) 3073.
11 A. Halkier, H. Koch, P. Jørgensen, O. Christiansen, I.M.B. Nielsen and T. Helhaker, Theor. Chem. Acc., 97 (1997) 150.
12 M. Schütz, S. Brdarski, P.-O. Widmark, R. Lindh and G. Karlström, J. Chem. Phys., 107 (1997) 4597.
13 J. Bertran, A. Oliva, L. Rodríguez-Santiago and M. Sodupe, J. Am. Chem. Soc., 120 (1998) 8159.
14 J. Sponer, J. Leszczynski and P. Hobza, J. Phys. Chem., 100 (1996) 1965.
15 N.C. Seeman, J.M. Rosenberg, F.L. Suddath, J.J.P. Kim and A. Rich, J. Mol. Biol., 104 (1976), 109.
16 J.M. Rosenberg, N.C. Seeman, R.O. Day and A. Rich, J. Mol. Biol., 104 (1976) 145.
17 C. Fonseca Guerra, J.-W. Handgraaf, E.J. Baerends and F.M. Bickelhaupt, J. Comput. Chem., 25 (2004) 189.
18 C. Fonseca Guerra, F.M. Bickelhaupt, J.G. Snijders and E.J. Baerends, Chem. Eur. J., 5 (1999) 3581.
19 C. Fonseca Guerra and F.M. Bickelhaupt, J. Chem. Phys., 119 (2003) 4262.
20 A.E. Reed and F. Weinhold, J. Chem. Phys., 78 (1983) 4066.
21 H. Umeyama and K. Morokuma, J. Am. Chem. Soc., 99 (1977) 1316.
22 C. Fonseca Guerra, F.M. Bickelhaupt and E.J. Baerends, Crystal Growth & Design, 2 (2002) 239.
23 C. Fonseca Guerra, T. van der Wijst and F.M. Bickelhaupt, Struct. Chem., 16 (2005) 211.
24 C. Fonseca Guerra and F.M. Bickelhaupt, Angew. Chem., 114 (2002) 2194; Angew. Chem. Int. Ed., 41 (2002) 2092.
25 C. Fonseca Guerra, F.M. Bickelhaupt and E.J. Baerends, Chem. Phys. Chem., 5 (2004) 481.

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 5

Quantum chemical modeling of charge transfer in DNA

Alexander A. Voityuk

Institució Catalana de Recerca i Estudis Avançats, Institute of Computational Chemistry, Universitat de Girona, 17071 Girona, Spain alexander.voityuk@icrea.es

Abstract

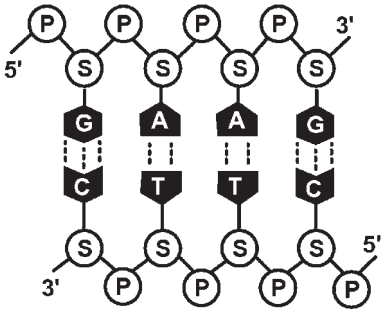
Charge migration in DNA is the subject of intense experimental and theoretical study motivated by its potential application in nanotechnology. Computational modeling of charge transfer (CT) mediated by DNA can be very useful to estimate the role of different factors determining the CT process and to predict CT parameters of specific systems.

It has been recently recognized that structural fluctuations of DNA and its environment can considerably affect the conducting properties of DNA and, therefore, have to be taken into account in computational models. In this review, we consider the sensitivity of key parameters of charge transfer in DNA (the donor–acceptor electronic coupling, the free energy of CT and the reorganization energy) to conformational changes of DNA and the dynamics of its environment and discuss computational approaches to estimating these parameters.

5.1 INTRODUCTION

Soon after Watson and Crick discovered the double-helix structure of DNA in 1953 [1], it was suggested that DNA with its stack of aromatic heterocycles could serve as an electronic conductor [2,3]. However, only 40 years later, long-range charge transport (over distances $\geqslant$ 100 Å) through DNA was experimentally observed [4].

The double helix of DNA forms when two single strands combine. Each strand is composed of a polyanionic sugar-phosphate backbone linking together the nucleobases adenine (A), guanine (G), cytosine (C), and thymine (T). Thus, within its interior, DNA contains an extended array of π stacked base pairs formed due to hydrogen bonds between A and T, and between G and C. A GAAG fragment in the double-stranded DNA is shown in Scheme 5.1. The array of stacked A–T and G–C pairs is a key structural element, which makes DNA a medium for long-range CT. The overlapping π orbitals of the bases provide a pathway for



Scheme 5.1

migration of charge carriers generated in the stack. The sugar–phosphate backbone does not play any remarkable role in the charge movement.

Recent experiments have provided contradictory results on electrical conductivity of DNA. The molecule ranges from a highly conducting wire to a semiconductor and an insulator [5–9]. Thus, the question–*Whether DNA is an electric conductor or not*–remains unsettled because of a number of factors that are very difficult to control. In particular, the nature of the contacts of DNA to metallic leads presents a considerable challenge.

Using a range of spectroscopic and biochemical methods, it has been established that double-helical DNA is a medium for efficient HT [10–15]. In such experiments one employs structurally well-defined systems consisting of an electron donor and acceptor covalently attached to a DNA oligonucleotide. By photoexcitation of the chromophore, an electron (or electron hole) is injected into the system. Then, one observes migration of the charge along the π stack. The CT efficiency is determined, either by measuring the fluorescence decay of the chromophore or by analysis of the relative yield of reaction products, which are detected as strand cleavage at different positions in the DNA stack.

Such measurements provide a possibility to examine the effects of different factors, e.g., the nature of the donor and the acceptor and the length and sequence of the bridge, on electron transfer through the stack.

There are two types of charge transfer in DNA:

(1) positive charge or electron hole transfer (HT) when a radical-cation state moves from one base pair to another; and (2) excess electron transfer (ET) when a radical-anion state migrates through the π-stack. The oxidation potential of nucleobases increases as $G < A \ll C \approx T$ [16]. Thus, the purine bases G and A are much easier to oxidize than the pyrimidine bases C and T. The cation radical G^+ is a key intermediate formed by one-electron oxidation of DNA. Created initially adjacent to an oxidant, the charge hops through the stack until it reacts irreversibly with water or molecular oxygen. As a result, G^+ can be generated in DNA far away from the oxidant because of the hole transfer process. Conversely, C and T are easier to reduce than are A and G [17] and therefore, an excess electron generated in a π-stack will be localized on T and C. The majority of experimental and theoretical studies of charge transfer in DNA have been related to radical-cation states. Considerably less is known about excess electron migration, although Several photochemical and spectroscopic studies of the negative charge transfer in DNA have been published recently [10,18].

Material scientists have only recently turned their attention to charge migration in DNA for the development of DNA-based molecular electronic devices [8,9–21]. The structures of DNA assemblies can be routinely prepared and encoded by the base pair sequence, and therefore, the double helix, which can also mediate charge transfer, may serve as a unique template in molecular electronics.

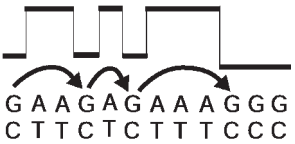
Besides being of potential importance in nanotechnology, CT in DNA is a very interesting phenomenon for molecular biology. Electron injection into DNA leads in many organisms to the efficient repair of DNA lesions induced by UV irradiation [22,23]. Also, the idea of doing 'chemistry at a distance' [24] has attracted much attention. In this context, the distance dependence of CT through DNA is the most important issue. Several types of distance dependence have been experimentally found for the hole transfer in DNA [10–15]. Theoretical studies have brought considerable progress in understanding the charge movement through π-stacks [25–30]. Now there is general agreement that two different mechanisms are relevant for short- and long-range CT.

1. Superexchange (or coherent transport, or tunneling) is characterized by a fast exponential decay of the HT rate with the donor–acceptor distance R.

2. $$k \quad \exp(-\beta R), \tag{1}$$

   where the decay parameter β is typically in the range of 0.6–1.4 Å^{-1}. When $\beta = 0.7$ Å^{-1}, the rate constant decreases by an order of magnitude with each additional (AT) base pair inserted between donor and acceptor and it drops by two orders of magnitude when $\beta = 1.4$ Å^{-1}. The superexchange mechanism is operative at short distances, $R \leqslant 15$Å. Grozema *et al.* [30] have proposed a model that relates the magnitude of β to the energy gap between the donor and bridge electronic levels. The fall-off parameter rapidly decreases with the lowering gap. For energy gaps of 0.55 eV and 0.27 eV the corresponding β parameter is found to be ~ 0.85 Å^{-1} and ~ 0.34 Å^{-1}. Although, in principle, single-step tunneling allows for long-range charge transfer, its efficiency becomes very low. In fact, the possibility of a long-range coherent charge transport through DNA can be ruled out.
3. Multi-step hole hopping gives rise to long-range charge transport ($R \leqslant 200$Å). The mechanism may be described as G-hopping, a series of tunneling steps between guanines separated by AT pairs (Scheme 5.2).

In this case, the hole transfer rate is given by $k \propto N^{-\eta}$ where N is the number of hopping steps and the power parameter η is between 1 and 2 [25]. The probability of individual hopping steps is directly controlled by superexchange. When Gs are separated by more than three AT pairs, hole hopping between adenine bases, A-hopping, may also become operative [31].



Scheme 5.2

Thus, single-step tunneling is relevant both for short- and for long-range CT in DNA stacks. Below we consider results of computational modeling of the CT process in short DNA stacks. A general approach to the quantum-mechanical treatment of electron transfer is discussed in detail by Newton [32]. Theoretical aspects of charge movement in DNA have recently been considered by Berlin *et al*. [33].

5.2 BASICS OF THE ELECTRON TRANSFER THEORY

The Marcus equation is a good starting point to consider basic parameters controlling the nonadiabatic charge migration process [34]. The rate constant of the nonadiabatic charge transfer from a donor *d* to an acceptor *a* can be written as:

$$k_{CT} = \frac{2\pi}{\hbar\sqrt{4\pi\lambda k_B T}} V_{da}^2 \exp\left[-(\Delta G^0 + \lambda)^2/4\lambda k_B T\right] \quad (2)$$

In line with this equation, three factors influence the reaction rate. V_{da} is the effective donor–acceptor coupling mediated by a bridging medium. The electronic couplings of neighboring base pairs are found to be 0.01–0.15 eV (see Table 5.2) [35,36]. The interactions between second neighbors are several times weaker, and therefore, the systems in which donor and acceptor are separated by one or more intervening base pairs, are expected to fall well within the nonadiabatic regime. The driving force ΔG^0 is the difference of redox potentials of the donor and acceptor sites. The reorganization energy λ is the change in the free energy to move the reactants to the product configuration without actually transferring the electron. Two main contributions to the reorganization energy are generally distinguished, $\lambda = \lambda_i + \lambda_s$; the internal reorganization energy λ_i is associated with fast changes in the intramolecular geometry of d and a initiated by CT. The solvent contribution λ_s is associated with slow changes in the polarization of the surrounding medium. The magnitude of λ_s can be calculated from a classical model in which the solvent medium is treated as a dielectric continuum.

The CT rate decays exponentially with d–a distance *R* (Eq. (1)). The fall-off parameter β depends both on the nature of the bridge and on the electronic parameters of donor and acceptor [37,38]. In the case of charge separation ($D\ A \rightarrow D^+\ A^-$), ΔG^0 is a function of the d–a distance (a negative correction accounting for bringing D^+ and A^- to the distance *R* has to be accounted for); however, for charge shifting ($D^+\ A \rightarrow D\ A^+$) ΔG^0 is nearly independent of *R*. Therefore, in line with Eq. (2), the parameter β for the charge shift process can be expressed approximately as a sum of electronic (el) and reorganization (r) contributions: $\beta = \beta_{el} + \beta_r$ and these effects may be considered separately by discussing the distance dependence of CT.

In the next section, we consider quantum chemical calculations ΔG^0 and V_{da}. Estimation of the reorganization energy for CT in DNA has been considered in detail in several recent papers [39–42] and will not be addressed below.

5.3 QUANTUM MECHANICAL MODELS

While π-stacks of nucleobase pairs hold some similarities to solid-state systems there are also crucial differences. Four different aromatic moieties (G, A, C, T) form an aperiodic

structure. The conformation of the π-stack is very flexible. The parameters ΔG^0 and V_{da} depend essentially on DNA sequence and the π-stack conformation as well as on the position and dynamics of surrounding water molecules and sodium ions. These structural and dynamic factors affect the efficiency of charge transport through DNA stacks and should be taken into account. However, an accurate quantum mechanical treatment of extended models entails substantial computational demands, and therefore, approximate methods and rather simple models are employed in computational studies of DNA-mediated CT reactions.

The most consequent approach is the *ab initio* molecular dynamics [43,44]. However, this method still cannot be applied routinely to microscopic modeling of charge transfer in DNA because of enormous computational demands. Hartree–Fock (HF) and density functional theory (DFT) calculation may be performed for systems consisting of several base pairs. These methods were employed for estimating the ionization potentials [45–47], electron affinities [48–51] and electronic couplings [36] in rather small models. Such calculations become completely impractical in combination with MD simulations where thousands of DNA snapshots have to be treated. Semi-empirical methods are considerably more efficient and can successfully be applied to extended models [38,52–54]. Combined QM/MD schemes based on semi–empirical methods have been employed in several studies on CT in DNA [55–57].

Geometry of single- and double-strand DNA fragments may be generated by different programs, for instance, X3DNA [58] employing atomic coordinates of nucleobases derived from high-resolution X-ray crystal structure [59] or quantum chemical calculations. By modeling the superexchange interaction in DNA the sugar–phosphate backbone can be neglected [52].

5.3.1 Driving force ΔG^0

Since electron holes are trapped at sites of minimum oxidation potentials, calculations of ionization energies (IE) of base pair sequences are useful for predicting the reactivity of different sites in DNA toward one-electron oxidation. In particular, the energy for hole transfer between two bases B and B′ prime can be estimated as the IE difference for these bases embedded in the duplex. Saito and co-workers [45,46] reported ionization potentials (IP) for XGY triplets calculated at the HF/6-31G* level. They concluded that the oxidation potential of G is strongly influenced by adjacent 3′- and 5′-base pairs. Semi–empirical calculations of all possible triplets 5′-XBY-3′ demonstrated that the oxidation potential of B in 5′-XBY-3′ is considerably affected by the nature of Y and b ecomes smaller in order C ~ T > A > G while the effect of the *preceding* base X is rather small [60]. Table 5.1 lists the relative energies of radical-cation states localized on G and A within different triplets calculated with the semi–empirical NDDO method [60].

The results suggest that the 5′-G in GG and both Gs on the 5′-side in GGG have the lowest oxidation potential, and thus, are the best hole traps. Similar estimates were derived from DFT calculations of triplets XBY [47,61]. The predicted differences of oxidation potentials [60,61] are in good agreement with experimental data [62].

Chemical modification of nucleobases changes their redox potentials and thereby can modulate CT properties of the system [63,64].

Table 5.1 The effect of neighboring base pairs X and Y on the relative oxidation potentials of G and A in the stacks 5'-XBY-3' (in eV)

Y	G	A	T	C
GGY	0.000	0.132	0.265	0.303
AGY	0.001	0.134	0.266	0.304
TGY	0.026	0.157	0.288	0.326
CGY	0.036	0.168	0.299	0.336
GAY	0.442	0.575	0.697	0.717
AAY	0.459	0.602	0.718	0.736
TAY	0.486	0.617	0.747	0.765
CAY	0.496	0.629	0.749	0.769

Schuster and co-authors [65,66] showed that the energetics of a hole state in DNA is strongly affected by the configuration of neighboring sodium ions. Because of that, charge transfer in DNA depends on the probability of forming certain ion configurations that are effective in changing the hole density over the duplex DNA. A combined QM/MD study [57] has demonstrated that the dynamics of water molecules strongly dominates the ΔG^0 fluctuations. These fluctuations are large enough to render electron hole transfer from G^+ to A energetically feasible [57]. Thus, the external electrostatic potential may essentially effect the redox potential of nucleobases and thereby the position of a localized hole in DNA.

5.3.2 Estimation of the electronic coupling V_{da}

Several quantum chemical methods have been developed to evaluate the electronic coupling of electron donor and acceptor [32,67,68]. Estimation of V_{da} in DNA has been recently considered in detail [69]. Here we outline several methods employed for the V_{da} treatment in DNA π-stacks.

A widely used method to compute the coupling matrix element is the estimation of the 'minimum energy splitting' of the donor and acceptor electronic levels [2]. When donor and acceptor are equivalent by symmetry

$$V_{da} = \frac{1}{2}(E_2 - E_1)$$

where E_1 and E_2 are the system energy of the ground and excited adiabatic states, respectively. If donor and acceptor are 'off resonance' an external perturbation must be applied to bring the electronic levels into resonance. This procedure is equivalent to minimizing the adiabatic splitting. For instance, the external electric field can be suitably adjusted to minimize the energy gap. This approach has been applied to estimate the coupling in small DNA fragments [35,36]. However, for systems, in which the d–a electronic interaction is weak, $|V_{da}| < 10^{-3}$ eV, minimization of the adiabatic gap to sufficient accuracy is cumbersome; it becomes a real challenge when treating open-shell systems. There are alternative approaches, which permit direct treatment of states that are off-resonance.

The generalized Mulliken–Hush (GMH) scheme [67] and the fragment charge method (FCM) [68] imply transformations of adiabatic states to diabatic states that are localized on donor and acceptor. In GMH, the transformation is employed that diagonalizes the adiabatic dipole moment matrix. In FCM, one determines the diabatic states, which are localized as much as possible on molecular fragments corresponding to the donor and the acceptor sites [68]. Once the unitary transformation T is defined, the electronic couplings (off-diagonal matrix elements of the diabatic Hamiltonian) can be calculated as

$$V_{da} = \Sigma T_{id} E_i T_{ia}$$

where E is the diagonal matrix of the adiabatic energies. For a two-state model, the GMH electronic coupling can be expressed via the vertical excitation energy $E_2 - E_1$, the difference of the adiabatic dipole moments $\mu_1 - \mu_2$, and the transition dipole moment μ_{12} [67]

$$V_{da} = \frac{(E_2 - E_1|\mu_{12}|}{\sqrt{(\mu_1 - \mu_2)^2 + 4\mu_{12}^2}}$$

The FCM yields a similar expression [68]

$$V_{da} = \frac{(E_2 - E_1|\Delta q_{12}|}{\sqrt{(\Delta q_1 - \Delta q_2)^2 + 4\Delta q_{12}^2}}$$

Here, Δq_1 and Δq_2 are the difference of charges on the donor and acceptor sites in the ground and excited states, respectively; Δq_{12} is transition charge density.

In the diabatic state (DS) approach, the electronic coupling may be estimated as

$$V_{da} = \frac{H_{da} - S_{da}(H_{da} + S_{da})/2}{1 - S_{da}^2}$$

In this scheme, the diabatic states ψ_d and ψ_a are defined independently, for instance, they may be calculated for separated donor and acceptor. In one-electron approximation, the diabatic states for hole transfer can be approximated with HOMO of neutral base pairs. Matrix elements of Hamiltonian H and the overlap integrals S_{da} are calculated for ψ_d and ψ_a using matrix elements computed in the AO basis for the whole system.

As noted above, these methods provide reliable results when applied to donor and acceptor states that are off-resonance. In Table 5.2, we compare the HT electronic couplings between adjacent pairs calculated using the DS, GMH and FCM methods. As can be seen, the methods yield similar V_{da} values, yet the GMH and FCM estimates are significantly closer to each other.

5.3.3 Intra- and inter-strand electronic couplings

Comparing the electronic couplings between Watson–Crick pairs with the corresponding values estimated for models comprising of just two nucleobases (Table 5.2) one may derive the following conclusions.

Table 5.2 HT couplings between base pairs, V(p–p), in dimers estimated with the DS, GMH and FCM schemes. Electronic couplings of purine bases, V(b–b), calculated for systems consisting of two bases (in meV)

System	Two-base pair model, V(p–p)			Two-base model[a]	
	DS	GMH	FCM	Bases	V(b–b)
[(AT), (AT)]	29.2	27.1	27.1	A–A	30.1
[(AT), (TA)]	53.8	55.8	55.8	A\A	55.8
[(TA), (AT)]	51.9	51.3	51.3	A/A	61.7
[(GC), (GC)]	85.2	81.8	81.2	G–G	83.7
[(GC), (CG)]	20.7	22.0	22.0	G\G	18.7
[(CG), (GC)]	71.8	76.7	76.7	G/G	43.4
[(AT), (GC)]	13.0	10.2	9.0	A–G	9.0
[(GC), (AT)]	128.5	130.2	129.2	G–A	89.4
[(AT), (CG)]	24.1	25.6	24.4	A\G	20.5
[(CG), (AT)]	21.9	26.2	25.7	A/G	4.3

[a] The position of two purine bases B_1 and B_2 in the stack are denoted as: B_1–B_2 for intra-strand arrangement; B_1/B_2 and $B_1\backslash B_2$ for the two inter-strand arrangements.

1. The purine–purine electronic coupling provides the dominant contribution to the hole transfer matrix elements, irrespective whether the nucleobases belong to the same or to opposite strands. For inter-strand couplings the effect of the complementary bases is significantly larger than for intra-strand matrix elements.
2. The V_{da} values exhibit base-order specificity. The electronic coupling between (GC) pairs with intra-strand Gs is significantly larger than the matrix elements between inter-strand guanines. The opposite trend is found for (AT) pairs because of the extraordinarily large inter-strand A/A and A\A couplings.
3. Hole transfer can exhibit a pronounced directional asymmetry. For instance, the coupling matrix in [(GC), (AT)] is about five times larger than that of [(AT), (GC)]; in both systems G and A are in the same strand with the orientations 5′-GA-3′ and 3′-AG-5′, respectively. On the other hand, very similar electronic couplings are found for the systems [(GC), (TA)] and [(CG), (AT)] with distinct inter-strand orientation of G and A.

Detailed analysis of the effects of pyrimidine bases on the HT electronic coupling of G and A bases has been performed [70]. Electrostatic and exchange interactions between the nucleobases in complementary pairs affect remarkably the matrix element V_{da}. Direct inspection of the spatial distribution of the HOMO and HOMO–1 orbitals of base pair complexes provided a qualitative understanding of the electronic coupling and its sensitivity to conformation changes. Also, hydrogen bonding between purine and pyrimidine bases clearly affects the electronic coupling [70]. Recently, systematic DFT calculations have been performed on electronic couplings of nucleobases in DNA stacks [61]. The matrix elements were determined using the diabatic state method. An atomic basis set of Slater-type orbitals of triple-ζ quality including polarization functions was employed. Although quantitative differences exist between V_{da} calculated within the Hartree–Fock or DFT methods, the trends in the values are found to be similar.

5.3.4 Effect of conformational dynamics on V_{da}

Electronic coupling between base pairs is found to be very sensitive to conformational changes of DNA [71,72]. Molecular dynamics simulations have shown that there exist remarkable fluctuations of base-pair position in DNA stacks. The conformational changes affect essentially the electronic couplings. In particular, the HT matrix elements decrease very rapidly with increasing distance between purine bases. For instance, the intra-strand coupling between (AT) pairs changes from 0.074 eV (R = 2.88 Å) to 0.010 eV (R = 3.88 Å) and may be well approximated by the function V(eV) = 0.027 exp[$-\alpha$ (R–3.38)] with α = 2.0 Å^{-1}. A "shift" of an (AT) pair in [(AT), (AT)] by $\pm$0.5 Å causes a considerable increase of the coupling. The helical twist parameter determines the extent of DNA winding. A symmetric structure of [(AT), (AT)], corresponding to completely unwound DNA fragments (twist = 0°), exhibits a very strong coupling between the pairs, V = 0.448 eV, which is considerably larger than V = 0.027 eV found for the regular structure (twist = 36°). A small variation of the twist from 36° to 41° causes the coupling to increase to 0.067 eV. Like the intra-strand interaction, inter-strand couplings are very sensitive to conformational changes. The charge transfer rate depends on the V_{da} squared (Eq. (2)), consequently, the conformation fluctuations can significantly alter the charge transfer efficiency. Troisi and Orlandi [72] estimated the effect of the fluctuations on the effective coupling between guanines separated by four AT pairs. The value of V_{da} increases by several orders of magnitude for a very short period of time (100 fs) when, by coincidence, several individual couplings in the stack are strong. On average, these fluctuations increase the CT rate by two orders of magnitude.

Grozema *et al*. [73] showed that the mobility of holes in ideally ordered poly(G)–poly(C) duplexes can decrease by two orders of magnitude due to twisting of base pairs and disorder in the site energies. For heterogeneous DNA duplexes containing both (GC) and (AT) pairs, the mobility of holes should be much lower because of the static disorder introduced by differences in ionization potentials. Using DFT and semi-empirical methods it has been shown that the main contribution from the molecular vibrations of DNA to its electronic parameters stems from the *transversal* motion consisting of stretching–squeezing the Watson–Crick hydrogen bonds. While twisting and helical pitch deformation mainly change the properties of the π-overlap between stacked base pairs, the deformation of hydrogen bonding drastically affects the separation of the frontier orbitals in the space and thus has a considerable influence on the DNA electronic properties [74,75].

Therefore, the effects of structural dynamics of DNA stacks should be included in computational models of CT in DNA. However, the analysis of static models, as pursued in many studies, remains very helpful for understanding fundamental issues of electron transfer in DNA.

5.3.5 Superexchange paths

Approaches based on perturbation theory may provide useful insights into the bridge effects within the superexchange model of CT in DNA [32,36,67]. However, interference of several paths has to be accounted for even when only one state per nucleobase is included.

The model becomes essentially more complicated when two or more electronic states of each site are considered.

In an alternative method, one invokes a 'divide and conquer' strategy, which relies on an effective Hamiltonian. Model Hamiltonians are very useful in giving insight into physical mechanisms; however, they are somewhat restricted in their predictive value. The effective Hamiltonian for hole transfer in a d-$b_1b_2...b_n$-a system may be defined as follows. A diagonal element H_{ii} is estimated by the relative energy of the state in which the excess charge is localized on the corresponding site. Electronic couplings between neighboring sites give the off-diagonal matrix elements. One obtains the donor–acceptor coupling V_{da} by solving the eigenvalue problem for the effective Hamiltonian or by using the Larsson scheme based on the partitioning technique [76]. Such models represent an approximation because only one or two states per bridge unit are usually included and all the states of higher energy, which also can essentially contribute to the donor–acceptor coupling, are neglected. Below, we analyze the role of different superexchange routes for hole transfer through a DNA π-stack and consider how electronic states of the nucleobase bridges affect the coupling between the donor and acceptor sites.

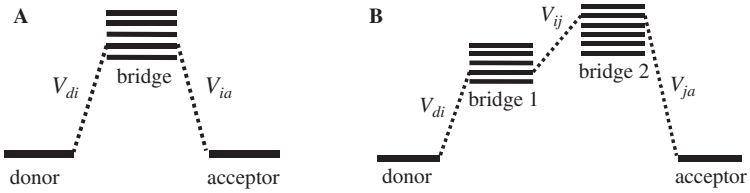
Within the superexchange model, the coupling V_{da} mediated by a bridge can be expressed as a sum over different pathways P_i. In the case of a single bridge (Scheme 5.3A)

$$V_{\mathrm{da}} = \sum_{i=1} P_i, \text{ where } P_i = \frac{V_{\mathrm{di}} V_{i\mathrm{a}}}{E_{\mathrm{t}} - E_i}$$

V_{di} and $V_{i\mathrm{a}}$ are coupling matrix elements between bridge state i and donor or acceptor, respectively, E_{t} is the tunneling energy, and E_i is the energy of the bridge electronic level i. In a system containing two sequential bridge units i and j (Scheme 5.3B), the summation extends over states of each unit

$$V_{\mathrm{da}} = \sum_{i=1} P_i, \sum_{i,j} P_{ij} = \frac{V_{\mathrm{di}} V_{ij} V_{j\mathrm{i}}}{(E_{\mathrm{t}} - E_j)(E_{\mathrm{t}} - E_j)}$$

Here V_{ij} is the electronic coupling between state i of bridge 1 and state j of bridge 2. In addition to the lowest lying electronic state, higher-energy states of the bridge may essentially contribute to V_{da}. As long as the energy gaps $|E_{\mathrm{t}} - E_i|$ and $|E_{\mathrm{t}} - E_j|$ are large and only the lowest virtual levels are accounted for, this model predicts that the rate of electron transfer will decay exponentially with the number of bridge units between donor and acceptor) [32].



Scheme 5.3

Let us consider the role of different electronic states of a (TA) bridge in the double-strand stack GTG, [(GC), (TA), (GC)] [77]. The electronic coupling between Gs is mediated by the (TA) pair. The intra-strand interactions $V(G_1–T)$ and $V(T–G_3)$ as well as inter-strand interactions $V(G_1\backslash A)$ and $V(A/G_3)$ determine the matrix element V_{da}. Table 5.3 lists the contributions of (TA) electronic states to the HT coupling. The lowest lying state 1-A(π) of the bridge is localized on adenine. While this state has a significantly lower energy than the second state 2-T(π) (the energy gaps are 0.45 eV and 1.40 eV, respectively), its contribution to V_{da}, $P_1 = -1.19$ meV, is an order of magnitude smaller than the contribution of the 2-T(π) state, $P_2 = 8.16$ meV. This reflects the weak inter-strand interactions $V(G_1\backslash A)$ and $V(A/G_3)$.

The intra-strand matrix elements are usually stronger than inter-strand couplings. σ-type electronic states of the bridge (states 4 and 9 in Table 5.3) play a very minor role in superexchange and their contributions can be neglected. Higher lying virtual states of the bridge can effectively mediate the coupling. Despite a large energy gap $\Delta_{12} = 5.62$ eV, the contribution of state 12, -1.24 meV, is similar to $P_1 = -1.19$ meV with $\Delta_1 = 0.45$ eV. The destructive interference of P_2 with several other terms P_i (i = 1,3,6,12) leads to V_{da} = 2.8 meV (the sum over 12 bridge states), which is notably smaller than the value of 7.2 meV estimated as $P_1 + P_2$ for the lowest states of the bridge. When 20 states are accounted for, the coupling value of 3.5 meV is in good agreement with V_{da} = 3.9 meV calculated for the whole system GTG using the GMH and FCM methods. From the calculated data one can see that *intra-strand* interactions dominate the donor–bridge–acceptor coupling.

Thus, the predicted V_{da} value considerably depends on the number of electronic levels of the bridge included in the superexchange model. Based on the comprehensive analysis of superexchange in several single- and double-strand DNA stacks [77] one can conclude that

1. HT pathways through thymine bases are characterized by substantial destructive interference while the influence of high-lying electronic states of adenine is found to be relatively weak.

Table 5.3 Interaction of bridge states with donor and acceptor of the duplex GTG of ideal DNA structure: energy gap $\Delta_i = |E_t - E_i|$, matrix elements V_{di} and V_{ia} between adjacent pairs and contribution P_i of bridge state i to the donor–acceptor hole transfer coupling

Bridge state i		Δ_i, eV	V_{di}, meV	V_{ia}, meV	P_i, meV
1	A(π)	0.45	24.2	−22.0	−1.19
2	T(π)	1.40	157.1	74.8	8.39
3	A(π)	2.42	29.0	−45.1	−0.54
4	A(σ)	3.33	5.6	−19.6	−0.03
5	A(π)	3.58	−57.6	6.4	−0.10
6	T(π)	3.48	−112.7	78.8	−2.52
7	A(σ)	4.22	10.5	32.6	0.08
8	T(σ)	4.38	−5.0	33.1	−0.04
9	T(σ)	5.01	−27.7	37.6	−0.21
10	A(π)	5.53	28.4	52.3	0.27
11	A(σ)	5.71	−7.3	43.1	−0.06
12	T(π)	5.62	39.5	−178.0	−1.24

2. The role of different bridge states depends essentially on the stack conformation.
3. Many pathways have to be taken into account for quantitative estimation of electronic couplings. At least three lowest-energy π-states of each nucleobase should be included in the model. However, qualitative results on the superexchange mediated coupling can be derived when only one state per nucleobase is considered.

5.3.6 Parameters for excess electron transfer

The ability of DNA to mediate an excess electron is associated with the formation of radical-anion states of nucleobases. Pyrimidine nucleobases T and C possess the lowest reduction potential [17], and therefore, may function as intermediate states in the ET reaction.

Dipole-bound and covalent states of radical anions. There exist two types of anionic states of bases: (i) electron-dipole-bound anions, in which an excess electron is located far outside the molecule and (ii) valence (or covalently bound) states with the negative charge delocalized over the molecule. The existence of dipole-bound anions was first predicted theoretically by Adamowicz [78,79] and subsequently detected in the gas-phase experiments by the groups of Schermann [80] and Bowen [81,82]. In the gas phase, the radical-anion states of nucleobases may be described as electron-dipole bound states [83–85]. However, the formation of hydrogen bonds between nucleobases and their surroundings stabilizes the valence-bound state [83–85]. For instance, the adenine–thymine (AT) Watson–Crick pair was predicted to form the covalent anion with the positive vertical electron detachment energy [83]. The energetics and structure of radical-anion states of nucleobases and their complexes has been intensively studied theoretically [48–51]. The EA values of the nucleobases were found to decrease in the order T $\approx$ C > A > G. The formation of Watson–Crick pairs does not essentially influence the relative EA values.

Driving force. The effects of adjacent base pairs on the driving force ΔG^0 for electron transfer between bases B and B′

$$\ldots X\ B^-\ Y\ldots\ X'\ B'\ Y'\ldots \rightarrow \ldots X\ B\ Y\ldots\ X'\ B'^-\ Y'\ldots$$

has been estimated as the differences of EA of B and B′ in trimers of base pairs 5′-XBY-3′ [86]. The *subsequent* base Y exerts a stronger influence than the *preceding* base. EA of B decreases when bases X and Y change in the order C ~ T>A>G . The strongest electron traps are predicted to be 5′-XCY-3′and 5′-XTY-3′, where X, Y are pyrimidine bases. Thus, the efficient ET may be expected in DNA duplexes, where the pyrimidine bases C and T are on the same strand, 5′-Pyr_i –...–Pyr_j– ...–Pyr_k-3′. In this case, the T and C bases exhibit very similar EAs and the corresponding radical-anion states are nearly in resonance, favoring excess electron migration. EA of T and C in stacks becomes considerably smaller when adjacent bases are purines. Adenine and guanine do not act as electron traps in DNA because of their essentially smaller EA values. The dependence of ΔG^0 on the sequence of a DNA stack affects the probability of electron transport. Experimentally, it has been demonstrated that an excess electron can move over a distance of 30 Å between donor and acceptor [87]. Because experimental studies have provided conflicting results on the ET efficiency [18], theoretical insights into the mechanistic and dynamic issues of excess electron transport are of special interest. However, this cannot be achieved without estimating electronic coupling matrix elements for excess electron transfer.

Electronic coupling. Recently, an efficient strategy has been discussed for calculating the ET matrix elements [88]. In a stack consisting of two base pairs, the diabatic donor and acceptor states for excess electron transfer can be approximated by radical-anion states of the separated base pairs. In one-electron approximation, a state of the excess electron is described by HOMO of the radical anion.

Figure 5.1(A) compares HOMOs of an AT radical anion calculated with the standard and extended basis sets. As seen, both calculations provide very similar pictures: the wave functions describe a covalently bound anion with an excess electron localized on thymine. The excess electron state in the radical anion can also be approximated by LUMO of the neutral AT base pair. As shown in Fig. 5.1(B), the calculated shape of LUMO depends dramatically on the basis set. When diffuse functions are explicitly excluded (6-31G* basis), the LUMO of the neutral system closely resembles HOMO of the radical anion (Fig. 5.1(A)), and therefore, should properly describe the state of an excess electron. By contrast, the LUMO calculated with diffuse functions (6-31++G** basis set) predicts a dipole-bound state, which drastically differs from HOMO of the radical anion. Evidently, these wave functions cannot be employed to calculate the coupling matrix element for excess electron transfer. Thus, LUMOs of neutral DNA π-stacks may be used to approximate ET diabatic states only when diffuse functions are explicitly excluded from the calculation. Similar results are obtained for a GC base pair. The GC pair forms a covalently bound anion with an excess electron localized on cytosine. The Hartree–Fock results on the charge distribution in AT and GC radical-anions are in good agreement with the DFT data of Richardson *et al*. [50,51]. The spatial separation of the frontier orbitals in an infinite poly(dA)–poly(dT) duplex including its hydration shell and sodium counter-ions has recently been analyzed on the bases of the extended Hückel and PM3 semi-empirical methods [89].

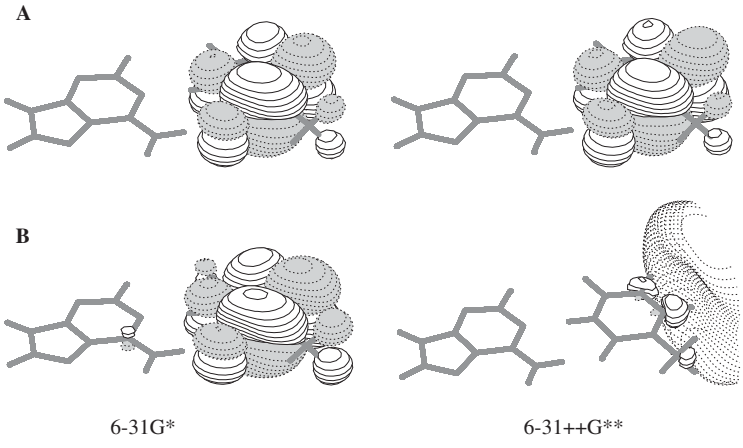


Fig. 5.1. (A) HOMO of the AT radical anion; (B) LUMO of the neutral AT base pair. The Hartree–Fock calculations were performed with the standard 6-31G* (on the left) and extended 6-31++G** (on the right) basis sets.

Recently, Li and Sevilla [48] showed that DFT calculations employing basis sets without diffuse functions produce reasonable estimates of the relative valence electron affinities.

Similar results are provided by calculations of base-pair stacks (Fig. 5.2). HF/6-31++G** calculations suggest that the excess electron in the stacks of base pairs forms covalently bound radical anions where the negative charge is almost completely confined to pyrimidine bases. LUMO and LUMO+1 of the neutral stack [(AT), (AT)] calculated with HF/6-31G* (Fig. 5.2(A)) describe covalently bound anion states delocalized over thymine bases. However, calculations with the extended basis set give LUMOs (Fig. 5.2(B)) corresponding to the dipole-bound states with a diffuse charge distribution around thymines (LUMO) and adenines (LUMO+1).

In isolated π-stacks, the diabatic states for excess electron transfer are predicted to be covalently bound anion states. Accounting for the polar environment should further stabilize the covalent states and suppress the formation of electron-dipole-bound states. Since diffuse functions are not very important for proper description of an excess electron in the covalently-bound anions they may be excluded from basis sets when radical anion states are treated and they have to be excluded if diabatic states for ET are approximated by LUMOs of neutral systems. This assumption is justified below by comparing ET couplings estimated using different computational approaches.

The matrix elements V_{da} were calculated for several complexes of base pairs employing the diabatic state (DS) method. In systems [(AT), (AT)], [(GC), (AT)], and [(GC), (GC)], the donor and acceptor sites for the excess electron (pyrimidine bases T and C) are on the same strand.

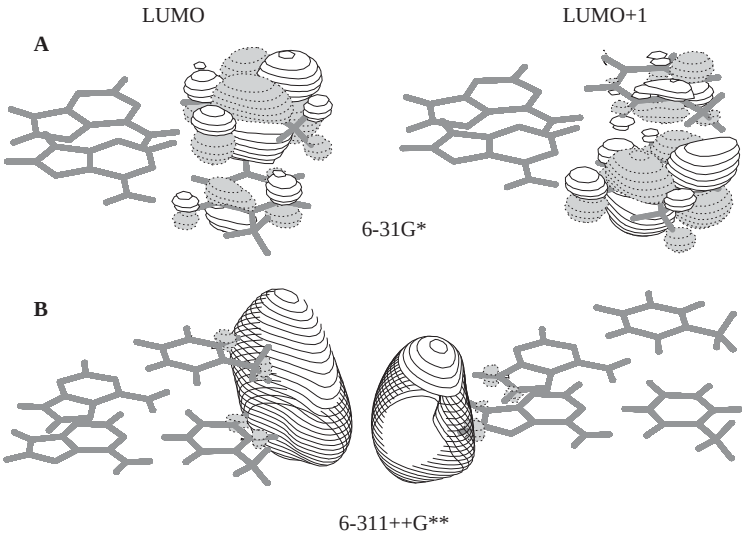


Fig. 5.2. LUMO and LUMO+1 of the neutral stack [(AT), (AT)] calculated with different basis sets: (A) HF/6-31G* calculation; (B) HF/ 6-31++G** calculation.

Table 5.4 Electronic couplings of AT and GC base pairs for excess electron transfer as calculated by the DS method for anion-radical states and by GMH and FCM for neutral systems (in eV)

Complex	DS method		GMH	FCM
	6-31++G**	6-31G*	6–31G*	
[(AT), (AT)]	0.034	0.025	0.029	0.030
[(GC), (AT)]	0.066	0.113	0.107	0.107
[(GC), (GC)]	0.078	0.1092	0.119	0.119
[(AT), (TA)]	0.022	0.018	0.023	0.023
[(GC), (TA)]	0.019	0.017	0.019	0.018
[(GC), (CG)]	0.016	0.017	0.019	0.019

In Table 5.4, we compare the couplings calculated with the standard and extended basis sets. The matrix element V_{da} for [(AT), (AT)], ~ 0.03 eV, is remarkably smaller than V_{da} found for [(GC), (AT)] and [(GC), (GC)] , ~ 0.10 eV. The calculated values depend on the employed basis sets: exclusion of diffuse functions leads to some decrease of the coupling for [(AT), (AT)], while the V_{da} values become larger in [(GC), (AT)] and [(GC), (GC)]. However, these differences are considerably smaller than the changes in V_{da} caused by conformational fluctuations of the π-stacks [88]. Because the radical-anion states of AT and GC are well approximated by LUMOs of the neutral base pairs calculated with no diffuse functions, these molecular orbitals may be employed to estimate ET coupling matrix elements. The GMH and FCM schemes give very similar values of V_{da} (Table 5.4). The V_{da} matrix elements estimated using LUMOs are in good agreement with the couplings obtained by the DS method for radical anion states of π-stacks.

Thus, electronic couplings for excess electron transfer in DNA can be reasonably well estimated in a similar manner to that used to calculate hole transfer matrix elements. Because the relative energetics for ET can also be estimated using small basis sets [48], modeling of negative charge migration in DNA can be performed using the Hartree–Fock and DFT methods with small basis sets. However, an extension of the basis set by including diffuse functions will lead to incorrect estimates of V_{da} based on the calculated results for neutral systems. The ET matrix elements for models containing intra-strand T and C bases are essentially larger than the couplings between inter-strand pyrimidine bases. In general, ET couplings were found to be considerably smaller than the corresponding values for HT [88].

5.4 CHARGE TRANSFER IN DNA STACKS AND CHROMOPHORE–DNA COMPLEXES

5.4.1 Quantum chemical study of DNA π-stacks

The first quantum chemical study of long-range HT in aperiodic DNA was carried out by Priyadarshy *et al*. [52]. The electronic coupling was determined from a CNDO/2 calculation that explicitly includes electron-transfer mediation arising from all valence orbitals of the

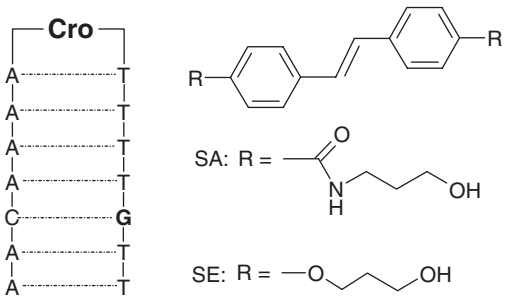
system. It was shown that the π-stacking interaction of nucleobases plays a dominant role in providing electronic couplings while the contribution of the sugar–phosphate backbone in V_{da} is small. Most important was the conclusion that the experimentally found long-range charge migration in DNA cannot be described in terms of superexchange tunneling.

Recently, Beratan and co-workers [90] have performed a comprehensive study of HT rates in DNA stacks including estimation of electronic couplings and reorganization energies. Systems consisting of AT sequences between two guanines or two 7-deaza-guanine (Z) were considered. The tunneling energy dependence of the donor–acceptor coupling has been analyzed in great depth. Because Z is easier to oxidize than G by 0.5 eV, the tunneling energy gap (the difference in energy of the bridge and *d–a* states) in (ZC) $(AT)_n$(ZC), Δ=0.7–0.8 eV, is substantially larger than in (GC)$(AT)_n$(GC), Δ= 0.2–0.3 eV. The values of β_{el} corresponding to decay of V_{da} were computed to be 1.46 and 0.95 $Å^{-1}$, respectively. Thus, the difference in the distance dependence of the HT process between donor and acceptor was attributed to the tunneling energy effect. Thus, the electron transfer rates depend critically on the tunneling energy gap Δ. Furthermore, it has been shown that Δ values estimated for isolated systems must be increased by half the ET reorganization energy. This shift results in a substantial increase in β_{el}. Because the predicted values of β are much larger than observed experimentally it was concluded that switching from the superexchange to an alternative HT mechanism takes place when more than two bridging AT base pairs separate the d–a sites. Indeed, the calculated HT rates in DNA across one and two intervening base pairs can be well described as single-step tunneling. However, for longer bridges the computed rates are much smaller than observed, and therefore, an alternative mechanism that implies thermal population of intermediate hole states should be considered.

5.4.2 Photoinduced charge separation in DNA hairpins

Beljonne *et al*. [38] have addressed the distance dependence of the HT matrix element for charge injection in the model hairpins on the chemical structure of the chromophore (see Scheme 5.4). The position of the GC pair was varied along the stack. A number of chromophores (including two different derivatives of stilbene) were studied.

Photoinduced hole transfer from the chromophore (donor) to the guanine (acceptor) was considered on the basis of semi-empirical INDO/S calculations. The initial d* a and



Scheme 5.4

final $d^+ a^-$ states in the charge separation process were described with a multi-configurational approach (CI singles). The calculations took into account all possible pathways for electron transfer. The electronic characteristics of d* a and $d^+ a^-$ states were used within the GMH scheme to estimate the coupling matrix element. It was shown that the nature of the donor has a significant influence on the distance dependence of the photo-induced hole transfer to guanine. For instance, the β values in the hairpins possessing the SA and SE chromophore (Scheme 5.4) are computed to be ~ 0.4 $Å^{-1}$ and 0.7 $Å^{-1}$, respectively. Thus, efficiency for hole injection depends on the energy gap between the donor and bridge electronic levels [38].

5.5 CHARGE LOCALIZATION IN DBA Π-STACKS

An important issue that has not been fully resolved yet is charge delocalization over neighboring base pairs. The description of charge transfer depends on the fact of whether charge carrier states are delocalized or not. The more extended the wave function, the better the conductance of the stack. In the limiting case of complete charge delocalization, the electron transfer is a coherent process with virtually no distance dependence [29,30]. Delocalization of the hole over several nucleobases is assumed in several CT mechanisms [10,91]. Moreover, different methods must be employed to describe theoretically propagation of localized and delocalized states. There has been an interesting discussion as to whether the hole charge in DNA is confined to a single base pair or delocalized over several adjacent base pairs (see, for instance, Ref. 10 and references therein). Recent computational studies have provided conflicting results.

1. Based on a simple cylindrical cavity model, where the charge was placed on the axis of the cylinder, Basko and Conwell [92] accounted for the solvation effects in the tight-binding Hamiltonian and concluded that the hole charge is spread over five or more adjacent base pairs. The electron-hole wave function calculated by using this model is similar to that found for the isolated DNA stack.
2. By contrast, Beratan and co-workers [93] showed that the interaction with surroundings considerably affects the charge distribution in DNA. They employed a heterogeneous dielectric model comprised of a DNA zone and a solvent zone. The solvation effects essentially favor localization of the hole, while hole states delocalized over two or three guanines are found to be energetically feasible [93].

In a stack containing several adjacent (GC) base pairs, delocalization of the hole charge over Gs is controlled by the interplay between the redox potentials of the nucleobases, the internal relaxation energy caused by the electron detachment, the solvation energy and the electronic coupling [94]. A scheme has been proposed for estimating the solvation effects for charge transfer in DNA within tight-binding Hamiltonians [95]. The corresponding linear response parameters were derived from the solvation energies of DNA stack calculated for different charge distributions in the stack. At variance with previous results [92], it was concluded that the hole states in $(GC)_n$ sequences are localized on individual guanine bases. The solvation effects essentially suppress the charge delocalization in DNA leading to radical cation states confined to single Gs.

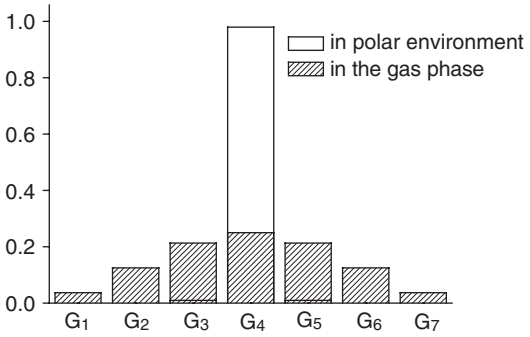


Fig. 5.3. The effect of solvation on the hole delocalization** in the stack $(GC)_7$.

The influence of the polar environment on the excess charge delocalization in DNA is shown in Fig. 5.3. The hole charge is well delocalized over the π-stack when the solvation effects are neglected. However, in the polar environment, the charge is almost completely, ~98 per cent, confined to the nucleobase G_4, the remainder of the charge is found on the nearest sites G_3 and G_5, while all the more distant Gs do not carry any charge. It should be emphasized that the hole can be localized on any guanine G_i within the stack. The calculated energies of state where the charge is localized on G_i (i ranges from 1 to 7) are within 0.1 kT. Thus, the solvation effects lead to very localized hole states which can reside on any single G base in $(GC)_n$ stacks. This result remains unchanged when considerable deviations of the electronic coupling from its average value are accounted for [95].

Recently, a combined QM/MD study has been performed on a poly(dA)–poly(dT) duplex to consider the influence of DNA dynamics on the hole delocalisation [56]. MD predicts that individual bases exhibit significant freedom of movement leading to large fluctuations in electronic couplings. The off-diagonal dynamical disorder results in charge localization of HOMO and LUMO of the stacks. Moreover, a large degree of localization for the wave functions is found for the states that are energetically close to the HOMO and LUMO. The off-diagonal disorder should reinforce the localization of an excess charge caused by the polar environment.

The internal (structural) reorganization of nucleobases caused by an excess charge is another factor that suppresses the hole delocalization over neighboring nucleobases. Olofsson and Larsson [96] considered this effect in detail and found well-localized hole states to be energetically stabilized due to internal reorganization of nucleobases.

5.6 CONCLUDING REMARKS

In this chapter, we have considered the dependence of the ET efficiency on the structure and electronic properties of systems consisting of donor and acceptor sites connected by a DNA π-stack. The results discussed here indicate some limitations of simple models used for calculating the donor–acceptor electronic coupling. In particular, the schemes that account for one or two states per base pair cannot provide quantitative estimates of V_{da} because the bridge states of higher energy considerably contribute to the coupling. The interference of several

superexchange pathways leads to essential changes in V_{da} depending on the structure and conformation of the bridge.

The structure and dynamics of solvated counter-ions play a very important role in CT process through DNA. Indeed, calculations of DNA fragments without a proper description of solvation effects suggest that an excess charge in a π-stack must be delocalized over several base pairs. However, the charge should be confined to a single base due to solvation and internal reorganization effects. Neglecting the solvation effects may result in misleading assumptions on mechanisms of charge transfer. Furthermore, the solvation effects increase the tunneling energy gap and thereby decrease the effective donor–acceptor coupling. The dynamics of surrounding counter-ion and water molecules lead to considerable fluctuations of both the driving force for CT and the tunneling gap.

Hole transfer to a guanine base may initiate proton transfer from the radical cation to its surroundings. The loss of a proton can, in turn, stop hole transfer through DNA. The interrelation between the hole and proton transfer in DNA has been experimentally demonstrated by Giese and Wessely [97]. Stabilization of the G^+ state by a proton shift from G^+ to C within the GC pair was found in DFT molecular dynamics calculations [44]. A general theoretical approach for treatment of proton-coupled electron transfer reactions was described by Hammes-Schiffer [98].

Thus, several factors, including π-stack dynamics, proton transfer coupled to ET, and the effects of surroundings, play important roles in distinguishing DNA-mediated charge transport. A very large computational effort is required when one aims to account for all the effects associated with charge transfer in DNA. Important and very promising steps on this road have already been made. Based on the DFT method Parrinello and co-workers [44] have performed the first-principles dynamic calculations on extended systems related to charge transfer in DNA. As semi-empirical methods are by several orders of magnitude faster than *ab initio* and DFT calculations and they have already proved to be very useful for treatment of CT in DNA, elaboration of a special parameterization to improve the performance of these computational schemes would be extremely useful.

Much experimental effort has been made recently to improve the conductance properties by combining DNA with transition metals in various ways [99]. Computational modeling of charge transfer in such systems appears to be a very attractive and challenging task for theoretical chemists.

5.7 REFERENCES

1 J.D. Watson and F. H. C. Crick, Nature, 171 (1953) 737.
2 A.S. Gyorgi, Proc. Natl. Acad. Sci., 46 (1960) 1444.
3 D.D. Eley and D.I. Spivey, Trans. Faraday Soc., 58 (1962) 411.
4 C.J. Murphy, M.R. Arkin, Y. Jenkins, N.D. Ghatlia, S.H. Bossmann, N.J. Turro and J.K. Barton, Science, 262 (1993) 1025.
5 A.Y. Kasumov, M. Kociak, S. Gueron, s., B. Reulet, V.T. Volkov, D.V. Klinov and H. Bouchiat, Science, 291 (2001) 280.
6 D. Porath, A. Bezryadin, S. deVries and C. Dekker, Nature (London), 403 (2000) 635.
7 C. Dekker and M.A. Ratner, Physics World, 29 (2001) 33.
8 R.G. Endres, D.L. Cox and R.R.P. Singh, Rev. Mod. Phys., 76 (2004) 195.
9 P.J. de Pablo, F. Moreno-Herrero, J. Colchero, J. Gómez Herrero, P. Herrero, A.M. Baró, P. Ordejón, J.M. Soler and E. Artacho, Phys. Rev. Lett., 85 (2000) 4992.

10 G.B. Schuster (Ed.), Long-range charge transfer in DNA. Topics in current chemistry, Springer, Berlin, 2004, Vol. 237, P. 238.
11 P.J. Dandliker, R.E. Holmlin and J.K. Barton, Science, 275 (1997) 1465.
12 S.O. Kelley and J.K. Barton, Science, 283 (1999) 375.
13 G.B. Schuster, Acc. Chem. Res., 33 (2000) 253.
14 B. Giese, Acc. Chem. Res., 33 (2000) 631.
15 F.D. Lewis, R.L. Letsinger and M.R. Wasielewski, Acc. Chem. Res., 34 (2001) 159.
16 S. Steenken and S.V. Jovanovic, J. Am. Chem. Soc., 119 (1997) 617.
17 C.A.M. Seidel, A. Schulz and M.H.M. Sauer, J. Phys. Chem., 100 (1996) 5541.
18 H.A. Wagenknecht, Angew. Chem. Int. Ed., 42 (2003) 2454.
19 S.J. Park, T.A. Taton and C.A. Mirkin, Science, 295 (2002) 1503.
20 A.P. Alivisatos, K.P. Johnsson, T.E. Wilson, C.J. Loveth, M.P. Bruchez and P.G. Schulz, Nature, 382 (1996) 609.
21 E.Winfree, F. Liu, L.A. Wenzler and N.C. Seeman, Nature, 394 (1998) 539.
22 B. Armitage, Chem. Rev., 98 (1998) 1171.
23 C.J. Barrows and J.G. Muller, Chem. Rev., 98 (1998) 1109.
24 J.K. Barton, Pure & Appl. Chem., 70 (1998) 873.
25 J. Jortner, M. Bixon, T. Langenbacher and M.E. Michel-Beyerle, Proc. Natl. Acad. Sci. USA, 95 (1998) 95, 12759.
26 M. Bixon, B. Giese, S. Wessely, T. Langenbacher, M.E. Michel-Beyerle and J. Jortner, J. Proc. Natl. Acad. Sci. USA, 96 (1999) 11713.
27 M. Bixon and J. Jortner, J. Am. Chem. Soc., 123 (2001) 12556.
28 M.A. Ratner, Nature, 397 (1999) 480.
29 Y.A. Berlin, A.L. Burin and M.A. Ratner, J. Am. Chem. Soc., 123 (2001) 260.
30 F.C. Grozema, Y.A. Berlin and L.D.A. Siebbeles, J. Am. Chem. Soc., 122 (2000) 10903.
31 B. Giese, J. Amaudrut, A.K. Köhler, M. Spormann and S. Wessely, Nature, 412 (2001) 318.
32 M.D. Newton, Chem. Rev., 91 (1991) 767.
33 Y.A. Berlin, I.V. Kurnikov, D.N. Beratan, M.A. Ratner and A.L. Burin, in G.B. Shuster (Ed.), Long-Range charge transfer in DNA. Topics in current chemistry, Springer, Berlin, 2004. Vol. 237, PP. 1–36.
34 R.A. Marcus and N. Sutin, N. Biochim. Biophys. Acta, 811 (1985) 265.
35 A.A. Voityuk, N. Rösch M. Bixon and J. Jortner, J. Phys. Chem. B, 104 (2000) 9740.
36 A.A. Voityuk, J. Jortner, M. Bixon and N. Rösch, J. Chem. Phys., 114 (2001) 5614.
37 J. Jortner, M. Bixon, A.A. Voityuk and N. Rösch, J. Phys. Chem. A, 106 (2002) 7599.
38 D. Beljonne, G. Pourtois, M.A. Ratner and J.L. Bredas, J. Am. Chem. Soc., 125 (2004) 14510.
39 H.L. Tavernier and M.D. Fayer, J. Phys. Chem. B, 104 (2000) 11541.
40 K. Siriwong, A.A. Voityuk, M.D. Newton and N. Rösch, J. Phys. Chem. B, 107 (2003) 2595.
41 D.N. LeBard, M. Lilichenko, D.V. Matyushov, Y.A. Berlin and M.A. Ratner, J. Phys. Chem. B, 107 (2003) 14509.
42 S. Tanaka and Y. Sengoku. Phys. Rev. E, 68 (2003) 031905.
43 F.L. Gervasio, P. Carloni and M. Parrinello, Phys. Rev. Lett., 89 (2002) 108102.
44 F.C. Gervasio, A.Laio and M. Parrinello, Phys. Rev. Lett., 94 (2005) 158103.
45 I. Saito, T. Nakamura, K. Nakatani, Y. Yoshioka, K. Yamaguchi and H. Sugiyama, J. Am. Chem. Soc., 120 (1998) 12686.
46 Y.Yoshioka, Y. Kitagawa, Y. Tukano, K. Yamaguchi, T. Nakamura and I. Saito, J. Am. Chem. Soc., 121 (1999) 8712.
47 K. Senthilkumar, F.C. Grozema, C.F. Guerra, F.M. Bickelhaupt and L.D.A. Siebbeles, J. Am. Chem. Soc., 125 (2003) 13658.
48 X. Li and M.D. Sevilla, J. Phys. Chem. A, 106 (2002) 1596.
49 N. Russo, M. Toscano and A. Grand, J. Comput. Chem., 14 (2000) 1243.
50 N.A. Richardson, S.S. Wesolowski and H.F. Schaefer III, J. Am. Chem. Soc., 124 (2002) 10163.
51 N.A. Richardson, S.S. Wesolowski and H.F. Schaefer III, J. Phys Chem. B, 107 (2003) 848.
52 S. Priyadarshy, S.M. Risser and Beratan, J. Phys. Chem., 100 (1996) 17678.
53 E.B. Starikov, Mod. Phys. Lett. B, 18 (2004) 825.
54 E.B. Starikov, Phil. Magaz. Lett., 83 (2003) 699.

55 J.P. Lewis, J. Pikus, Th.E. Cheatham III, E.B. Starikov, H. Wang, J. Tomfohr and O.F. Sankey, Phys. Stat. Sol. B, 223 (2002) 90.
56 J.P. Lewis, Th. E. Cheatham III, E.B. Starikov, H. Wang and O.F. Sankey, J. Phys. Chem. B, 107 (2003) 2581.
57 A.A.Voityuk, K. Siriwong and N. Rösch, Angew. Chem. Int. Ed., 43 (2004) 624.
58 X.J. Lu, M.A. El Hassan and C.A. Hunter, J. Mol. Biol., 273 (1997) 681.
59 L. Clowney, S.C. Jain, A.R. Srinivasan, J. Westbrook, W.K. Olson and H.W. Berman, J. Am. Chem. Soc., 118 (1996) 509.
60 A.A. Voityuk, J. Jortner, M. Bixon and N.Rösch, Chem. Phys. Lett., 324 (2000) 430.
61 K. Senthilkumar, F.C. Grozema, C.F. Guerra, F.M. Bickelhaupt, F.D. Lewis, Y.A. Berlin, M.A. Ratner and L.D.A. Siebbeles, J. Am. Chem. Soc.**,** 127 (2005) 14894.
62 F.D. Lewis, J.Q. Liu, X.B. Zuo, R.T. Hayes and M.R. Wasielewski, J. Am. Chem. Soc., 125 (2003) 4850.
63 A.A. Voityuk and N. Rösch, J. Phys. Chem. B, 106 (2002) 3013.
64 S. Yokojima, W. Yanoi, N. Yoshiki, N. Kurita, S. Tanaka, K. Nakatani and A. Okada, J. Phys. Chem. B, 108 (2004) 7500.
65 R.N. Barnett, C.L. Cleveland, A. Joy, U. Landman and G.B. Schuster, Science, 294 (2001) 567.
66 R.N. Barnett, C.L. Cleveland, U. Landman, E. Boone, S. Kanvah and G.B. Schuster, J. Phys. Chem. A, 107 (2003) 3525.
67 R.J. Cave and M.D. Newton, J. Chem. Phys., 106 (1997) 9213.
68 A.A. Voityuk and N. Rösch, J. Chem. Phys., 117 (2002) 5607.
69 N. Rösch and A.A. Voityuk, in G.B. Shuster (Ed.), Long-range charge transfer in DNA. Topics in current chemistry, Springer, Berlin, 2004, Vol. 237, PP. 37–72.
70 J. Rak, A.A. Voityuk, A. Marquez and N. Rösch, J. Phys. Chem. B, 106 (2002) 7919.
71 A.A.Voityuk, K. Siriwong and N. Rösch, Phys. Chem. Chem. Phys., 3 (2001) 5431.
72 A. Troisi and G. Orlandi, J. Phys. Chem. B, 106 (2002) 2093.
73 F.C. Grozema, L.D.A. Siebbeles, Y.A. Berlin and M.A. Ratner, Chem. Phys. Chem., 3 (2002) 536.
74 S.S. Alexandre, E. Artacho, J.M. Soler and H. Chacham, Phys. Rev. Lett., 91 (2003) 108105.
75 D. Hennig, E. Starikov, J.F.R. Archilla and F. Palmero, J. Biol. Phys., 30 (2003) 227.
76 S. Larsson, J. Am. Chem. Soc., 103 (1981) 4034.
77 A.A.Voityuk and N. Rösch, Isr. J. Chem., 117 (2004) 5607.
78 N.A. Oyler and L. Adamowicz, Chem. Phys. Lett., 219 (1994) 223.
79 L. Adamowicz, J. Phys. Chem., 97 (1993) 11122.
80 C. Desfrançois, H. Abdoul-Carime and J.P. Schermann, J. Chem. Phys., 102 (1996) 7792.
81 J.H. Hendricks, S.A. Lyapustina, H.L. de Clercq, J.T. Snodgrass and K.H. Bowen, J. Chem. Phys., 104 (1996) 7788.
82 J.H. Hendricks, S.A. Lyapustina, H.L. de Clercq and K.H. Bowen, J. Chem. Phys., 108 (1998) 8.
83 I. Al-Jihan, J. Smets and L. Adamowicz, J. Phys. Chem. A, 104 (2000) 2994.
84 L. Adamowicz, J. Phys. Chem., 100 (1996) 14655.
85 D.M. Smith, Y. Elkady and L. Adamowicz, J. Phys. Chem. A, 103 (1999) 5784.
86 A.A. Voityuk, M. E. Michel-Beyerle and N. Rösch, Chem. Phys. Lett., 342 (2001) 231.
87 T. Carell, C. Behrens and J. Gierlich, Org. Biomol. Chem., 1 (2003) 2221.
88 A.A. Voityuk, J. Chem. Phys., 123 (2005) 034903.
89 E.B. Starikov, Phys. Chem. Chem. Phys., 4 (2002) 4523.
90 G.S.M. Tong, I.V. Kurnikov and D. N. Beratan, J. Phys. Chem. B, 106 (2002) 2381.
91 E. Conwell, in G.B. Shuster (Ed.) Long-range charge transfer in DNA. Topics in current chemistry, Springer, Berlin, 2004., Vol. 237, p. 73.
92 D.M. Basko and E.M. Conwell, Phys. Rev. Lett., 88 (2002) 098102.
93 I.V. Kurnikov, G.S.M. Tong, M. Madrid and D.N. Beratan, J. Phys. Chem. B, 106 (2002) 7.
94 A.A. Voityuk, J. Phys. Chem. B, 109 (2005) 10793.
95 A.A. Voityuk, J. Chem. Phys., 122 (2005) 204904.
96 J. Olofsson and S. Larsson, J. Phys. Chem. B, 105 (2001) 10398.
97 B. Giese and S. Wessely, Chem. Commun., 20 (2001) 2108.
98. S. Hammes-Schiffer, Acc. Chem. Res., 34 (2001) 273.
99. J. Richter, Physica E (Amsterdam), 16 (2003) 157.

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 6

Solvent effects on biomolecular dynamics simulations: A comparison between TIP3P, SPC and SPC/E water models acting on the Glucocorticoid receptor DNA-binding domain

Johan Bredenberg, Pekka Mark and Lennart Nilsson

Karolinska Institutet, Department of Biosciences, Center for Structural Biochemistry, Hälsovägen 7, SE-141 57 Huddinge, Sweden

Abstract

Three widely used solvent models (TIP3P, SPC and SPC/E) have been compared in the context of molecular dynamics simulations performed on the Glucocorticoid receptor DNA-binding domain (GR DBD), using eight different initial coordinate sets. The analysis involved hydration properties, protein dynamics, entropy, order parameters, reorientational diffusion times and thermal fluctuations of non-hydrogen protein atoms. The three latter admit a direct comparison with experimental observations. It was found that only reorientation times showed significant solvent dependence.

6.1 INTRODUCTION

Nuclear receptors belong to a superfamily of ligand-induced transcription factors [1] which are important for cell differentiation, growth and metabolism [2,3]. They share a common DNA-binding domain motif (DBD) (see [4] for a review) with a high degree of homology and structural similarity. DBD is globular and built of about 70–80 amino acids. Two perpendicularly packed helix loop regions form the core protein. Eight out of nine invariant cysteine residues are in a tetrahedral coordination by two zinc ions that form two distinct Cys_4 zinc fingers. The N-terminal finger directs three residues, the P-box, to recognize a specific DNA sequence, while the C-terminal finger is involved in the protein–protein dimerization and non-specific DNA contacts. For the Glucocorticoid receptor DBD (in short GR DBD), a wealth of experimental [5–11] and theoretical

[12–15] studies exist in the literature, which therefore makes the GR DBD a good system for exploring theoretical models that may aid in understanding observations made from experiments.

An important feature in molecular dynamics (MD) simulations of biomolecules is the inclusion of explicit solvent atoms. Three widely used water models are the transferable interaction potential (TIP3P) [16], the simple point charge (SPC) [17] and the simple point charge extended (SPC/E) [18]. They are three centre models with a single point charge at each nucleus, but differ in charge distributions, van der Waals parameters and geometry. This study is a follow-up on previously reported comparisons between these three water-models used in MD on pure solvent and on smaller systems [19–21]. Solvent viscosity [22], translation and rotational diffusion for proteins in SPC and SPC/E have also been reported [23]. In this work we address two main issues: physical effects upon changing the solvent model in terms of protein stability, dynamics, energy and entropy changes in the solute upon using the TIP3P, SPC or SPC/E solvent model, and coordinate-dependent sampling of the conformational space. For this purpose, eight sets of coordinates were drawn from a 20 ns simulation of a fully hydrated GR DBD monomer. Each of these eight structures was then subjected to 2 ns MD simulations using TIP3P, SPC and SPC/E for modelling the solvent. Thus a total of 24 MD simulations were carried out for a total 48 ns.

6.2 METHODS

Eight coordinate sets were extracted from a previous [14] 20 ns MD simulation done on the crystal GR DBD monomer [10] at $t = 0$, 2, 4, 6, 10, 15, 18 and 20 ns with the solute centred in a water sphere (TIP3P, SPC, SPC/E) of radius 30 Å and four chloride ions to render overall charge neutrality. The shortest distance between any of the solute atoms and the sphere edge was at least 11 Å, and the total number of atoms in each coordinate set was 10 398. Prior to any molecular dynamics simulation on these structures, 100 steps of steepest descent minimization with harmonic constraints of force constant 5 kcal mol^{-1} $Å^{-2}$ on solute non-hydrogen atoms were done, followed by 200 steps of adopted basis Newton–Raphson (ABNR) minimization. Then, the force constant was gradually reduced by 1 kcal mol^{-1} $Å^{-2}$, followed by another 200 steps of ABNR, until the constraint forces were zero.

All 24 simulations (each of 2 ns duration) were carried out with the CHARMM [24] program, the CHARMM22 all-atom force field [25] and the modified thiolate to zinc non-bonded interaction potential [26,27]. We used a spherical geometry, with stochastic boundary molecular dynamics [28] at 300 K, which in previous studies [29,30] has been shown to yield comparable results as periodic boundary conditions. The temperature was regulated with a heat bath, and a frictional constant, β, of 50 ps^{-1} was assigned to oxygen water atoms. SHAKE [31] was applied on explicit bonds to hydrogen atoms, allowing a 2 fs time-step in the numerical integration, for which atoms within a radius of 27 Å were treated according to Newton's laws using the leap-frog algorithm, while the remaining atoms were treated as Langevin particles. Electrostatic interaction energies and forces were smoothly shifted to zero at 12 Å, using an atom-based spherical cut-off [32] with a non-bonded list for atoms within 14 Å. This list was regenerated whenever an atom was displaced > 1 Å since the last update.

For comparison, two trajectories generated in the isothermal-isobaric ensemble (NPT) and the microscopic ensemble (NVE) were analysed by means of rotational times. A detailed description of the simulation protocol has been given elsewhere [14], but for easy review the protocol is described here. The second monomer of the 1glu crystal structure was solvated in a TIP3P water box with initial dimensions 53×51×54 Å^3. Water molecules closer than 2.8 Å to the solute or the crystal waters were deleted and four chloride ions were added for charge neutrality. All electrostatic and Lennard–Jones interaction energies and forces were truncated at 13 Å, and long-range electrostatic energies and forces beyond 13 Å were treated with the particle mesh Ewald summation method from Darden and co-workers [33]. We used a lattice grid of 48 points extending in the x, y and z directions, and a sixth-order interpolation spline function for calculating the reciprocal summation. The kappa value was set to 0.4. The system was equilibrated for 40 ps at 298.15 K and targeted an isotropic pressure of 1.0 bar, while all C_α atoms were harmonically constrained to their initial positions with a force constant of 1.0 kcal mol^{-1} Å^{-2}. A time step of 2 fs was used and coordinates were stored for analysis every 0.4 ps. SHAKE [31] was applied on all explicit bonds to hydrogen atoms. The temperature was controlled with a heat bath and the pressure was regulated with a Langevin piston mass of 500 atom mass units and a collision frequency, γ, of 20 ps^{-1}. After 40 ps, the harmonic constraints on the C_α atoms were removed and one simulation, carried on keeping the pressure with the Langevin piston, i.e., the NPT ensemble, while another trajectory was generated at constant volume, i.e., without a mass on the piston, thus resulting in an NVE ensemble. The volume of the equilibrated system was 109.5 nm^3, and the fluctuations of the volume in the NPT simulation was 0.41 nm^3. The average temperature was 299.1 (NPT) and 298.9 (NVE) fluctuating about 2 K in both simulations, well in line with the targeted temperature of 298.15.

Notably, the SBMD simulations were targeted at 300 K, and in this case the average temperature for all simulations was close to 301 and fluctuated approximately 5 K. The pressure is of main concern in the SBMD approach and therefore we also calculated the pressure for the 20 ns SBMD simulation, which resulted in 90 ± 500 bar. Only TIP3P was used for representing the solvent in the PBC/PME simulations that were approximately 2.5 times slower (1 ps required 57 CPU minutes on a single DEC alpha processor) than the non-periodic boundary spherical simulations.

6.2.1 Analysis

All analyses were done on the production phase of the simulations, where we considered the trajectories equilibrated after 1 ns. All calculations, except for rotational reorientation times, were preceded by a mass-weighted fit of the non-hydrogen protein atoms onto a reference structure. The 20 ns simulation was subdivided into ten blocks and then averaged over all blocks, whereas averages for the 2 ns simulations were done over all eight trajectories.

6.2.1.1 Atomic displacement and fluctuations

The crystallographic temperature factors were converted into atomic mean square fluctuations using the relation:

$$3B_i = 8\pi^2 <\Delta r_i^2> \quad (1)$$

where <…> denotes time averaging over the trajectory, and Δr_i is the displacement of atom *i* from its average position.

6.2.1.2 Order parameters

The generalized order parameter, S^2, in the model-free formalism from Lipari and Szabo [34] was calculated according to

$$C_{2i}(t) = <P_2(h_i(\tau + t) \cdot h_i(\tau))> \tag{2}$$

and

$$S^2 = \lim_{t \to \infty} C_{2i}(t) \tag{3}$$

P_2 is the second order Legendre polynomial and *h* is a unit bond vector along the amide N–H bond of residue *i*. An order parameter varies between 1 and 0, reflecting the degree of restriction on the residue, where 0 means a total unrestricted motion. The estimate of S^2 was based on the plateau value of the resulting curve of Eq. (3). However, in some cases no plateau was reached, implying insufficient conformational sampling of these residues (data not shown). The only residues suffering this fate in more than one simulation were at positions 509 and 510 (i.e., the C-terminal region of the GR DBD). Autocorrelation curves with such behaviour were discarded in the evaluation.

6.2.1.3 Hydration

The solvent accessible surface area (ASA) was calculated according to Lee and Richards [35] using a water probe of radius 1.4 Å.

A hydrogen bond was considered present if the lifetime was 20 ps or longer and the distance between the hydrogen and acceptor was 2.4 Å or less for snapshots equally spaced by 2 ps.

6.2.1.4 Energies

Frames separated by 10 ps were extracted for energy component analysis.

The configurational entropy was estimated with quasi-harmonic analysis (see [36] and references therein), where the frequencies, ω_i, were obtained from the diagonalization of the mass-weighted covariance-matrix, with *n* atoms and $3n-6$ non-zero degrees of freedom at the absolute temperature, *T*.

$$S_{ho} = k_b \sum_{i}^{3n-6} \frac{\hbar\omega_i | k_b T}{\exp(\hbar\omega_i | k_b T) - \mathbf{1}} - 1\mathrm{n}[1 - \exp(-\hbar\omega_i / k_b T)] \tag{4}$$

The co-variance matrix is defined as:

$$\sigma_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \tag{5}$$

where x_i is one Cartesian coordinate in the system. All solute atoms were included in the entropy estimates, and frames separated by 0.4 ps in the production phase were used to construct the covariance matrix, σ.

6.2.1.5 Rotational times

Rotational reorientation times for the protein in water were calculated using the time-correlation function,

$$C(t_m) = \langle P_2(\mu(0)\cdot\mu(t_m))\rangle \approx \frac{1}{N-m}\sum_{n=1}^{N-m} P_2\left[\mu\,(t_n)\cdot\mu\,(t_n+t_m)\right] \qquad (6)$$

N is the total number of time steps in the simulation and m is the number of time steps passed at time t_m. $P_2\,(x)$ is the second-order Legendre polynomial and μ is a unit vector, defined in the protein molecule. The orientations of μ at times t_n and $t_n + t_m$, respectively, were calculated from the trajectory. Three unit vectors, $V1$, $V2$ and $V3$ were defined in the protein molecule: $V462^{C\alpha}$ - $A467^{C\alpha}$, $R496^{C\alpha}$ - $L510^{C\alpha}$, and between the two zinc ions (see Fig. 6.1). We calculated the rotational reorientation time τ from the inverse of the slope of the linear part, 1 ps $< t_m <$ 10 ps, of the decay of $\ln[C(t_m)]$ with time t_m. The fast initial (<0.2 ps) drop in $C(t_m)$ due to librational motions seen in some of the correlation curves does not interfere with the calculation of the rotational reorientation time τ.

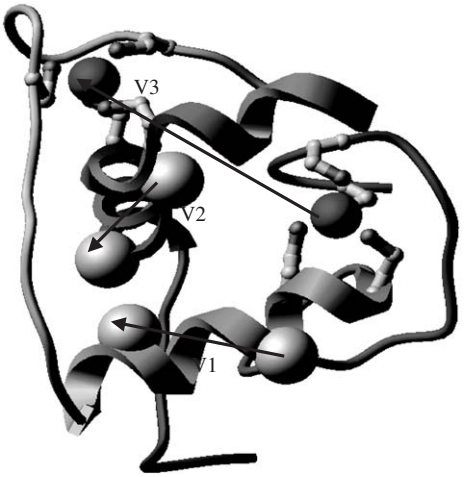


Fig. 6.1. Schematic representation of the crystal GR DBD second monomer (pdb accession code 1 glu). The thiolate ligands to the zinc atoms are shown (hydrogens are omitted), and the arrows display the three unit vectors ($V1$, $V2$ and $V3$) used for calculating the re-orientational times. $V1$ and $V2$ are defined through the C_α atoms (light spheres) of Val 462 to Ala 467 ($V1$), Arg 496 to Leu 501 ($V2$) and $V3$ through the two zinc ions (dark spheres) (see Color Plate 13).

6.3 RESULTS AND DISCUSSION

6.3.1 Protein stability and rotational diffusion

All simulations show similarities in spatial displacement, radius of gyration and accessible surface area of the protein atoms, regardless of the water model or compared with the 20 ns simulation (Table 6.1). Hence, the solvent model does not significantly influence the stability of the simulations. The average number of solute intra-molecular hydrogen bonds is also virtually the same for simulations in all three water models (Table 6.1). A different picture is seen for the protein-water inter-molecular hydrogen bonds, for which the lifetimes and numbers of protein–water hydrogen bonds increase for SPC and SPC/E compared with TIP3P. This increase is mainly due to Asp and Glu sidechain carboxylate–solvent interactions in conjunction with an increased tetrahedral formation for the two former water models.

The correlation time for the rotational diffusion of the protein, evaluated from a set of three vectors spanning the protein, depends on the water model (Table 6.2) in a fashion that is consistent with observations on small molecules simulated with these water models [19,21]. The slower dynamics, and hence larger viscosity, of the SPC/E model compared with TIP3P is also reflected in the protein rotational diffusion. Correlation times obtained from the SPC/E simulations are also the ones closest to experimental estimates. The slightly anisotropic shape of the protein is reflected in the slower rotation observed for vector *V*3 in all three water models, and also seen in the calculation from the 20 ns TIP3P simulation. Although the 2 ns simulations (of which the last ns is used in the analysis) are fairly short on the timescale of the rotational diffusion of the protein, and there is a spread between the correlation functions obtained from eight simulations (Fig. 6.2), averaging over all simulations improves the statistics, and the difference between the water models is thus significant. The difference between the SBMD and the PBC/PME simulations seen in the rotation diffusion (Table 6.2), on the other hand may be within the statistical fluctuations; since we only performed one (1.5 ns) PBC/PME simulation we can only say that the behaviour of the protein in TIP3P water under these conditions is similar to the spherical boundary conditions used in the other simulations.

6.3.2 Structural fluctuations in the protein

As seen in Fig. 6.3, the root mean square fluctuations (rmsf) of the C_α atoms per residue are lowest for SPC/E. Regions that exhibit larger fluctuations than the experimentally derived B-factors are those that form protein–protein and protein–DNA contact interfaces in the crystal structure. Since the simulations were done on the crystal monomer, we would expect larger fluctuations in these regions in the simulation than in the crystal structure, owing, for instance, to conformational relaxation of the protein in the absence of the crystal contacts. Previous studies [37,38] on the influence of pressure on atomic positional fluctuations show that several thousand atmospheres are required before the fluctuations are significantly altered. Since the magnitude of the simulated rmsf is comparable to the magnitude of fluctuations inferred from the experimental B-factors, we can exclude the possibility that the differences between the water models are due to pressure effects.

Table 6.1 Mean simulation statistics for TIP3P, SPC and SPC/E from the production phase of the trajectories averaged over all eight trajectories

Model	Rgyr (Å)	ASA (Å^2)	Number of H-bonds[a]/τ (ps)		RMSD[b] (Å)	
			Intra-molecular	Inter-molecular	Backbone	Sidechain
TIP3P (20 ns)	11.88 (0.09)	4 898 (64)	45 (2) / 53.2 (7.5)	43 (1) / 39.0 (6.4)	1.83 (0.25)	3.36 (0.39)
TIP3P	11.92 (0.13)	4 944 (105)	46 (2) / 64.0 (19.6)	47 (1) / 26.7 (2.2)	2.02 (0.19)	3.63 (0.37)
SPC	11.90 (0.09)	4 926 (106)	45 (3) / 59.6 (11.5)	70 (2) / 39.7 (2.4)	1.74 (0.25)	3.16 (0.42)
SPC/E	11.89 (0.11)	4 956 (123)	47 (5) / 61.2 (8.5)	92 (2) / 46.9 (2.7)	1.86 (0.23)	3.39 (0.41)
Initial	11.85	4 812	43	57	—	—

[a]Intra-molecular protein–protein; inter-molecular protein–solvent.
[b]Relative to the initial coordinates from the crystal structure.

Table 6.2 Rotational diffusion time (ns)

	Solvent model							
System:	TIP3P (PBC/PME)[a]		TIP3P20[b]	TIP3P	SPC	SPC/E	Experimental[c]	Ref.
Ensemble:	NPT	NVE	NVT	NVT	NVT	NVT	6.3 ± 0.1, 5.76 ± 0.001	[5,8]
Vector								
V1	1.0	1.0	1.7	2.0 (0.2)[d]	2.5 (0.2)	3.4 (0.2)		
V2	1.1	1.0	2.2	2.5 (0.2)	3.1 (0.5)	4.3 (0.6)		
V3	1.4	1.0	3.4	4.0 (0.2)	5.1 (0.3)	7.1 (0.5)		

[a]Periodic boundary conditions and particle mesh Ewald summation for long-range electrostatics.
[b]20 ns simulation.
[c]isotropic tumbling times from NMR experiments.
[d]The mean value and standard deviation (in parentheses) of the rotational reorientation of the *V*1, *V*2 and *V*3 vectors were calculated using all eight trajectories. The last 1.0 ns of the trajectories were used in all calculations. See Fig. 6.1 for the definition of the vectors.

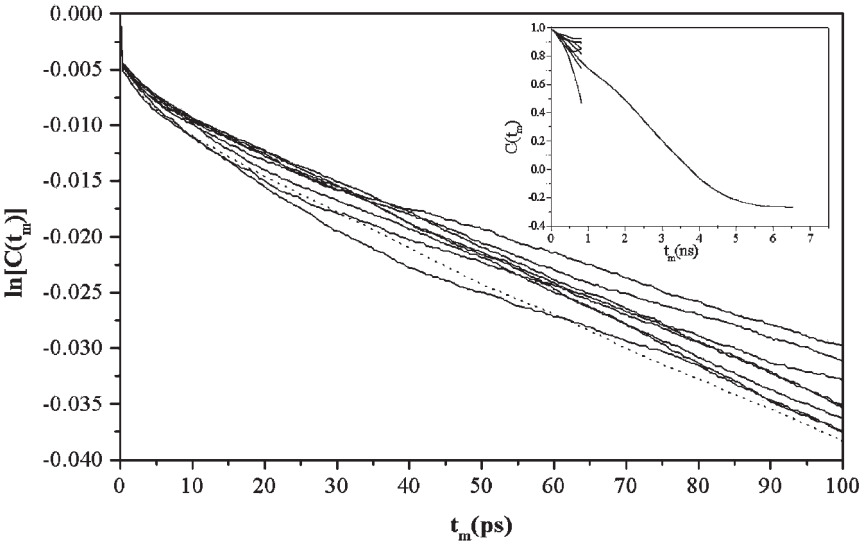


Fig. 6.2. ln[$C(t_m)$] was calculated with vector $V1$ from simulations with TIP3P water. 1.0 ns blocks from eight simulations (line) and 20.0 ns simulation (dot). The inset shows the time-correlation function $C(t_m)$ from simulations with TIP3P water.

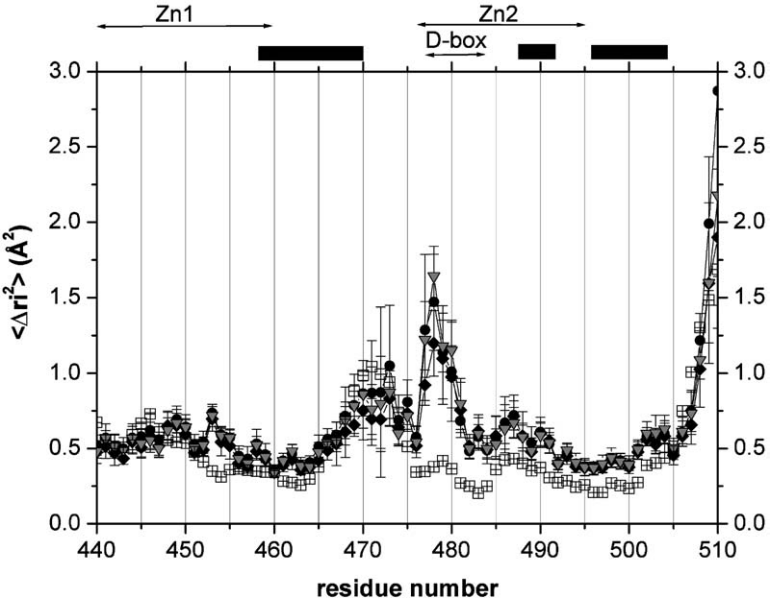


Fig. 6.3. Ensemble-averaged C_a atomic fluctuations per residue and standard deviations shown as bars for all 24 simulations with the three water models: TIP3P, SPC, SPC/E. The corresponding experimental B-factors are represented by filled triangles, circles, black diamonds and crossed squares, respectively. Both zinc finger domains (Zn1 and Zn2) and the three helices are indicated on top.

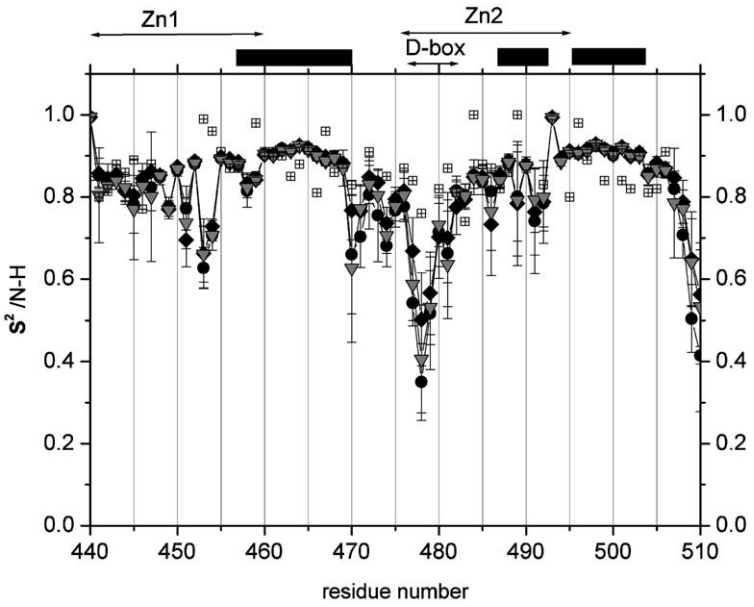


Fig. 6.4. Ensemble-averaged S^2 N–H order parameters per residue and standard deviations shown as bars. TIP3P, SPC, SPC/E and the corresponding experimental S^2-values are represented by filled triangles, circles, black diamonds and crossed squares, respectively. Both zinc finger domains (Zn1 and Zn2) and the three helices are indicated on top.

A similar picture is seen for the calculated generalized S^2-order parameters, depicted in Fig. 6.4. We see the trend of increased order of the amide bond, where generalized order parameters S^2 in the SPC/E simulations on average are closer to the corresponding experimentally derived NMR N–H relaxation order parameters.

The most pronounced solvent effects on the S^2-order parameters is found for surface exposed residues (Fig. 6.5), which are distant from the core of the protein. Residues forming the core protein or helices remain very similar among the solvent models in terms of fluctuations and S^2-order parameters. In line with the increased order of the SPC/E water model compared in particular with the TIP3P model, the restriction increases on the solvent exposed sidechains in the SPC/E simulation. Apparently, this is also the case for the backbone structure, as seen in the Ramachandran map (Fig. 6.6), where the spread of points is smaller for SPC/E than for TIP3P, with SPC in between.

6.3.3 Entropy and energy

In TIP3P and SPC the quasi-harmonic estimates of the conformational entropy of the protein are very similar, but for SPC/E the entropy is significantly lower (Table 6.3), consistent with the lower fluctuations and higher S^2-order parameters, all of which basically reflect the enhanced structure in SPC/E compared with the other two models. However,

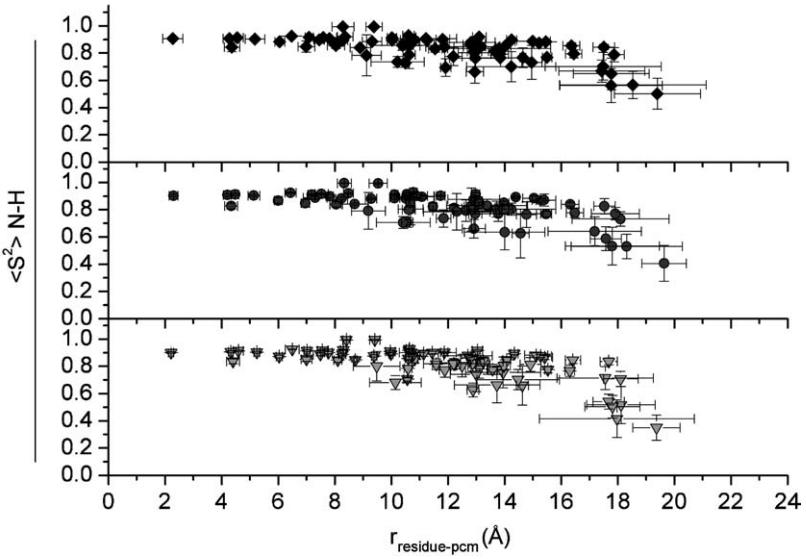


Fig. 6.5. Generalized N–H order parameters as a function of the ensemble average distance for the residue to the protein centre of mass. The horizontal bars show the fluctuation of the distance, and the vertical bars the standard deviation of the ensemble-averaged S^2 order parameter.

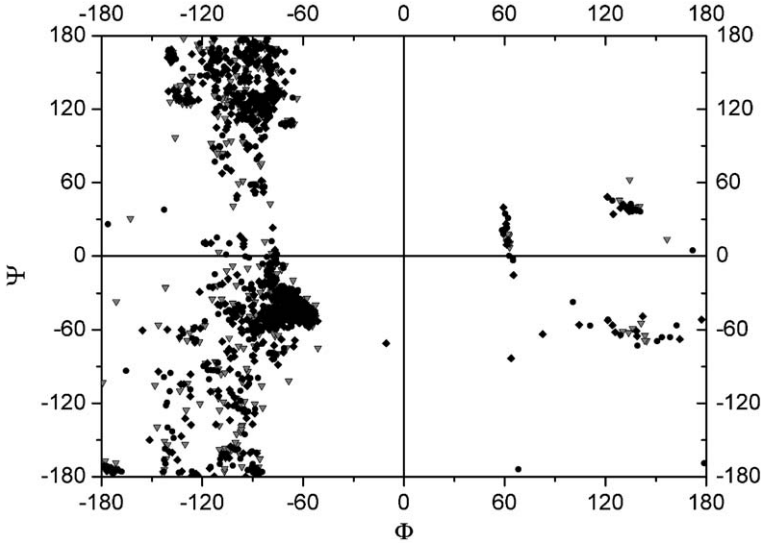


Fig. 6.6. Ramachandran $\Phi-\Psi$ map with TIP3P, SPC and SPC/E represented by filled triangles, circles and black diamonds, respectively.

Table 6.3 Ensemble-averaged[a] quasi-harmonic protein entropy and total potential energy for the whole system at 300 K

Solvent	<*E*> pot kcal/mol	<*S*> kcal/mol, K
TIP3P	−33 386 ± 15	3.17 ± 0.05
SPC	−34 085 ± 16	3.14 ± 0.07
SPC/E	−38 300 ± 24	3.06 ± 0.04

[a]Average calculated over all eight simulations.

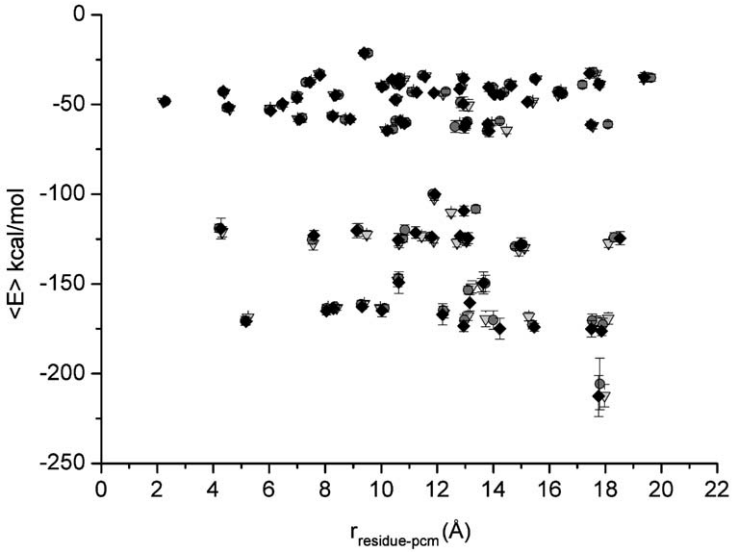


Fig. 6.7. Mean interaction energies for each residue with its environment (kcal/mol) as a function of the ensemble average distance for the residue to the protein centre of mass for the three water models, with TIP3P, SPC and SPC/E represented by filled triangles, circles and black diamonds, respectively.

the energy of each protein sidechain interacting with its environment (protein and water) is very similar for the three water models (Fig. 6.7). Hence the increased degree of solvent order around the protein in the SPC/E simulation, which results in favourable solute–solvent interactions, is accompanied by an increase in the solute internal energy.

6.4 CONCLUSION

We conclude that it is sufficient to average over several short trajectories for calculating reorientation times. In most of the curves there is an initial fast (< 0.2 ps) drop in $C(t_m)$, which is related to librational motion, as indicated by the undershoot before the slower part starts. This does not, however, interfere with the analysis of the rotational reorientation time τ. The overall rotation correlation time for the protein depends on the solvent viscosity.

Calculated atomic fluctuations and generalized S^2 order parameters for residues on the protein surface are most susceptible to changes of the water model but in general the overall structural properties of the protein remain quite unaffected.

6.5 ACKNOWLEDGEMENT

This work was supported by the Swedish Research Council.

6.6 REFERENCES

1 R.M. Evans, Science, 240 (1988) 889.
2 L.P. Freedman, B.F. Luisi, R. Korzun, R. Basavappa, P.B. Sigler and K.R. Yamamoto, Nature, 334 (1988) 543.
3 V. Laudet, C. Hänni, J. Coll, F. Catzeflis and D. Sthehelin, EMBO J., 11 (1992) 1003.
4 R. Kumar and E.B. Thompson, Steroids, 64 (1999) 310.
5 H. Berglund, H. Kovacs, K. Dahlman-Wright, J.-Å. Gustafsson and T. Härd, Biochemistry, 31 (1992) 12001.
6 J. Zillicaus, K. Dahlman-Wright, A. Wright, J.-Å. Gustafsson and J. Carlstedt-Duke, J. Biol. Chem., 266 (1991) 3101.
7 J. Zillicaus, K. Dahlman-Wright, J. Carlstedt-Duke and J.-Å. Gustafsson, J. Steroid Biochem. Molec. Biol., 42 (1992) 131.
8 A. Wikström, H. Berglund, C. Hambræus, S. van den Berg and and T. Härd, J. Mol. Biol., 289 (1999) 963.
9 M.A.A. van Tilborg, J.A. Lefstin, M. Kruiskamp, J.-M. Teuben, R. Boelens, K.R. Yamamoto and R. Kaptein, J. Mol. Biol., 301 (2000) 947.
10 B.F. Luisi, W.X. Xu, Z. Otwinowski, L.P. Freedman, K.R. Yamamoto and P.B. Sigler, Nature, 352 (1991) 497.
11 H. Baumann, K. Paulsen, H. Kovacs, H. Berglund, A.H. Wright, J-Å. Gustafsson and T. Härd, Biochemistry, 32 (1993) 13463.
12 M.A.L. Eriksson and L. Nilsson, Protein Eng., 11 (1998) 589.
13 M.A.L. Eriksson, T. Härd and L. Nilsson, Biophys. J., 68 (1995) 402.
14 J. Bredenberg and L. Nilsson, PROTEINS: Structure, Function, and Genetics, 49 (2002) 24.
15 T.C. Bishop and K. Schulten, Proteins, 24 (1996) 115.
16 W.L. Jorgensen, J. Chandrasekar, J.D. Madura, R.W. Impey and M.L. Klein, J. Chem. Phys., 79 (1983) 926.
17 H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren and J. Hemans, in B. Pullman (Ed.), Intermoleular forces, Reidel, Dordrecht, Interaction models for water in relation to protein hydration, (1981) 331–342.
18 H.J.C. Berendsen, J.R. Grigera T.P. Straatsma, J. Phys. Chem., 91 (1987) 6269.
19 P. Mark and L. Nilsson, J. Phys. Chem. B, 105 (2001) 8028.
20 P. Mark and L. Nilsson, J. Phys. Chem. A, 105 (2001) 9954.
21 P. Mark and L. Nilsson, J. Phys. Chem. B, 106 (2002) 9440.
22 R. Walser, P.H. Hünenberger and W.F. van Gunsteren, Proteins: Structure, Function and Genetics, 43 (2001) 509.
23 P.E. Smith and W.F. van Gunsteren, J. Mol. Biol., 236 (1994) 629.
24 B.R. Brooks, E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan and M. Karplus, J. Comp. Chem., 4 (1983) 187.
25 A.D. MacKerell Jr, D. Bashford, M. Belott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus, J. Phys. Chem. B, 102 (1998) 3586.

26. R.H. Stote and M. Karplus, Proteins, 23 (1995) 12.
27. J. Bredenberg and L. Nilsson, J. Int. Quant. Chem., 83 (2001) 230.
28. C.L.I. Brooks, A.T. Brunger and M. Karplus, Biopolymers, 24 (1985) 843.
29. J. Norberg and L. Nilsson, Biophys. J., 79 (2000) 1537.
30. J. Sarzynska, T. Kulinski and L. Nilsson, Biophys. J., 79 (2000) 1213.
31. J.P. Ryckaert, G. Ciccotti and H.J.C. Berendsen, J. Comp. Chem., 23 (1977) 327.
32. P.J. Steinbach and B.R. Brooks, J. Comp. Chem., 15 (1994) 667.
33. T. Darden, D. York and L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J. Chem. Phys., 98 (1993) 10089–10092.
34. G. Lipari and A. Szabo, J. Am. Chem. Soc., 104 (1982b) 4559.
35. B. Lee and F.M. Richards, J. Mol. Biol., (1971) 379.
36. I. Andricioaei and M. Karplus, J. Chem. Phys., 115 (2001) 6289.
37. J. Norberg and L. Nilsson, Chem. Phys. Lett., 224 (1994) 219.
38. J. Norberg and L. Nilsson, J. Chem. Phys., 104 (1996) 6052.

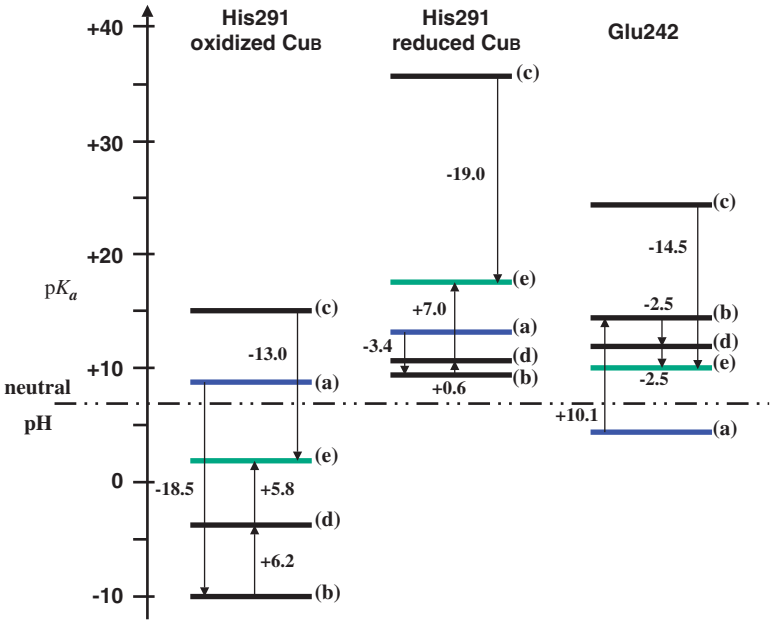


Plate 12. A schematic of the p*K*a values of His291 in the oxidized and reduced Cu_B complex and the pK_as of Glu242, for different models ((a)–(e)) applied in our study. Blue bars are related to the pK_a values in the aqueous phase and green bars represent the calculated pK_as in the protein (cytochrome oxidase). The arrows with the numbers represent the magnitude of the pK_a shifts caused by the different structural effects. See the text for the explanation (illustration appears on page 75 of this volume).

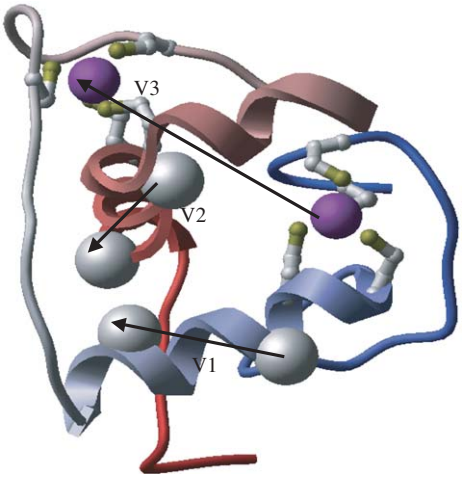


Plate 13. Schematic representation of the crystal GR DBD second monomer (pdb accession code 1 glu). The thiolate ligands to the zinc atoms are shown (hydrogens are omitted), and the arrows display the three unit vectors (*V*1, *V*2 and *V*3) used for calculating the re-orientational times. *V*1 and *V*2 are defined through the C_α atoms (light spheres) of Val 462 to Ala 467 (*V*1), Arg 496 to Leu 501 (*V*2) and *V*3 through the two zinc ions (dark spheres) (illustration appears on page 127 of this volume).

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 7

Computer simulations of DNA stretching

Raimo A. Lohikoski[1], Martin Dahlberg[2] and Aatto Laaksonen[2]

[1]University of Jyväskylä, Department of Physics, FIN - 40014
University of Jyväskylä, Finland
[2]Division of Physical Chemistry, Arrhenius Laboratory,
Stockholm University, S 106 91, Stockholm, Sweden

Abstract

In this chapter we will give short review of computer modelling/simulations of DNA manipulation as a complementary tool to current single molecule manipulation experiments in order to follow the impact on molecular structure during the manipulation experiments. As an example we report molecular dynamics simulations of a 22 base-pair DNA fragment in an explicit water solution with counter-ions to mimic a torsionally unconstrained single-molecule stretching experiment. Positions of the O5′ and O3′ atoms at one end of the 22-mer were fixed while an external linearly increasing tensile force was applied on the corresponding atoms at the other end. Changes in the intramolecular potential energy components reflect large structural changes during the stretching. Large increase in the specific torsional angles of the elongated double-strand DNA is observed as the angles find new stationary values, quite different from those in the dynamically stable original structure before the stretching. Upon pulling the DNA gradually looses its twist developing a spectacular base-stacked structure with the backbones are stretching out in parallel, and the bases were in a zipper-like stack in the position of the major groove of the original un-stretched structure. The results are discussed in the light of recent experiments and some biological consequences of the simulation results are indicated.

7.1 INTRODUCTION

Conventional studies of interactions, structure and dynamical processes in biomolecular systems are performed on a large number of molecules at equilibrium conditions. The observed properties are obtained as averages over the operational time window of the

used instrumental technique. The exact opposite is the case in recently developed single-molecule methods allowing studies of individual molecules over time and changing conditions and thereby giving information that would have been lost in averaging over time. Individual molecules undergoing fluctuations in a variety of regimes can be used to report specific events and processes taking place in their local environment. Optical single molecule techniques, applicable in liquids at room temperature, can be traced back to the 1970s and detection by fluorescent markers, first reported by Hirschfeld [1]. Development of non-optical techniques was triggered by the invention of the scanning tunnelling microscope (STM) by Binnig and Rohrer [2,3]. STM was initially applied on conducting surfaces in *vacuum* but was soon developed to a family of methods even operating at conditions needed in studies of biological systems in solution. The immensely popular atomic force microscope (AFM) [4] is just one of these more general techniques. Single molecule manipulation experiments are normally divided into two categories: 'imaging and force measurement' techniques. AFM, for example, can be applied for both purposes.

7.1.1 Straining biopolymers one by one

For force measurements in particular, the development of experimental single-molecule manipulation techniques during the last 10–15 years has made it possible to observe the physical and/or physico-chemical properties of individual biopolymers (*i.e.*, nucleic acids and proteins) using a variety of different nano-mechanical tools and methods. Besides AFM, there are optical and magnetic tweezers, micro pipettes and micro needles, as well as methods based on flow fields and hydrodynamic drag. These all address the very same goal but with different set-ups and provide different sensitivity on the scale of microscopic forces. Central to all these techniques is anchoring of one end of the molecule, for instance by bonding it to a surface, and connecting the other end to a force sensor device which is moved, *i.e.*, translated and/or rotated. The result of a measurement is a dynamical force curve. Together these different experimental techniques can cover a broad range of microscopic forces; from those corresponding in magnitude to Brownian motion (pN = 10^{-12} Newtons) to those capable of breaking covalent bonds (nN = 10^{-9} Newtons). It should be mentioned however, that entering the nano-scale, there are two complications in comparison with conventional studies affecting the accuracy and the means to interpret the results: kT and the very technique itself. While thermal motions can limit the resolution of the measured force, the fact that thermodynamics does not apply at all well on single molecules obscures a reliable quantitative analysis of the results. These newly developed possibilities to manipulate and study individual molecules in their natural surroundings have already taken the experimental investigations to a new level of ingenuity and it has all only just begun.

7.1.2 The highly flexible DNA

Concerning DNA, the focus of this chapter, its characteristic structure offers three interesting regimes on the force axis: force in unzipping the strands while breaking the hydrogen bonds between the bases, stretching the sugar-phosphate backbone and twisting

of the helical structure. And indeed, DNA samples have been stretched, bent, under-, and over-wound (negatively and positively supercoiled, respectively), even braided or having had the strands unzipped in a large number of different ways [5–28]. While the unzipping observations reflect the hydrogen bonding nature of the base pairs, the stretching and related twisting experiments have shown a spectrum of results depending on how the stretching is carried out, and whether the chain is torsionally free or contains an initial under- or over-twist [15]. The very purpose of these systematic studies is, of course, to learn how biological molecules behave and respond in various controlled conditions and environments. For the current state-of-the-art and various aspects of the single-molecule treatment of biomolecules, the reader is referred to several excellent reviews [20,24,26,28–41].

7.1.3 The highly robust DNA

The aim when studying DNA using micromanipulation methods is, besides investigating its topological response, to gradually learn more about the key molecular mechanisms of how DNA maintains its genetic information and how it can be passed further without damage. Indeed, many biological processes can cause severe structural stress to DNA. Therefore, it is highly important to apply these sensitive single-molecule manoeuvring methods also under extreme conditions, even close to denaturation, and follow the recovery process back to the initial form, closely studying its capability to restore its natural structure. Over 10 years of single-molecule studies of both double- and single-stranded DNA have given much new valuable information of this highly flexible, extensible and amazingly adaptable molecule.

7.1.4 Pulling the cord

When stretched, the behaviour of a double-helical DNA is initially no different from that of simple polymers. Several studies show that when a small force (10 pN) is applied, the elasticity can be described in purely entropic terms. Statistical mechanics, applied on chain molecules, provides simple models, such as the 'freely jointed chain' (FLJ) or the 'worm-like chain' model (WLC). In fact, the ideal inextensible WLC model seems to reproduce the force/length curves of double-stranded DNA (dsDNA) surprisingly well (see Fig. 7.1 of Ref. [9]) to a point where the stretching becomes purely elastic, starting to follow Hooke's law. Remarkably, a very distinct deviation from the elastic behaviour was observed in the early experiments to stretch a nicked DNA upon increasing the applied force [6]. A rapid transition of cooperative character was reported to take place when the pulling force reached a value close to 70 pN, suddenly extending the dsDNA within a narrow force interval of 2–3 pN to a length of roughly 1.7 times the original equilibrium length of B-DNA (the most common form of DNA in living organisms). This transition was found to be reversible, and when allowed to relax, a large hysteresis effect could be observed. Whether it is a structural transition [13] or a force-induced melting [42] appears to be still a matter of debate. Both explanations have found support. In the case of a structural transition, a new metastable form of B-DNA was suggested and referred to as S-DNA. Based on molecular modelling, a detailed structure of the overstretched

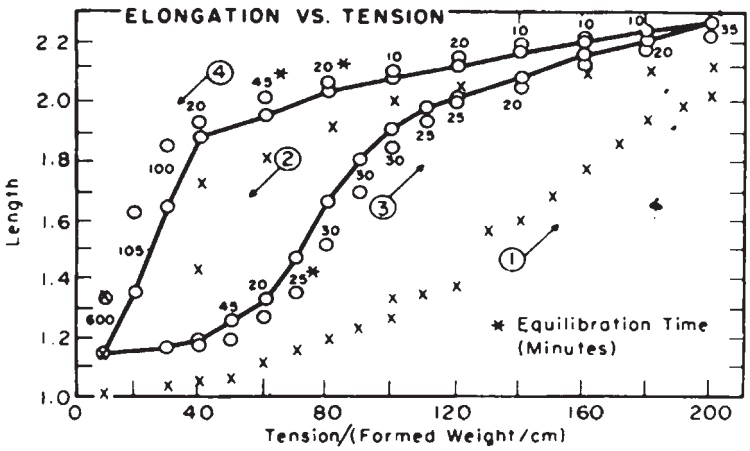


Fig. 7.1. Results from stretching DNA bundles by Wyckoff from 1955 [48]. See text for details.

B-DNA was obtained [13,43,44]. The B→S transition has gained much popularity in both experimental and theoretical studies. S-DNA can be visualized as a slightly twisted ladder with the bases paired but tilted. A spontaneous reason to question the B→S transition model was the cooperativity and hysteresis effects [45]. A more formal critical evidence for the melting model was built up based on thermodynamical analysis of careful experiments with various denaturation-promoting factors [42,46]. For unnicked DNA, the same type of transitional plateau occurs higher, close to 110 pN of pulling force [15]. Transitions to other forms of DNA have been shown to occur when the twisting angle is constrained (see the tension–torque 'phase diagram' of dsDNA in Ref. [22]).

7.1.5 Early mechano-chemical work

Already, before Watson and Crick published the base-paired double-helical structure of DNA [47], it was clear that DNA is indeed a highly extensible molecule. The early work by Wilkins *et al.* [48] showed that DNA could be extended to double its length, and while undergoing a transition to a new structure with tilted bases. Figure 7.1 shows a result from another similar early experiment from the same period [49] where a DNA fibre bundle in 85 percent ethanol bath is stretched by adding small weights of 10 g at a time until a total weight of 200 g, resulting a maximum relative elongation of 2.14. The figure shows two elongation–relaxation experiments with the same bundle. The first one (marked with 'x') was carried out by successively adding weights every 2–3 minutes, allowing it to relax thereafter. In the second experiment, the bundle was allowed to reach an equilibrium length after each weight addition/removal.

The solid curve is drawn through the equilibrated points. The elongation curve shows a sigmoid shape, suggesting a transition between two phases, each with its own extensibility and rest length. The relaxation curve (not a sigmoid) shows distinct hysteresis effects. The experiment was repeated (not shown in the figure) with a thinner DNA bundle of 6 microns and lighter weights, exhibiting the same behaviour which also could be converted

by scaling to the curve shown in Fig. 7.1, showing that the reversible extensibility is a property of the individual fibre rather than that of the bundle [49]. In these pioneering mechano-chemical investigations [48,49], and in numerous similar studies done afterwards on DNA fibres at varying temperatures, salt/ion concentrations have been material to pave the way to today's single-molecule morphological studies.

7.2 STRUCTURAL CHANGES IN DNA DURING SINGLE-MOLECULE MANIPULATIONS

Single-molecule manipulations of the DNA molecules by stretching, twisting, unzipping and/or even ripping apart the backbone strands have offered a rich collection of DNA's elasticity and binding information. However, the more fundamental understanding of the external force-induced events is still lacking, owing to the unobservable itemized structural actions. A typical example is the force–length curve of a stretched double-stranded (ds) DNA. This has a nice plateau-structure showing that a small (few pN) increase in external force gives a 70 percent increase in the molecule's length. The structures of the overstretched S-DNA, as well as the P-form found from over-wound DNA, have been suggested merely based on molecular modelling [13,44,50]. Thus far, the method of molecular dynamics simulations (MD) is the only general approach capable of giving all (necessary) intermediate structures of an atomistic or molecular process. The concessions made to facilitate the mechanical manipulation computations of DNA coupled with known progressions made in protein and nucleic acid MD techniques have led us to pose the question of today's realities to simulate a complete stretching experiment of a normal DNA in its aqueous environment.

7.2.1 Computer modelling of the stretched structure of DNA

Although it is possible to obtain rough images of single molecules using a scanning force microscope, none of the current single-molecule methods can at the moment provide the detailed structural parameters of the manipulated molecules. Therefore, it is not possible to follow the conformational changes directly in real-time. The structures of the overstretched S-DNA, as well as the P-form found from over-wound DNA, have been suggested merely based on molecular modelling [13,44,50]. Modelling has become routine work with rapidly increasing computing power and with new sophisticated computational chemistry tools developed to explore complex biomolecular systems. Quantum chemical calculations and computer simulations, combined with spectroscopic and crystallographic studies, are powerful tools in structure determination. Molecular dynamics (MD) simulations will become more and more useful, even in combination with nano-mechanical single-molecule experiments.

Stretching of proteins and DNA has been modelled on several occasions. Concerning DNA, there is an energy minimization study dealing with all possible ways to stretch one turn of a double helix to its double length, modelled using the JUMNA software [43], and carried out in internal and helicoidal coordinates. An implicit solvent was described with a distance-dependent dielectric function while the phosphate charges were reduced

to −0.5e [44]. The implicit solvent model with distance-dependent dielectric was also used in the first MD simulation to stretch a 12-mer starting from a canonical form [51]. The stretching was done with gradual stepwise increments of the force between the opposing three ends after each 10 ps. A ladder structure with the base pairs largely intact but the bases tilted in an inter-strand stacking was observed with an applied force of 220 pN after a simulation of 200 ps. Potential of mean force (PMF) calculations based on NPT-MD simulations with periodic boundaries were reported with explicit solvent molecules using a double-stranded d(ACTG)3-d(CAGT)3 dodecamer [52]. The CHARMM22 force field [53] was used for the dodecamer and TIP3 model for water [54]. The C5 terminal atoms on separate strands were pushed and thereafter pulled apart by increasing the distance by 0.5 Å followed by a short equilibration of 10 ps before calculation of the PMF. The water box was successively made longer during the stretching. The PMF calculations show that the dodecamer can be stretched 2.4 times from a length shorter than A-DNA without any significant energy barrier. An extensive and systematic DNA stretching/compression study was reported for regular poly(dG).poly(dC) and poly(dA).poly(dT) double helices extending them from compressed states of 2.0 Å per base-pair step to 7.0 Å per residue [55]. Both strands were stretched simultaneously without torsional constraints on the helix. A distance-dependent dielectric function was used instead of explicit solvent molecules and the atomic charges were modified to mimic the ion concentration.

7.3 COMPUTER SIMULATION OF A SINGLE DNA STRETCHING EXPERIMENT

Can a complete stretching experiment of a double helical DNA in aqueous solution be simulated realistically enough? Or simply, how closely to a real DNA does a computer-created electronic DNA behave? Concerning the quality of current atomistic force fields used in simulations, we can assume that they do produce reliable results at normal equilibrium conditions. Results from hundreds of DNA simulations, found in literature dealing with various structural and dynamic aspects of nucleic acids, give convincing evidence that computer experiments are useful as a complement to experimental observations, providing the well-needed details on atomic length scale. However, typical timescales used in the stretching experiments and those applicable in conventional atomistic MD simulations are very different for obvious practical reasons.

In this work, we have carried out all-atom MD simulations to mimic a typical stretching experiment of a double-stranded DNA where both the 3′ and 5′ terminal atoms in one end of the DNA are anchored to fixed positions while the corresponding atoms on the other termini are pulled with a linearly increasing force. No constraints are set on unwinding the helix. In other words, the double helix is allowed to open to a conformation with straight parallel strands when stretched long enough. The overall simulation set-up should be molecularly similar to conditions in an iso-tensional pulling experiment in water solution. However, no beads are attached to DNA in the simulation, and no other microscopic instrumental details from the experiment are included in the model.

7.3.1 Computational details

Our DNA, a d(gtctgaattctaatgtagtata) duplex, is placed in a 300 Å long and 45 Å wide hexagonal container. The container, which was almost four times longer than the 22-mer itself, was filled with water and Na^+ counter-ions (one per phosphate group). The whole simulated system contained totally 49 442 atoms (see Fig. 7.2).

Since this hexagonal cell is wide enough, DNA becomes hydrated with several layers of water, including the bulk phase. The thickness of the water layer around the DNA was more than 11 Å. The latest revision of the CHARMM force field, developed for nucleic-acid systems [53,56,57] was used for the DNA and the flexible SPC water model [58] was used for the solvent molecules. This particular water model reproduces both the dielectric constant and diffusion coefficient for water correctly; both important properties for a solvent in biological systems. All computations were performed at 300K, using the general-purpose parallel MD simulation package M.DynaMix [59]. A reversible double time-step algorithm was applied with the time step of 0.2 fs for fast motion (bond stretching, angle bending and the closest non-bonded interactions) and 2.0 fs for slow motion (all remaining degrees-of-freedom). Ewald summation was used for electrostatic interactions.

Before starting the pulling, the system was equilibrated 40 ps in NVT ensemble while the DNA structure was kept rigid. Equilibration was continued another 20 ps by allowing DNA to become fully flexible. Thereafter the equilibration was continued in NPT ensemble for 60 ps to obtain a right density for the system. During constant pressure simulation, the box finally stabilized to a new dimension of roughly 240 Å by 48 Å. An additional equilibration in NVT for 10 ps was carried out before the stretching was started. The average distance between the O3′ and O5′ atoms at the opposite ends of the strands was 78.7 Å before the pulling was started. The average, dynamically stable initial structure of the DNA was close to that of B-DNA.

The stretching was carried out in NVT ensemble in order to keep the ionic strength constant in the simulation cell. It should also be commented that the NPT simulation

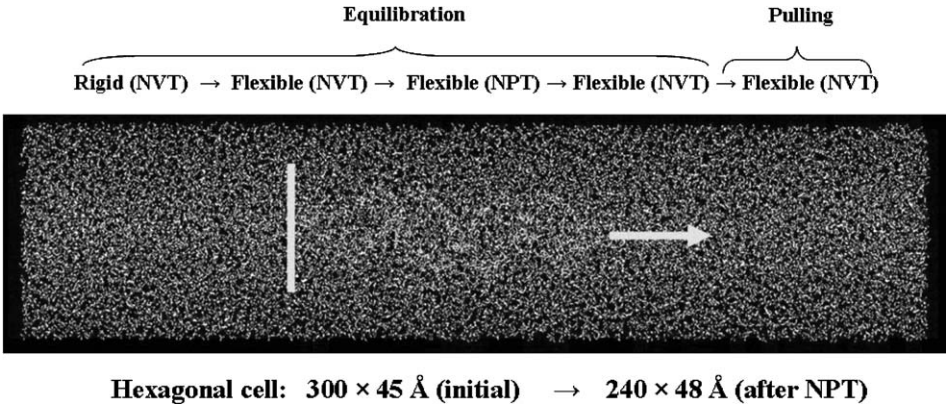


Fig. 7.2. Simulation set-up, protocol and details of methods (see Color Plate 14).

would not be possible because the box would be rapidly shrinking along the direction parallel to the helix axis at the same time as the DNA is stretched due to counter-balancing virial forces in the calculation of the pressure tensor. The stretching was started by adding a same tensile force simultaneously on the O3′ and O5′ atoms at the free end of the DNA along the helix axis. To make the force increase smoothly it was coupled to the simulation time step. The force increment was chosen as 0.25 fN at each long MD step of 2.0 fs. The initial tensile force of 18.75 fN and the size of the force increment of 0.25 fN were both estimated as reasonable compromises.

It should be stressed that the fastest stretching rates applied in a single-molecule experiment [17] have been roughly 1 m/s. This rate is far too slow for any realistic time scales covered by atomistic simulations. It would take roughly 10 000 years to perform our simulation with the fastest currently available computer at the same speed as in the experiments. In our simulation the stretching rate started from roughly 1 m/s growing monotonously to 30 m/s (67mph – gives easily a speeding ticket on the highway in UP of MI). Clearly the simulated situation thus corresponds to a more dynamical process than the quasi-static elongation, carried out in the experiments. However, the main features obtained from the simulations should be similar to those obtained in experiments. Simulations should still reveal the major structural changes, although the conformational phase-space may not be sampled for all the smallest details. In separate control simulations, the pulling was stopped at several points to control how the force changed when the system was allowed to relax. Very small changes were observed, indicating that the system, including solvent water, was not far from its equilibrium during the entire simulation. The thermal velocity from Maxwell–Boltzmann distribution for water at 300K is close to 600 m/s.

For a discussion on non-equilibrium work in systems like DNA under longitudinal stress, and specifically the non-equilibrium work relation of Jarzynski, the interested reader is referred to the Appendix.

7.3.2 Tensile load versus elongation

The pulling force (in pN) against the elongation, relative to the average equilibrium length of the simulated (initially straight) double-helical 22-mer is plotted in (Fig. 7.3). In the simulation the pulling force was increased with 0.25 fN at every 2 fs, monitoring continuously the length of the 22-mer. At the end of the simulation, the DNA is about 2.3 times longer than it was in the beginning, stretched almost completely with an external force of 600 pN without any constraints put on its torsional motion. Although the absolute tensile force values are certainly model-dependent, it is encouraging to see that it fits nicely in the same force window as reported from the corresponding experiments. The simulated curve displays two distinct regimes characterized by a fast elongation and a slow length extension with an intermediate region in between.

The corresponding force intervals are roughly from 37.5 to 85 pN, 85 to 165 pN and 165 to over 600 pN, respectively. The rapid extension of the 22-mer in the fast elongation regime is roughly 1.75-fold. At the end of the intermediate region, it is about twice its normal average length. Finally, at an external force of 600 pN, when the resisting force from the DNA force field is very large, the strain in the 22-mer is 2.3. Compared with

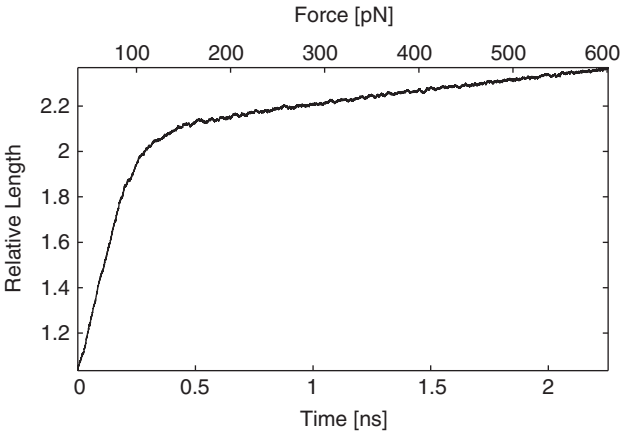


Fig. 7.3. Relative length of the double helical 22-mer as a function of external tensile force (and time).

corresponding curves from elongation of dsDNA without a nick, we can see that the experimental fast elongation regime ends at about 115 pN when the DNA is about 1.6 times its initial length. More force is required in the simulation to complete the fast elongation region and the plateau corresponding to a phase transition is therefore not as flat as in the experiments. This is most likely as a result of the high elongation rate in the simulation.

7.3.3 Path of least resistance

When the DNA is stretched in the simulation its resisting force comes from the used force field, an empirically parametrized function which is (somewhat arbitrarily) divided into components describing the bond stretching, angle bending, and rotation barriers around the bonds; the torsional angles. In addition, there are both short-range van der Waals and long-range electrostatic interactions between the atoms separated by three or more bonds in DNA. The functional form of the CHARMM force field for nucleic acids [53,57] is given as:

$$
\begin{aligned}
V = & \sum_{\text{bonds}} K_{\text{b}}(r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_{\text{a}}(\theta - \theta_{\text{eq}})^2 \\
& + \sum_{\text{torsions}} K_{\text{b}}(1 + \cos(m\phi)) \\
& + \sum_{i<j} \left\{ 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i\, q_j}{4\pi\varepsilon_0 r_{ij}} \right\}
\end{aligned}
\tag{1}
$$

Time development of all these contributions (given in Eq. (1)) to the total internal energy are plotted in Fig. 7.4. The energies (in kJ/mol/base-pairs) give an overview of how the pulling is responded to by the internal degrees-of-freedom of the model DNA.

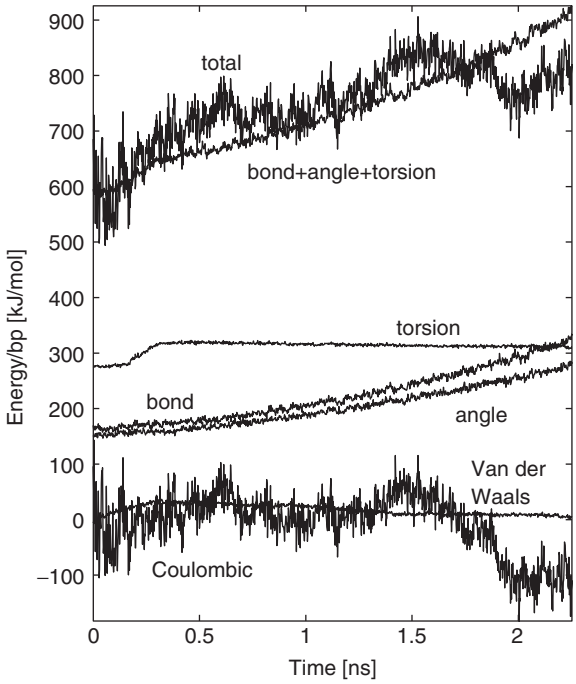


Fig. 7.4. Internal interaction energies (in kJ/base-pair) of the 22-mer as a function of time. Covalent bonds, covalent angles and torsional angles are marked as bond, angle and torsion, respectively. Lennard–Jones 1-4 bonded interactions and Coulombic 1–4 bonded interactions are marked as Van der Waals and Coulombic, repectively. Total bonded, excluding the 1–4 interactions, are marked as bond+angle+torsion. All bonded intramolecular interactions summed up are marked as total.

Each point in the curve is a running time-average from the past 2 ps. We can see that the increase in the covalent bond (blue) and angle (red) energies is monotonously quadratic, reflecting the harmonic character of the force field terms for these two contributions when their deviation from the equilibrium positions r_{eq} and θ_{eq} increases. The characteristic features from the elongation curve (see Fig. 7.3) are best reflected in the torsional energies (green), showing three different regions coinciding with those seen in the force curve. Besides the fairly sharp rise (16 percent), taking place in the intermediate region of the pulling force (from 200 ps to 500 ps), there is no further tension build-up due to rotations around the bonds building up the DNA conformation. A closer look at the dynamics of torsional angles shows that the sharp rise is due to a collective fast conformational transition, taking place in the sugar–phosphate backbone (see separate section below). From 500 ps towards the end of the simulation the torsional energy is even decreasing slightly. Initially the sum of the covalent bond and angle energy components is roughly equal to the torsion component. Successively the bond energy becomes equal to the torsion energy after about 2.1 ns when the strain in the 22-mer is about 2.25. The total energy from these three bonded interactions is given as the purple curve in the upper part of Fig. 7.4.

Based on the internal energies we can conclude that the studied 22-mer, can easily, i.e., with a small external load, be elongated to a strain of 1.7–1.8. This strain is possible through energetically cheap changes in covalent bonds/angles and torsion angles. In the strain range from 1.8 to 2.05, the torsional angles still dominate the response of the DNA to a steadily increased load. Above a strain of about 2.05, the bond and angle contributions become, however, increasingly important. The location of this last crossover point on the strain axis obviously depends on the harmonic form of the bond and angle force field components and the used force constants.

7.3.4 Can we rely on a stretched force field?

An important question is the general applicability of the modelling force fields originally developed for equilibrium conditions. In other words, how far can their validity be stretched? Like most force fields, CHARMM has harmonic functional forms for the bonds and angles (Eq. (1)). At ambient temperatures and otherwise normal conditions for biological systems these potentials are populated close to the well bottom, thermally fluctuating around their minimum positions. Anharmonic potentials for bond stretching would be more realistic at off-equilibrium conditions, such as a very large tension.

In order to evaluate the validity of the harmonic approximation, we have compared the average backbone bond and angle values of the initial and final conformations. The mean bond stretch shift between them was found to be 0.08 Å. For an equilibrium bond length of 1.5 Å, this corresponds to a about 5 percent strain. This means that the bond variables are still safely near the bottom of the potential well. If the stretching were to be continued further, it would become important to replace the harmonic bonds with more realistic anharmonic bond stretching potentials, which also allow a pulling until the strands break. Comparing in the same way, the backbone angles show an average opening of 7°. Evaluation of the validity of the harmonic form in the case of angles is more difficult. If, however, the energies of individual backbone bonds and angles are compared in Fig. 7.4, it can be noticed that their increases are comparable (blue and red curves). We can assume that the harmonic approximation is operational but it is approaching its validity limit in the final over-stretched conformation of the DNA of this simulation.

7.3.5 Internal interactions short and long range

Figure 7.4 also contains the energies for interactions between the DNA atoms, which are not mediated through the covalent bonds, but rather through the space. These can be conceptually divided between short–range van der Waals type of interactions, calculated using the Lennard-Jones potential term, and the long-range Coulombic interactions (last two terms in Eq. (1)). Major individual contributions to these interactions in biomolecules come from steric contacts due to close encounters and intramolecular hydrogen bonds. Exactly how the stretching affects these interactions is difficult to analyse in detail. In the case of the van der Waals energies, shown in Fig. 7.4 (cyan), there is a rise up until 500 ps as in the torsional energies, although it is not as sharp and it starts from the beginning

of the simulation. A major part of its origin is in the steric, so-called 1–4 interactions, involving the terminal atoms in each torsional angle. Like the torsional energy (green), the van der Waals energy is stabilized after the transition regime, and even slightly decreases as the simulation proceeds towards its completion.

On account of the short-range interactions, there is a major restructuring taking place among the bases. As will be discussed below, the base pairs will open and the original stacking of bases is lost, starting at an early stage of the stretching. This is initially a costly process, as can be seen from van der Waals (cyan) and Coulombic (yellow) terms. A new rather spectacular stacking structure is created when the backbone is stretched further. This helps to stabilize the short-range energy. Coulombic interactions (yellow) come both from short and long distances, causing them to fluctuate more heavily than the more smooth van der Waals energies (cyan). The Coulombic energy shows, besides both high- and low-frequency oscillations, the same type of behaviour as the van der Waals energy until 1.7 ns of the stretching, corresponding to a 2.1 relative elongation. Interestingly, it goes thereafter down to a low (strongly attractive) level for the over-stretched DNA. In the final structure the negatively charged phosphates find their maximum distances, thereby minimizing their mutual repulsion. The distance between the strands decreases somewhat during the pulling; allowing oppositely charged atomic charges to come closer to each other. Based on Fig. 7.4, an informative picture can be obtained how dsDNA responds to pulling and how the new conformation is stabilized as a complex interplay of interactions and mechanisms. The total internal energy after adding all the contributions in Eq. (1) is shown as the black curve in Fig. 7.4. DNA needs both water and counter-ions to stabilize its structure under normal conditions. Hydration and specific water-DNA interactions are very important even for the stretched form, where the building blocks of DNA: phosphate, sugar and the bases, are all exposed to surrounding water. Besides strong hydrogen bonds to water, the base stacking is stabilized by hydrophobic interactions.

7.3.6 Fast transition in sugar-phosphate backbone

As a complement to the behaviour of internal energies, we have analysed the backbone torsion angles (see Fig. 7.5 for definitions), responsible for the conformational transition already indicated in Fig. 7.4. The analysis of torsional angles was carried out using the Curves program [60–62]. The dynamical trajectories of the conformational parameters reveal that α and γ parameters already undergo large transitions 200–300 ps after the stretching was started. The α torsion angle around the P-O5 bond changes about 110° from 300° to 190° while the γ angle around the C–C bond connected to the sugar changes roughly the same amount from 50° to 160°. These large changes in the torsion angles explain the form of the length versus force (time) curve in the intermediate regime (*cf.* Fig. 7.3). In Fig. 7.6(a), the population distribution of α torsion angle is given as a function of time after the stretching was started, showing a very rapid transition. This takes place when DNA is elongated about 1.7 times from its initial length. The distribution around the new value stabilizes rapidly and becomes narrower towards the end of the simulation. The same is shown for the γ angle in Fig. 7.6(b), showing a rapid transition at the same point on the time (force) axis but in the other direction. In fact, they have an

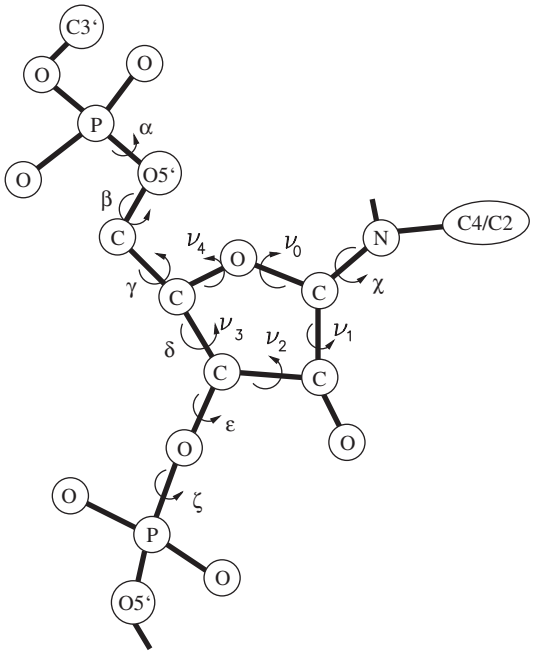


Fig. 7.5. Definition of torsional angles in DNA.

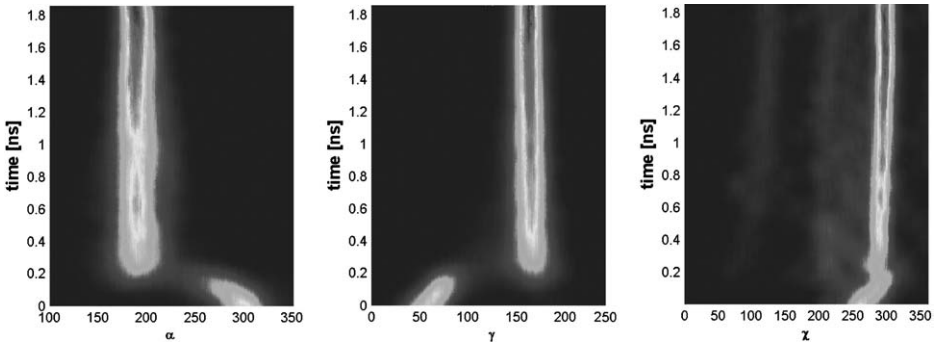


Fig. 7.6. Time development of selected torsional angles. From the left: (a) α, (b) γ and (c) χ (see Fig. 7.5 for definitions) (see Color Plate 15).

appearance of mirror images. These two angles are strongly correlated in the helix structure. A third torsional angle change connected to the transition is that for the angle χ around the glycosyl bond to the base, shown in Fig. 7.6(c). It changes from about 100° to about –75°. Its position stabilizes somewhat slower as it is connected to base tilt and the stacking of the bases. All the other torsional angles (except α, γ and χ) stayed very close to their initial values during stretching.

7.3.7 Sugar–phosphate chains zipped up by the bases

Figure 7.7 illustrates the entire final structure of the DNA after 2.6 ns. The pulling direction in the figure is to the right. The entire DNA is uniformly elongated, giving confidence that the pulling speed has not been too high and the system has not been too far from equilibrium during the stretching. A zoomed stereo image is shown in Fig. 7.8, giving a closer

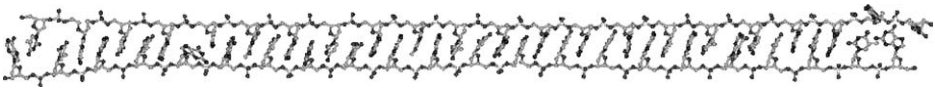


Fig. 7.7. Final conformation of the over-stretched 22-mer double helix DNA. The pulling direction is from the left to the right.

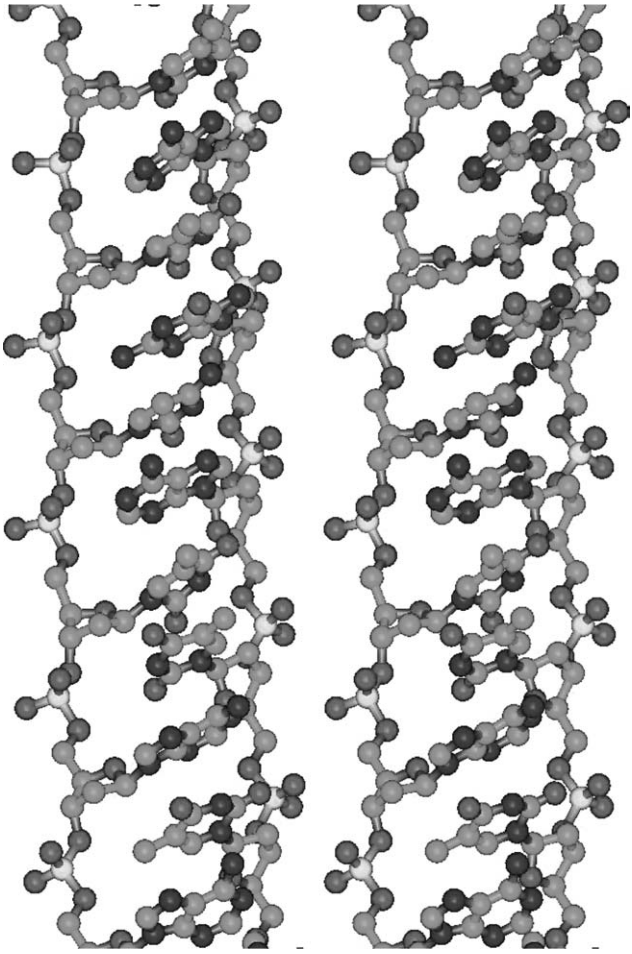


Fig. 7.8. Stereoscopic projection of a part of the base-stacked structure (see Color Plate 16).

look at the remarkable base-stacked zipper-like structure, formed spontaneously and in a collective manner as the result of stretching in explicit water solution. In the final structure, the screw angle of the 22-mer is less than 45°. In this base-stacked structure, the individual bases are both tilted and stacked on top of each other by the straightened backbone strands. The bases are exposed on the major groove side of the original helical structure. Having access to virtual reality equipment, we analysed simulation data in a VR-CAVE. The three-dimentional animation confirms a partial appearance of the structure known as S-DNA [13]. It appeared soon in the fast elongation regime in separate substructures of about five base-pairs long. Some base pairs break before they form the S-like structure. The already formed S structure elements soon disappear as their base pairs start to break. The last base pairs break in the transition regime. The breaking of base pairs is connected to the narrowing distance between the backbone strands. This narrowing can be seen in Fig. 7.7. The inter-strand distance has been kept constant in the anchored end of the dsDNA during the stretching. We can also see that the base stacking is not completed in the end regions because of the constraint. Development towards the structure shown in Fig. 7.7 starts early and progresses rapidly. The first base-stacked unit was already formed after 90 ps. Later, when the distances between adjacent bases in both strands increase, the stack formation takes place spontaneously. Visualization shows how separate stacked units of several bases are created and how they merge into each other as pulling continues. One important question is why this rather remarkable base-stacked structure has not been observed previously in computer simulations? One probable reason is that an implicit solvent has been used in earlier computer modelling works [44,51,55]. Water molecules and counter-ions are an integral part of nucleic acids and their function. Specific DNA-water interactions are important for hydration of DNA, stabilizing the conformation with a balanced interplay of hydrophilic and hydrophobic interactions. The base-stacked structure, observed in our work, is to a large extent driven by surrounding water.

7.3.8 Comparison with experiment

Our computer experiment, together with experimental single-molecule investigations (and previous modelling studies), show that a small amount of force (100 pN), within a narrow interval is needed to extend a dsDNA in a torsionally unconstrained elongation to its double length. Our simulation also shows that the conformational transition is characterized by changes of 110° for both α and γ torsional angles in opposite directions (Fig. 7.6). The transitions in both these angles basically allow an elongation of the phosphate-sugar backbone, while the latter transition affects even the sugar and consequently the base attached to it, thereby destabilizing the hydrogen bonds to its complementary base. Changes in the χ angles are connected to a strong a tilt of the associated base. All the bases become exposed on the same (major groove) side of the backbone and the extended length of the backbone and shorter distance between the strands allow the bases to become stacked on top of each other. Has anything of this been observed in experiments? Nishinaka *et al.* [63] observed rotation of bases in their NMR study of extended dsDNA. The bases are strongly tilted (Figs 7.7 and 7.8), which was observed by Nord en *et al*. [64,65] in their linear dicroism study, where they propose a similarly tilted model as is obtained in our work. Also, the X-ray diffraction study of Greenall and co-workers [66]

on DNA fibres under tension gives zero or positive birefringence from the stretched regions, indicating that the bases are highly tilted.

7.3.9 Biological relevance?

The RecA and RecA-related proteins are essential in DNA strand exchange processes such as homologous recombination but also in repair and replication processes [67–78]. The RecA-mediated strand exchange mechanism can be modelled as four stages and summarized by the following scheme [73]:

$$\begin{aligned} &n\text{RecA} + \text{ssDNA}^{\mathbf{I}} \Rightarrow (\text{A}) \Rightarrow \text{RecA}_n\text{:ssDNA} + \text{dsDNA}^{\mathbf{CO}} \Rightarrow (\text{B}) \Rightarrow \\ &\text{RecA}_n\text{:}\{\text{ssDNA...dsDNA}\} \Rightarrow (\text{C}) \Rightarrow n\text{RecA} + \text{dsDNA}^{\mathbf{IC}} + \text{ssDNA}^{\mathbf{O}} \end{aligned} \tag{2}$$

I, **C** and **O** denote incoming, complementary and outgoing strands, respectively, while ":" is used here for the filament complex between RecA polymer and DNA strands and denotes hydrogen bonds between the strands involving the bases. A nucleotide cofactor will become involved in the reaction. In the first step, (**A**), monomers of RecA are attached to the incoming single-stranded DNA (ssDNA), covering roughly three bases of it each. An approaching dsDNA is associated with the helical filament (**B**). At this stage the entire filament is extended and the DNA strands are somewhat unwound. The strands are swapped after a match is found between the incoming single strand and its complementary strand in the double helix. Finally, the dsDNA and ssDNA are released and the RecA polymer dissociates to monomers (**C**).

Electron microscopy, NMR, linear dicroism, as well as modelling studies, have all contributed to a picture of the RecA-mediated homologous recombination [67–78]. Several aspects in the picture are still less clear, however. The origin of the free energy to drive the reaction is one. Another, is the dsDNA conformation in the RecA_n: ssDNAdsDNA complex. The most important problem to be solved is of course the molecular mechanism for homologous recombination. In the case of the dsDNA conformation, the early modelling [44,79], showed that dsDNA opened its bases on the minor groove side upon stretching [68], although the stability of the major groove model was also tested for the sake of completeness. Kosikov *et al*. [55] found the opened major groove as a favourable interaction site for ssDNA in the RecA_n: ssDNAdsDNA complex.

7.4 CONCLUSIONS

In our simulation, the dsDNA opens spontaneously on the major groove side. In addition, we already observe a collective opening of the base pairs at a *low* pulling force and the zipper-like stacking of bases as a natural consequence after the distance between the adjacent intra-strand bases is doubled. This behaviour has, to the best of our knowledge, not been seen or suggested previously. Still, it does not really come as a surprise. DNA is an amazing molecular construction of nature from the beginning of evolution. Its ability to adapt and survive is truly impressive. It still keeps many secrets from us. The base-stacked conformation in Figs 7.7 and 7.8 is both similar and different from S-DNA. It is simply

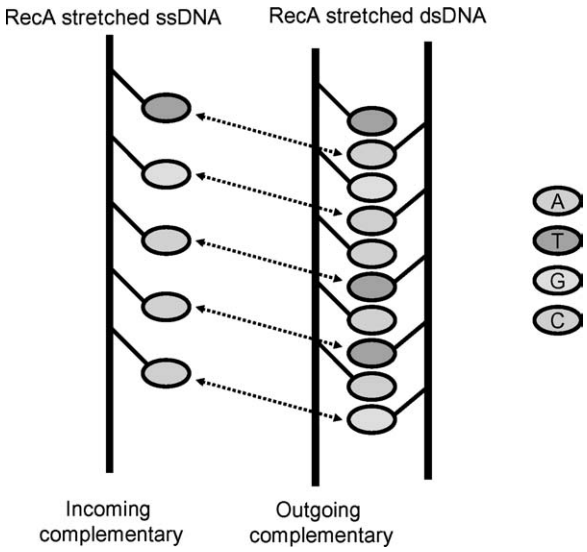


Fig. 7.9. Schematic close encounter between the stretched DNA and the incoming single strand looking for a match (see Color Plate 17).

one step further from it. Because the transition takes place in an early stage of stretching, requiring little energy, it may well occur even in biological processes. For example, it would give a simple platform for a homologous matchmaking in the process of strand exchange between ssDNA and dsDNA (see Fig. 7.9). This is naturally just a hypothesis, which needs to be confirmed independently.

Appendix

One of the latest developments in fundamental thermodynamics concerns the relation between equilibrium free-energy changes and their coupling to non-equilibrium changes. Much work remains to be done on the topic of non-equilibrium systems, although the difficulty from the onset looks staggering. Whereas a system at equilibrium is defined by a relatively small number of parameters (say the number of particles, the temperature and the pressure), the non-equilibrium state can in general depend on any and all the parameters of the system pertaining to its set-up. Such an everyday property as the temperature of the system is not in general well defined away from equilibrium, making analysis of the system arduous. In some sense, there are many more ways in which the system can be prepared away from equilibrium than in equilibrium. However, there turns out to be ways of overcoming some of the difficulties in the case of relating free-energy changes of equilibrium processes to their non-equilibrium counterparts.

The non-equilibrium work theorem of Jarzynski in 1997 [80] gives us the remarkable identity $<e^{-\beta W}> = e^{-\beta \Delta F}$ where W is the work performed in taking the system between two states, let us call them A and B. ΔF is the equilibrium free-energy change of the process. β is $1/kT$, where T is the temperature of the heat reservoir which keeps

our system at thermal equilibrium in the initial state. We average over the ensemble of initial conditions (for Hamiltonian systems; in the case of stochastic processes we instead average over the realizations of the change). The implications of this relation are not fully understood in terms of their practical use, but the identity is general in the sense that the work does not have to be performed at equilibrium. As we reach state B, we might be away from equilibrium, but the heat reservoir will relax the system without additional work being performed by us. Averaging work in this way weights the contributions so that low- work trajectories (essentially negative irreversible work) are more important. The apparent deviations from the second law of thermodynamics, the fact that we sometimes have to use less work than in the reversible case, are expected as the system size and the observed time are small. This is the subject of study in fluctuation theory and so-called fluctuation-dissipation theorems handle these well at small deviations from equilibrium.

Phenomenologically, this non-equilibrium work relationship tells us that we can find free energy changes for a process from multiple data sets on the work that is needed to transfer the system between the two states. With DNA pulling experiments in mind, this would mean that we can get free energies of extension by performing many sets of pulling (at any speed) and then use the Jarzynski relationship to average the data. However, for the equality to be valid in the strict sense, the amount of trajectories to average over must be infinite. Of course this is something we cannot afford experimentally (neither in simulations nor in the laboratory). If we restrict ourselves to a finite number of trajectories, we can use the relationship above as an estimator of the free energy. For this to be of significant value, we need to know what kind of convergence criteria we can expect and the analysis of the bias that we introduce is discussed in the work of Gore *et al*. [81]. Generally, since the instances at the tail of the work distribution are rare (by definition), and the theorem tells us that these are important, we expect experiments with small deviations from equilibrium to be preferred over strongly perturbed systems. In essence, this puts us in the linear response regime. As we perturb the system further from equilibrium, non-Gaussian behaviour in the work distributions is to be expected. It should be noted here that the use of the Jarzynski relationship might be associated with large systematic and statistical uncertainties [82] exactly for this reason. These effects can be minimized by using the cumulant expansion rather than the exponential itself [83].

The cumulant expansion, correct to the fourth order, is given below. $W_{0\to1}$ denotes the average of W from point 0 to 1 on the free-energy coordinate, and $\sigma^2_{0\to1}$ denotes the variance of this work.

$$\Delta F^{(2)} = \frac{1}{2}(W_{0\to1} - W_{1\to0}) - \frac{\beta}{12}(\sigma^2_{0\to1} - \sigma^2_{1\to0}) \tag{3}$$

From the point of view of computational evaluation of the free energy, one should in general use a method of switching from one state to the next with a variance in the work used around $(kT)^2$. By also dividing the free-energy coordinate into intervals, which are evaluated piece by piece, forwards and backwards, such a fast switching method is trivial to parallelize.

For a system where the trajectory consists of extension, like in the case of AFM experiments, the free energy profile of extension along z, $G_0(z)$, is expressed as:

$$G_0(z) = -\beta^{-1}\ln < \delta(z - z_t)\mathrm{e}^{-\beta\left[\int_c F(z,t)\,\mathrm{d}z - u(z_0,0)\right]} > \tag{4}$$

where z_t is the distance at time t, $F(z,t)$ is the pulling force, and $u(z_0, 0)$ is the position of the harmonic spring at the beginning of the measurement (thus we include the energy stored in the spring at time 0). The method for computing the free energy from multiple force-extension curves is given in [84].

In summary, we note that smaller perturbations from equilibrium (less dissipated work) lead to a reduced number of trajectories to average over [84,85]. The question of few and slow or many and fast therefore leans towards few and slow. Also, several authors note that the errors for a given amount of computational time are not smaller than traditional methods for evaluating free-energy changes.

7.5 REFERENCES

1 T. Hirschfeld, Appl. Opt., 15 (1976) 2965–2966.
2 G. Binnig and H. Rohrer, Helv. Phys. Acta, 55 (1982) 726–730.
3 G. Binnig, H. Rohrer, C. Gerber and E. Weibel., Phys. Rev. Lett., 49 (1982) 57–61.
4 C. Binnig, C. Quate and C. Gerber, Phys. Rev. Lett., 56 (1986) 930–933.
5 S.B. Smith, L. Finzi and C. Bustamante, Science, 258 (1992) 1122–1126 .
6 S.B. Smith, Y. Cui and C. Bustamante, Science, 271 (1996) 795–799.
7 T.R. Strick, J.F. Allemand, D. Bensimon, A. Bensimon and V. Croquette, Science, 271 (1886) 1835–1837.
8 U Bockelmann, B. Essevaz–Roulet and F. Heslot, Phys. Rev. Lett., 79 (1997) 4489–4492.
9 C. Bustamante, J.F. Marko, E.D. Siggia and S.B. Smith, Science, 265 (1994) 1599–1600.
10 A. Bensimon, A. Simon, A. Chiffaudel, V. Croquette, F. Heslot and D. Bensimon, Science, 265 (1994) 2096–2098.
11 T. Thundat, D.P. Allison and R.J. Warmack, Nucl. Acid Res., 22 (1994) 4224–4228.
12 D. Bensimon, A. Simon, V. Croquette and A. Bensimon, Phys. Rev. Lett., 74 (1995) 4754–4757.
13 P. Cluzel, A. Lebrun, C. Heller, R. Lavery, J.-L. Viovy, D. Chatenay and F. Caron, Science, 271 (1996) 795–799.
14 B. Essevaz-Roulet, U. Boeckelmann and F. Heslot, Proc. Natl. Acad. Sci. USA, 95 (1997) 11935–11940.
15 J.F. Leger, G. Romano, A. Sarkar, J. Robert, L. Bourdieu, D. Chatenay and J.F. Marko, Phys. Rev. Lett., 83 (1999) 1066–1069.
16 M. Rief, H. Clausen-Schaumann and H.E. Gaub, Nature Struct. Biol., 6 (1999) 346–349.
17 H. Clausen-Schaumann, M. Rief, C. Tolksdorf and H.E. Gaub, Biophys. J., 78 (2000) 719–724.
18 T.E. Fischer, P.E. Marzalek and J.M. Fernandez, Nature Struct. Biol., 7 (2000) 719–724.
19 C. Bustamante, S.B. Smith, J. Liphardt and D. Smith, Curr. Opin. Struct. Biol., 10 (2000) 279–285.
20 T. Strick, J.-F. Allemand, V. Croquette and D. Densimon, Progr. Biophys. Mol. Biol., 74 (2000) 115–1140.
21 L.H. Pope, M.C. Davies, C.A. Laughton, C.J. Roberts, S.J.B. Tendler and P.M. Williams, Eur. Biophys. J., 30 (2001) 53–62.
22 A. Sarkar, J.-F. Leger, D. Chatenay and J.F. Marko, Phys. Rev. E, 63 (2001) 051903-1 - 051903-10.
23 E.B. Brauns, L. Madaras, R.S. Coleman, C.J. Murphy and M.A. Berg, Phys. Rev. Lett., 88 (2002) 158101–1 - 158101-4.
24 C. Bustamante, Z. Bryant and S.B. Smith, Nature, 421 (2003) 423–427.
25 Z. Bryant, M.D. Stone, J. Gore, S.B. Smith, N.R. Cozzarelli and C. Bustamante, Nature, 424 (2003) 338–341.

26 R. Strick, M.-N. Dessinges, G. Charvin, N.H. Dekker and J.-F. Allemand, Rep. Progr. Phys., 66 (2003) 1–45.
27 C. Danilowicz, V.W. Coljee, C. Bouzigues, D.K. Lubensky, D.R. Nelson and M. Prentiss, Proc. Natl. Acad. Sci. USA, 100 (2003) 1694–1699.
28 G. Charvin, A. Vologodskii, D. Bensimon and V. Croquette, Biophys. J., 88 (2005) 4124–4136.
29 J.F. Marko, Proc. Natl. Acad. Sci. USA, 94 (1997) 11770–11772.
30 A. Rich, Proc. Natl. Acad. Sci. USA, 95 (1998) 13999–14000.
31 A. Engel and D.J. Müller, Nature Struct. Biol., 7 (2000) 715–718.
32 E. Evans, Ann. Rev. Biophys. Biomol. Struct., 30 (2001) 105–128.
33 C. Gosse and V. Croquette, Biophys. J., 82 (2002) 3314–3329.
34 M. Rief and H. Grubmüller, Chem. Phys. Chem., 3 (2002) 255–261.
35 M. Hegner and W. Grange, J. Muscle Res.Cell Motil., 23 (2002) 367–375.
36 J. Zlatanova and S.H. Leuba, J. Muscle Res. Cell Motil., 23 (2002) 377–395.
37 M.C. Williams and I. Rouzina, Curr. Opin. Struct. Biol., 12 (2002) 330–336.
38 J.-F. Allemand, D. Bensimon and V. Croquette, Curr. Opin. Struct. Biol., 13 (2003) 266–274.
39 W. Zhang and X. Zhang, Progr. Polym. Sci., 28 (2003) 1271–1295.
40 J. Zlatanova and S.H. Leuba, Biochem. Cell Biol., 81 (2003) 151–159.
41 S.A. Harris, Contemporary Phys., 45 (2004) 11–30.
42 M.C. Williams, J.R. Wenner, I. Rouzina and V.A. Bloomfield, Biophys. J., 80 (2001) 874–881.
43 R. Lavery, K. Zakrzewska and H. Sklenar, Comput. Phys. Commun., 91 (1995) 135–158.
44 A. Lebrun and R. Lavery, Nucl. Acids Res., 24 (1996) 2260–2267.
45 I. Rouzina and V.A. Bloomfield, Biophys. J., 80 (2001) 882–893.
46 M.C. Williams, J.R. Wenner, I. Rouzina and V.A. Bloomfield, Biophys. J., 80 (2001) 1932–1939.
47 J. Watson and F. Crick, Nature, 171 (1953) 737–738.
48 M. Wilkins, R. Gosling and W. Seeds, Nature, 167 (1951) 759–760.
49 H.W. Wyckoff, X-ray diffraction analysis of the structure of deoxyribonucleic acid, PhD thesis, Mass. Inst. of Technol., 1955.
50 J.F. Allemand, D. Bensimon, R. Lavery and V. Croquette, Proc. Natl. Acad. Sci. USA, 95 (1998) 14152–14157.
51 M.W. Konrad and J.I. Bolonick, J. Am. Chem. Soc., 118 (1996) 10989–10994.
52 A.D. MacKerell and G.U. Lee, Euro. Biophys. J., 28 (1999) 415–426.
53 A.D. Mackerell, J.Wiorkiewicz-Kuczera and M. Karplus, J. Am. Chem. Soc., 117 (1995) 11946–11975.
54 W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey and M.L. Klein, J. Chem. Phys., 79 (1983) 926–935.
55 K.M. Kosikov, A.A. Gorin, V.B. Zhurkin and W.K. Olson, J. Mol. Biol., 289 (1999) 1301–1326.
56 N. Foloppe and A.D. MacKerell, J. Comput. Chem., 21 (2000) 86–104.
57 A.D. MacKerell and N.K. Banavali, J. Comput. Chem., 21 (2000) 105–120.
58 K. Toukan and A. Rahman, Phys. Rev. B, 31 (1985) 2643–2648.
59 A.P. Lyubartsev and A. Laaksonen, Comp. Phys. Commun., 128 (2000) 565–589.
60 R.Lavery and H. Sklenar, J. Biol. Struct. Dynam., 6 (1989) 63-91.
61 R. Lavery and H. Sklenar, J. Biol. Struct. Dynam., 6 (1989) 655–667.
62 R. Lavery and H. Sklenar. Curves 5.2 – helical analysis of irregular nucleic acids. Technical report, Laboratoire de Biochimie Theorique, CNRS URA 77, Institut de Biologie Physico-Chimique, Paris, 1997.
63 T. Nishinaka, Y. Ito, S. Yokoyama and T. Shibata, Proc. Natl. Acad. Sci., 94 (1997) 6623–6628.
64 B. Nord en, C. Elvingson, M. Kubista, B. Sjöberg, H. Ryberg, M. Ryberg, K. Mortensen and M. Takahashi, J. Mol. Biol., 226 (1992) 1175–1192.
65 B. Nord en, P.Wittung-Stafshede, C. Ellouze, H.-K. Kim, K. Mortensen and M. Takahashi, J. Biol. Chem., 273 (1998) 15682–15686.
66 R.J. Greenall, C. Nave and W. Fuller, J. Mol. Biol., 305 (2001) 669–672.
67 S.C. Kowalczykowski and R.A. Krupp, Proc. Natl. Acad. Sci., 92 (1995) 3478–3482.
68 G. Bertucat, R. Lavery and C. Prévost, Biophys. J., 77 (1999) 1562–1576.
69 M. Hegner, S.B. Smith and C. Bustamante, Proc. Natl. Acad. Sci., 96 (1999) 10109–10114.
70 A. Kuzminov, Proc. Natl. Acad. Sci., 98 (2001) 8461–8468.
71 E.H. Engelman, J. Mol. Biol., 309 (2001) 539–542.

72 M. Feingold, Physica E, 9 (2001) 616–620.
73 S.F. Singleton and J. Xiao, Biopolymers, 61 (2002) 145–158.
74 A.A. Volodin, E.A. Smirnova, T.N. Bocharova and R.D. Camerini-Otero, FEBS Lett., 546 (2003) 203–208.
75 F.S. Katz and F.R. Bryant, J. Biol. Chem., 278 (2003) 35889–35896.
76 S. Patel and J.S. Edwards, DNA repair, 3 (2004) 61–65.
77 S.L. Lusetti and M.M. Cox, Ann. Rev. Biochem., 71 (2002) 71–100.
78 E.H. Engelman, Nature Revs. Mol. Cell Bio., 4 (2003) 621–630.
79 G. Bertucat, R. Lavery and C. Prévost, J. Biol. Struct. Dynam., 16 (1998) 535–546.
80 C. Jarzynski, Phys. Rev. Lett., 78 (1997) 2690–2693.
81 J. Gore, F. Ritort and C.R.C. Bustamante, Proc. Natl. Acad. Sci. USA, 100 (2003) 12564–12569.
82 J.R. Gullingsrud, R. Braun and K. Schulten, J. Comp. Chem., 151 (1999) 190–211.
83 G. Hummer, J. Chem. Phys., 114 (2001) 7330–7337.
84 J. Liphardt, S. Dumont, S.B. Smith, I. Tinoco and C. Bustamante, Science, 296 (2002) 1832–1835.
85 R.C. Lua and A.Y.G. Grosberg, J. Phys. Chem. B, 109 (2005) 6805–6811.

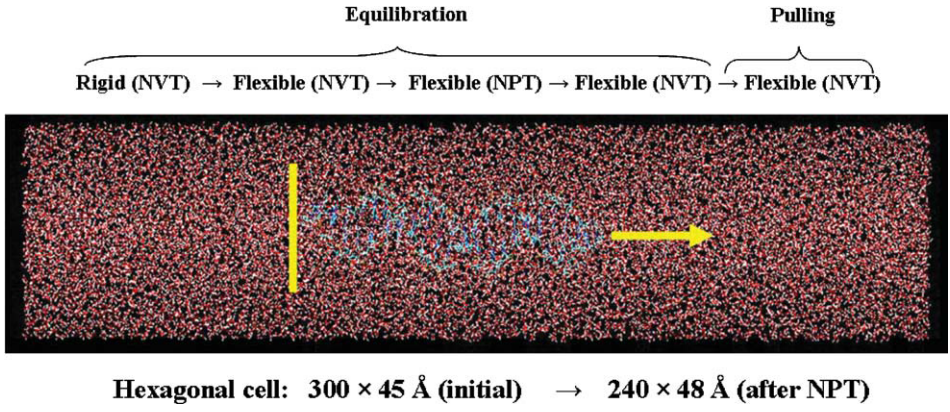


Plate 14. Simulation set-up, protocol and details of methods (illustration appears on page 143 of this volume).

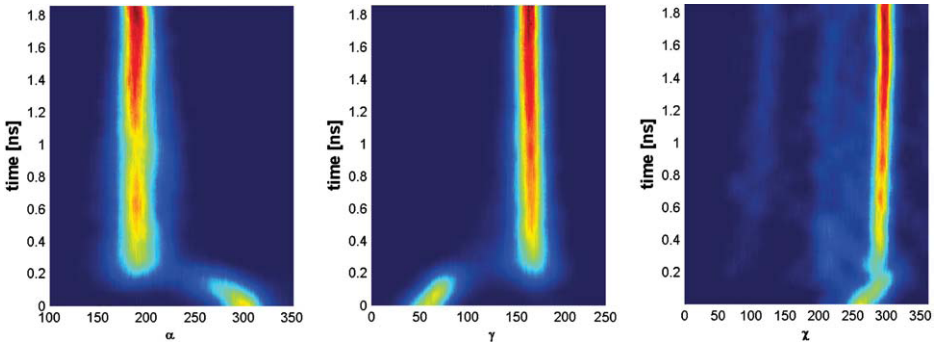


Plate 15. Time development of selected torsional angles. From the left: (a) α, (b) γ and (c) χ (see Fig. 7.5 for definitions) (illustration appears on page 149 of this volume).

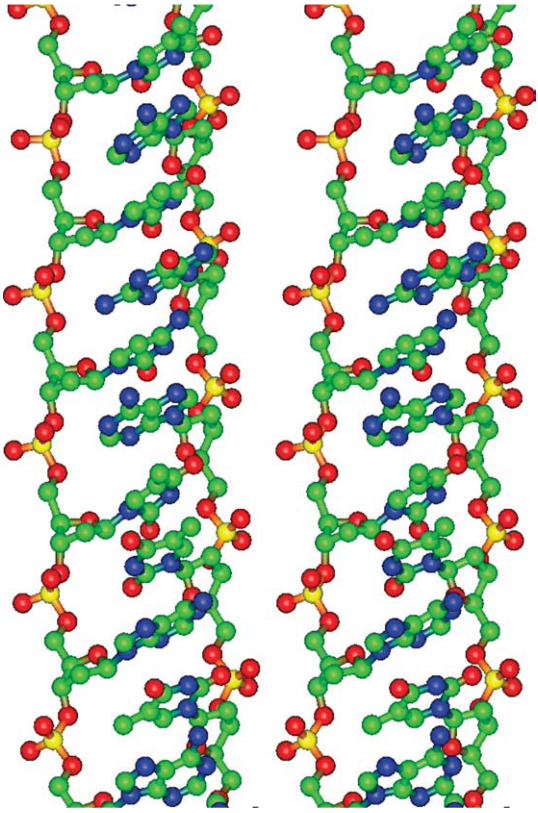


Plate 16. Stereoscopic projection of a part of the base-stacked structure (illustration appears on page 150 of this volume).

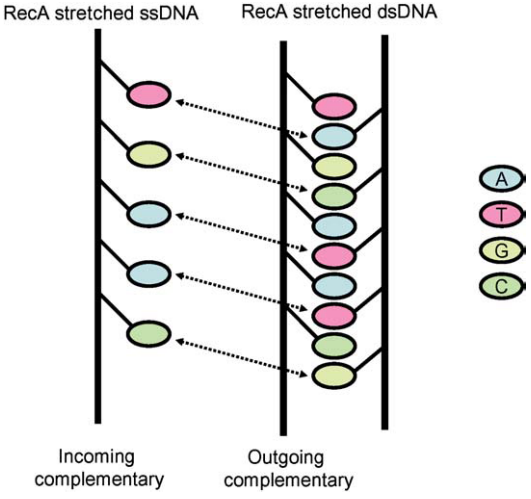


Plate 17. Schematic close encounter between the stretched DNA and the incoming single strand looking for a match (illustration appears on page 153 of this volume).

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 8

On the art of computing the IR spectra of molecules in the condensed phase

Matthias Schmitz and Paul Tavan

Theoretische Biophysik, Lehrstuhl für Biomolekulare Optik, Ludwig-Maximilians Universität München, Oettingenstraße 67, 80538 München, Germany

Abstract

Infrared (IR) spectroscopy is a most important technique for identifying compounds and monitoring their reactions in bio-chemistry. Usually spectra are obtained from condensed-phase samples containing molecules embedded in solvents of varying polarity or in complex and polar protein environments. Upon transfer from the gas phase to the condensed phase, the vibrational frequencies may become sizeably shifted and the lines inhomogeneuosly broadened. For molecules comprising up to about 100 atoms, accurate computations of their gas-phase IR spectra were enabled by the development and widespread accessibility of density functional theory (DFT) about a decade ago. Concerning molecules in the condensed phase, however, the art of accurately computing the IR spectra is an ongoing development, whose computational basis is provided by hybrid methods combining DFT descriptions of molecules with molecular mechanics (MM) models of their condensed phase environment. Here we review the involved physics, the available computational procedures, the achievements, and the perspectives of this DFT/MM based development, which promises to provide detailed insights into the structural dynamics of condensed phase bio-chemical reactions by quantitative descriptions of corresponding IR spectra.

8.1 INTRODUCTION

When a molecule is transferred from the gas phase into a polar solvent or into a protein environment, its vibrational bands, as probed by IR and Raman spectroscopy, are shifted and broadened. These effects are caused by the polarization of the solute molecule, which is induced by the electric field originating from the thermally fluctuating charge distribution of its solvent environment, and by the coupling of the vibrations of the solute to those of the neighbouring solvent molecules. Because the polarization of the molecule and, therefore, also the sizes of the shifts and broadenings depend on the spatial arrangement

of the surrounding charge distribution, the vibrational spectra of a solute molecule may strongly vary with the environment. The left part of Fig. 8.1 provides a relevant example for such solvatochromic frequency shifts.

As a consequence, the vibrational spectra of solute molecules contain information on the structure and dynamics of the charge distribution characterizing their solvent or protein environment. Because IR spectroscopy can provide access to this information, it has become an important tool in the study of biological systems at a molecular level [4,5]. For example, the IR absorption of the quinone chromophores in the photosynthetic reaction centre of bacteria revealed their reduction states and binding modes [6]; in the case of the visual pigment rhodopsin, time-resolved IR studies enabled identification of the structures of the retinal chromophore in short-lived intermediates, which follow the primary photoisomerization triggering vision [7]; phosphate vibrational bands were used to study the hydrolysis of GTP as catalysed by Ras p.21 [10]; the vibrational modes of flavines were shown to provide markers for their binding modes in flavoproteins, e.g., in photolyase or in adrenodoxin reductase [8–11]; information on the secondary structure compositions of proteins and peptides is frequently derived from the amide-bands of the protein backbone [12,13].

In all these cases, a precise computational prediction of the spectrum of the given molecule in its respective environment is necessary, if one wants to decode the experimental IR spectra in order to obtain structural information. It is the aim of this chapter to discuss to what extent currently available computational methods can give access to such predictions and, hence, to the decoding.

A key merit of density functional theory (DFT), as compared with previous Hartree–Fock or costly configuration interaction methods, is the fact that the gas-phase IR spectra of quite large molecules can be speedily computed to very high accuracy,

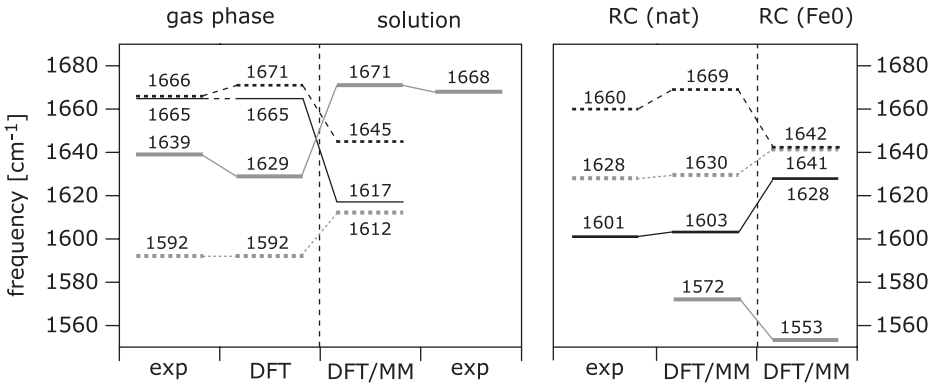


Fig. 8.1. Left graph: spectral locations of vibrational bands observed (exp) for the symmetric (solid) and asymmetric (dashed) C=C (grey) and C=O (black) stretching modes of p-benzoquinone in the gas phase and aqueous solution, respectively, are compared with corresponding harmonic DFT and DFT/MM frequencies. For explanation see the text. Right graph: experimental frequencies of the C1=O (black dashed), C4=O (black solid), C2=C3 (grey solid), and C5=C6 (grey dashed) vibrations of the quinone QA in the photosynthetic reaction centre (RC) of *Rb. sphaeroides* are compared with DFT/MM results for a natively charged (nat) and for an uncharged iron cofactor (Fe0). Data taken from Refs [1–3].

which is required for vibrational analysis [1,14,15]. The B3LYP functional [16,17], for instance, can reproduce harmonic frequencies and anharmonic shifts within a few wavenumbers [15]. Harmonic intra-molecular force fields calculated with the BP86 functional [18,19], though overestimating the experimental harmonic force constants, match the anharmonic fundamentals very well. The reason is a fortuitous cancellation of errors, which was first noted for the p-benzoquinone example [1] illustrated in the left part of Fig. 8.1 and was subsequently confirmed for a whole class of molecules by Neugebauer and Hess [15]. The latter authors concluded: '[...] it may be legitimate from a pragmatic point of view to compare harmonic frequencies from BP86 calculations directly to experimental wave numbers (keeping in mind that this legitimation is based on a statistically demonstrated error cancellation effect)' [15].

It is due to the high accuracy of the DFT procedures discussed above and to their accessibility in widely distributed software packages [20–22] that the application of DFT has become a standard tool in the description of gas-phase IR spectra. Occasionally, gas-phase results can even provide clues to the decoding of condensed phase IR spectra [7,23]. In general, however, this approach will fail, because gas-phase models cannot account for the complicated and specific effects of the respective condensed-phase environments on the vibrational modes of embedded solute molecules [24,25].

For the description of condensed-phase IR spectra, a realistic modelling of the solvent environment and of its interaction with the solute molecule is crucial. To meet that challenge DFT/MM methods have been developed [25–28], in which the solute is treated by an accurate DFT Hamiltonian and the solvent environment by a computationally efficient molecular mechanics (MM) force field [29]. Despite the well-known inaccuracies of MM force fields [30,31], such a simplified treatment of the environment (MM fragment) may suffice, because the electric field generated by the MM fragment and polarizing the solute molecule (DFT fragment) is imported as an external perturbation into the DFT Hamiltonian of that molecule. Inaccuracies in the description of the perturbation will be reflected in corresponding errors of the solvatochromic shifts. However, because such shifts usually amount only to a few per cent of the total frequency of a given vibrational band, the errors concerning frequency predictions will be relatively small. The right part of Fig. 8.1 illustrates the validity of this hypothesis for one band of one molecule in a specific environment. Clearly, many more sample applications involving perturbations of varying magnitude are required for further confirmation.

In the computation of the polarizing external electric field, a major issue of concern is the adequate treatment of the long-range interactions between the DFT fragment and the MM fragment particularly when considering a molecule embedded in a protein or a polar solvent [25]. In this respect, during the past decade considerable progress has been achieved by the development of efficient multiple scale methods, which either apply Ewald summations [32–34] or combine linearly scaling hierarchical multipole expansions with reaction field corrections [28,35–38]. As a result, corresponding DFT/MM hybrid methods appear to be sufficiently accurate and efficient for computations of IR spectra of molecules in the condensed phase. However, as we shall see below, the mere existence of DFT/MM hybrid methods does not yet guarantee that one can reliably compute such IR spectra. Instead, a whole series of additional technical points has to be adressed in order to achieve that goal.

As of today, only a few applications of hybrid DFT/MM methods to the computation of condensed-phase IR spectra have been presented [2,3,26,28,39–45]. Hybrid methods of substantially lower accuracy, using, e.g., semi-empirical (SE) [46] or *ab initio* Hartree–Fock (HF) [40,47,48] methods instead of DFT, have also been applied for this purpose, despite their well-known deficiencies in the description of intra-molecular force fields. Although the art of computing IR spectra of molecules in the condensed phase has not yet reached the status of a routine application characteristic for computations of gas- phase IR spectra, the quoted studies have provided valuable insights, particularly concerning the accuracy of the various hybrid DFT/MM approaches. However, because of the small number of previous studies and because of sparse comparisons with experimental data, the field is not yet technically settled.

It is the aim of this chapter to sketch the state-of-the-art and to explain the origin of open questions. The central issues will be: by which procedures can one compute IR spectra of molecules in the condensed phase? How accurate are these calculations at the given state-of-the-art, i.e., how do they compare with experiments? What are the main difficulties in these calculations, and how may they be removed by future developments?

In order to tackle these questions, we will first outline the theoretical framework of IR spectra calculations. Subsequently, we will give an overview and a discussion of selected results of recent DFT/MM calculations of IR spectra.

8.2 METHODS TO CALCULATE IR SPECTRA OF MOLECULES IN THE CONDENSED PHASE FROM DFT/MM-MD SIMULATIONS

Our sketch of the theoretical framework for the methods to calculate IR spectra from DFT/MM molecular dynamics (MD) simulations will be guided by an experimental viewpoint. First recall that a typical IR experiment aims to measure the absorption $\alpha_m(\omega)$ of a specific molecule in a given condensed-phase environment. Here, the focus lies on the mid-infrared (MIR) spectral region, which ranges from 600 cm^{-1} to 4 000 cm^{-1}, because this frequency range contains the characteristic intramolecular vibrational bands of the probed molecule. Furthermore, the absorption $\alpha_s(\omega)$ of the environment ('background absorption') is subtracted by difference techniques from the observed absorption $\alpha(\omega)$ and one obtains the spectral changes

$$\tilde{\alpha}_m(\omega) = \alpha(\omega) - \alpha_s(\omega) \tag{1}$$

caused by the given molecule. These changes $\tilde{\alpha}_m(\omega)$ represent the absorption $\alpha_m(\omega)$ of the molecule in the condensed-phase environment, if the total absorption

$$\alpha(\omega) = \alpha_m(\omega) + \alpha_s(\omega) \tag{2}$$

is separable into the absorption $\alpha_m(\omega)$ of the molecule and the absorption $\alpha_s(\omega)$ of its environment.

According to Ref. [43], the separation (2) applies to the MIR absorption of a molecule, if there is no resonance between the MIR vibrational modes of that molecule and

those of its environment, and if the polarization induced by the molecule into the environment is negligible on the timescale of an MIR vibration.

These conditions are fulfilled for a molecule, which is dilutely soluted in a polar solvent, if the electronic polarization of the solvent by the solute is negligible and if the vibrational bands of the solvent molecules are spectrally separated from those of the solute, such that resonant couplings are avoided [43]. Furthermore, they apply to DFT/MM simulations in which a single solute molecule is described by DFT and the solvent molecules by a rigid, non-polarizable MM model [43].

All computational procedures suggested so far for the computation of MIR spectra from DFT/MM-MD simulations rely on the approximations and assumptions sketched above. These procedures are the instantaneous normal mode analysis (INMA) [2,26,43], the Fourier transform of the time correlation function (FTTCF) of the dipole moment [2,43], and the time-resolved, bandpass filtered generalized virial (TF-GV) frequency approach [44]. We will now sketch the merits and deficiencies of each of these methods for the above-described scenario of a molecule dilutely soluted in a polar solvent. In Section 8.2.5, we will discuss to what extent these methods are also applicable to the more general case of a molecule in condensed phase, e.g., to a chromophore in a protein.

8.2.1 FTTCF

The FTTCF approach starts out from the expression

$$\alpha_m(\omega) = \frac{c_0\omega}{6\pi\hbar}\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\omega t}\,\langle \vec{d}\,(t')\vec{d}\,(t'+t)\rangle\,\mathrm{d}t \tag{3}$$

which has been derived by Fermis golden rule for the MIR spectrum of dilutely soluted molecules isotropically distributed in a polar solvent at room temperature ($k_\mathrm{B}T \ll \hbar\omega$) [43,49]. Here, the velocity of light c and the refractive index n are collected into $c_0 = 4\pi^2/(cn)$, and $\vec{d}$ is the dipole moment of a solute molecule. For evaluation, the time correlation function of $\vec{d}\,(t)$ is calculated from a DFT/MM-MD trajectory of one solute molecule and is subsequently Fourier transformed [2,43].

By considering an extended trajectory one has to make sure that the statistics suffice. The spectral resolution $\delta\omega \sim 1/T$ obtained by the FTTCF method is limited by the duration T of the analysed MD trajectory as stated by Nyquists theorem for Fourier transforms [50]. As a consequence, a 30 ps trajectory is required for a 1 cm^{-1} spectral resolution.

In addition, one has to take care that the time step Δt chosen for the integration of the Newtonian equations of motion is sufficiently small. The background for this *caveat* is the fact that a too coarse time step Δt applied to the usual Verlet integrator [51] leads to a systematic overestimate $\Delta\omega = \omega^3\Delta t^2/24$ of frequencies ω calculated by FTTCF [44]. For example, the stretch frequency of a C–H bond, which typically measures about 3 000 cm^{-1}, will be artificially blue-shifted by about $\Delta\omega = 15$ cm^{-1}, if an integration time step of $\Delta t = 0.5$ fs is chosen.

As a last *caveat* we note that the spectral accuracy of the FTTCF method is limited by underestimates of anharmonic shifts. These underestimates are caused by the fact that a classical MD trajectory at room temperature samples only a small part of the anharmonic

potential energy surface as compared with a real molecule, which is excited by an MIR quantum at an energy $\hbar\omega \gg k_B 300$ K [44].

As a key advantage, the FTTCF method includes dynamical effects such as motional narrowing and, therefore, can render exact line widths. The intensities, on the other hand, when calculated by (3) from an MD simulation, may be erroneous because of the following two effects.

First, the thermal equilibrium assumed in the derivation of the FTTCF expression (3) is usually absent in necessarily short DFT/MM-MD simulations. Therefore, IR intensities calculated by FTTCF may be grossly wrong [44]. Fortunately, these errors may be corrected by the introduction of mode-specific temperatures [44]. In addition, the classical description of the nuclear motion in MD simulations introduces systematic errors into the calculated IR intensities [52–54]. For a removal of these errors, the so-called harmonic approximation correction [55] should be used, as has been shown in a series of studies [44,55–58].

8.2.2 INMA

As compared with FTTCF, the INMA approach [2,26,43] additionally assumes that the solute's vibrations are decoupled from its rotational and translational motions in its solvent cage and from the dynamics of the solvent. Here, the vibrational states of the solute are described, as if the solvent molecules were fixed at given positions, which we collect into a vector $\vec{r}_S$. In this 'frozen solvent approximation' (see Ref. [43] for a discussion) the MIR spectrum of a thermally equilibrated and isotropically distributed solute in a polar solvent at room temperature is given by

$$\alpha_m(\omega) = \frac{c_0\omega}{3\hbar} \int \mathrm{d}V\, \rho(\vec{r}_S) \sum_f \left| \left\langle f(\vec{r}_S) \middle| \vec{d} \middle| 0(\vec{r}_S) \right\rangle \right|^2 \delta(\omega - \omega_f(\vec{r}_S)) \tag{4}$$

where $|f\rangle$ are the eigenstates of the molecular vibrational Hamiltonian $H_{\mathrm{vib}}(\vec{r}_S)$ of the solute at a given solvent configuration $\vec{r}_S$ [43]. The associated absorption frequencies $\omega_f = (E_f - E_0)/\hbar$ follow from the difference of the energy E_f of the eigenstate $|f\rangle$ from the energy E_0 of the vibrational ground state $|0\rangle$. The integral in Eq. (4) represents an ensemble average over these configurations as defined by the probability density $\rho(\vec{r}_S)$ and $\mathrm{d}V$ is the volume element in configuration space.

In order to evaluate the INMA Eq. (4), an ensemble $\{\vec{r}_S\}$ of solvent cages is generated by an MD simulation. Each of these cages is subsequently kept fixed and the vibrational spectrum of the embedded solute molecule is calculated from the DFT/MM force field by the following procedure [2,26]: first, the potential energy of the polarized solute is minimized in each cage. Subsequently, the harmonic vibrational frequency ω_j of each vibrational mode j of the solute is calculated from the Hessian. Finally, the derivatives $\partial\vec{d}/\partial a_j$ of the solute's molecular dipole moment $\vec{d}$ with respect to the normal coordinates a_j are calculated. These frequencies and dipole derivatives are different for each solvent cage, because the vibrational Hamiltonian $H_{\mathrm{vib}}(\vec{r}_S)$ contains the interactions of the solute with its respective solvation shell.

Restricting to the fundamentals and applying the harmonic approximation, one can then identify the resulting ensemble $\{\omega_j\}$ of harmonic vibrational frequencies with the ensemble $\{\omega_f\}$ of absorption frequencies, and can calculate the ensemble $\{\langle f|\,\vec{d}\,|0\rangle\}$ of transition matrix elements in linear approximation from the associated dipole derivatives [43]. If the vibrational band of each mode is well separated from the others, then each mode can be characterized by a central frequency $\overline{\omega}_j$ and a line width σ_j, both of which can be estimated by applying second-order statistics to the ensemble $\{\omega_j\}$ of vibrational frequencies. Averaging also the dipole derivatives and assuming a Gaussian line shape, the INMA absorption spectrum (4) becomes

$$\alpha_m(\omega) = \frac{c_0}{6} \sum_j \left\langle \left(\frac{\partial \vec{d}}{\partial a_j} \right)^2 \right\rangle \frac{\exp\left(-\frac{(\omega-\overline{\omega}_j)^2}{2\sigma_j^2} \right)}{\sqrt{2\pi\sigma_j^2}} \tag{5}$$

Using perturbation theory, anharmonic corrections can be additionally applied to the absorption frequencies [15,46,59] and to the transition matrix elements [60].

For large molecules in viscous solvents or in the binding pocket of a protein, the decoupling of the vibrations from the rotational motions of the solute, as assumed by the INMA approach, usually holds very well. However, because INMA completely decouples the solute vibrations from the solvent dynamics, it does not capture dynamical effects like motional narrowing [61–63] or vibrational dephasing [64], which modify the distribution of the absorption frequencies and, thus, the line widths. Therefore, one would expect that line widths of molecules in the condensed phase are largely overestimated by INMA. In contrast to this expectation, INMA line widths have shown good agreement with FTTCF reference line widths [2,44] and with experimental data [42].

A recent study [44] has shown that this good agreement is due to the compensation of the expected INMA overestimate by the following effect: according to the INMA protocol, the molecular energy is minimized in the solvent cages prior to frequency calculations. Owing to these minimizations, the molecule avoids strong interactions with its solvent cages and, therefore, the fluctuations of the frequencies are reduced [44].

Note that this favourable cancellation of errors is absent in other protocols [65–67] related to INMA, which do not prescribe an energy relaxation of the molecule prior to frequency calculation. Apart from increasing the calculated line widths, the omitted energy minimization can lead to imaginary values for some of the vibrational frequencies, which can then no longer be interpreted as absorption frequencies and entail unphysical divergences in the computation of time-correlation functions [65].

8.2.3 TF-GV

The frequency of a harmonically oscillating normal coordinate a derives according to

$$\omega^2 = \langle \dot{a}^2 \rangle / V(a) \tag{6}$$

from the variance $V(a)$ and from the second moment of the time derivative $\dot{a}$. Based on the ‘generalized virial’ expression (6), an alternative method, called TF-GV, has recently

been developed to calculate the IR spectra of molecules in polar solvents from DFT/MM trajectories [44]. Here, the MD trajectory is projected onto (approximate) normal modes of the solute molecule yielding trajectories $a_j(t)$ of the normal coordinates for each mode j.

Unfortunately, one cannot evaluate the frequencies ω_j directly from Eq. (6), because slow fluctuations of the equilibrium geometry, of the intramolecular force constants, and of the directions of the normal modes as well as diagonal and off-diagonal anharmonicities of the molecular potential, which are sampled by the classical MD trajectory, introduce errors into the generalized virial frequencies (6) [43]. However, because all these sources of errors give rise to distinct spectral features in the Fourier transform $\hat{a}_j(\omega_j)$ of $a_j(t)$, the application of a band pass filter to $a_j(t)$ can remove all those sources of errors, whose spectral signatures are clearly separated from the band marking the fundamental vibration. Applying Eq. (6) to the band pass filtered trajectories $a_j^*(t)$ then yields accurate vibrational frequencies [44].

In addition, the TF-GV approach even allows the computation of time-resolved IR spectra. By computing the expectation values in (6) for small gliding time windows of width τ, time-resolved vibrational frequencies $\omega_j(t|\tau)$ are obtained. Gaussian convolution kernels are particularly useful for such gliding temporal averages, because they yield highly accurate frequencies even at a time resolution τ, which is as small as the oscillation time $\tau_j = 2\pi/\omega_j$ of the monitored vibration. Correspondingly, the grey line in Fig. 8.2 shows the trajectory of the strongly fluctuating carbonyl stretching frequency of a formaldehyde (H_2CO) molecule in aqueous solution at a time resolution of $\tau = 20$ fs, as calculated from a 35 ps DFT/MM simulation [44].

Now the question arises, how one can calculate time-resolved IR spectra from such trajectories, keeping in mind that the IR spectra may be affected by motional narrowing, which has not yet been accounted for by TF-GV. Recall that the parameter $\alpha = \delta\omega_f\tau_f$, which depends on the size $\delta\omega_f$ and the time scale τ_f of the frequency fluctuations, indicates whether these fluctuations are within ($\alpha < 1$) or outside the motional narrowing limit ($\alpha > 1$) thus generating inhomogeneous line broadenings [61].

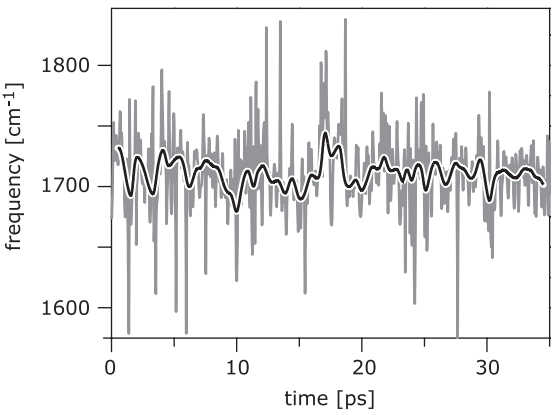


Fig. 8.2. Trajectories $\omega_j(t/\tau)$ of the carbonyl stretching frequency of formaldehyde in aqueous solution, as calculated by TF-GV at time resolutions of τ = 20 fs (grey line) and τ = 240 fs (black) from a 35 ps DFT/MM simulation. Data taken from Ref. [44].

An analysis of the TF-GV time series $\omega_j(t/\tau = 20$ fs) shown in Fig. 8.2 has demonstrated that large parts of the frequency fluctuations are induced by fast vibrational motions of the molecule in its solvent shell. Slower fluctuations are caused by conformational transitions within that shell. The corresponding parameters $\alpha = \delta\omega_f\tau_f$ indicate that only the slower fluctuations contribute to the inhomogeneous broadenings [44].

If the averaging time scale τ for the computation of the TF-GV frequencies $\omega_j(t|\tau)$ is now chosen in such a way that the fastest, motionally narrowed fluctuations are averaged out, then only those fluctuations remain, which are in the inhomogeneous limit ($\alpha > 1$) and, in principle, are accessible to time-resolved spectroscopy. For the sample case of the formaldehyde molecule in aqueous solution, the black line in Fig. 8.2 shows the correspondingly smoothened TF-GV trajectory of the carbonyl stretching frequency, which represents the time-resolved single-molecule IR spectrum of that mode.

The absorption spectrum can be constructed according to

$$\alpha_m(\omega) = \frac{c_0}{6} \sum_j \left(\frac{\partial \vec{d}}{\partial a_j} \right)^2 p_j(\omega) \tag{7}$$

Here, the dipole derivatives follow from the trajectories of the dipole moment and of the normal coordinates [43] and the distribution $p_j(\omega)$ of absorption frequencies can be modelled by a histogram $p_j(\omega_j(t|\tau))$ of the TF-GV frequencies. If a sufficiently large DFT/MM trajectory has been sampled, the line shape of an IR band can also be computed [44] from the autocorrelation function of the TF-GV frequency $\omega_j(t|\tau)$ instead from (7).

8.2.4 Summary of Methods

As we have shown above, each of the various methods to compute MIR spectra from DFT/MM trajectories has its specific merits: in the INMA procedure, a computationally cheap MM-MD simulation may be used to generate the required ensemble of solvent cages (cf. Refs [25,42,45]). Furthermore, INMA enables an analysis of anharmonic effects on the MIR spectra (cf. Ref. [46]). The FTTCF approach yields excellent line shapes and is best applicable to study larger systems with sloppy, strongly anharmonic modes (cf. Refs [41,54]). The TF-GV method allows us to focus on the origins, the sizes, and the time scales of the frequency fluctuations (cf. Ref. [44]).

Figure 8.3 displays the MIR spectrum of a formaldehyde molecule in aqueous solution as calculated by INMA, TF-GV, and FTTCF from a DFT/MM trajectory. Apart from a slight underestimate of the maximum absorption by INMA, which is caused by a slight overestimate of the corresponding line widths by this approach, the figure demonstrates that all three methods give consistent results, if all the corrections enumerated above are applied.

8.2.5 IR spectra of molecules in a protein environment

Having defined the assumptions underlying the various computational approaches toward the IR spectra of molecules in polar solvents, we will now check whether these methods can be applied to the more complicated case of a molecule in a protein environment.

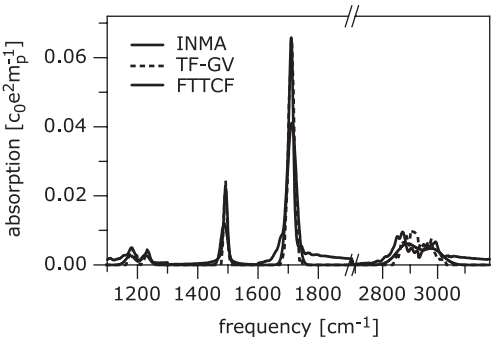


Fig. 8.3. MIR spectrum of a formaldehyde molecule in aqueous solution as obtained by INMA, by TF-GV, and by FTTCF from a DFT/MM simulation. Data taken from Ref. [44].

In the MIR spectral region up to 1 700 cm^{-1}, a protein has a high density of vibrational states. The region contains mainly the bending and stretching modes of the covalent bonds connecting the heavy atoms and the hydrogen bending modes. Near 3 000 cm^{-1}, one finds the hydrogen stretches. In both frequency regions, the resonant couplings between these protein modes with the vibrational modes of an embedded solute molecule can no longer be excluded.

If we consider, for instance, a single peptide unit of the protein backbone as such a solute molecule, the necessity of including resonant couplings becomes immediately obvious. Here, the normal modes of all other peptide units in the backbone will have similar frequencies and, therefore, will mix with those of the given unit through transition dipole coupling [68,69]. For β-sheet structures, for example, this coupling leads to the splitting of the amide-I band near 1 650 cm^{-1} into two bands centred at 1 630 cm and at 1 690 cm^{-1}, respectively [8]. In addition, the amide-I band couples to the bending mode of surrounding water molecules. For these reasons, a separation (2) of the total absorption $\alpha(\omega)$ into contributions of local modes is impossible.

In order to calculate the MIR spectrum of a molecule in a protein environment, one may use two approaches. The first and most straightforward approach is to *neglect* the vibrational couplings discussed above. This may be legitimate, if the dipole–dipole coupling of the resonant modes is weak, e.g., due to small dipole derivatives of the respective modes and/or due to a large spatial separation. If resonant couplings cannot be neglected, as is the case for large vibrational dipoles in close proximity, one must choose a different approach. For instance, one may calculate the vibrational frequencies of each fragment individually and account for the couplings afterwards. Various methods for such calculations have been suggested [8,68–71].

For example, the transition dipole coupling (TDC) method [68,69] is based on the assumption, that the peptide backbone can be treated as a set of oscillators with identical force constants, thus neglecting polarization-induced variations of the force constants. Then the amide oscillators are coupled via the dipole–dipole interaction while neglecting (i) interactions mediated by higher-order multipoles or by covalent bonds, and (ii) couplings with the solvent oscillators. Owing to these crude approximations, TDC fails to be

sufficiently reliable for the prediction and structural decoding of protein and peptide amide bands. Only recently, attempts were made to determine the force constants from DFT [72,73] or DFT/MM calculations [28,74], as well as to include through-bond interactions of neighbouring amide groups into TDC [71].

8.3 RECENT RESULTS AND ACCURACY ISSUES

Having outlined the theoretical framework for calculations of MIR spectra from DFT/MM simulations, we will now give an overview of selected results. We will use this overview to focus on specific aspects in the computational set-up, i.e., the partitioning of the system into a DFT fragment and an MM fragment, the description of the MM fragment, and the interaction between the two fragments. This will allow us to show how different computational set-ups and modelling procedures affect the accuracy of the calculated vibrational spectra and how the remaining difficulties could be removed by future developments.

8.3.1 DFT/MM partitioning

In the set-up of a DFT/MM calculation one first has to address the question as to how one should partition the system into DFT and MM fragments. For a molecule in solution or a chromophore in a protein binding pocket, the natural choice seems to be to describe the solute or chromophore by DFT and the solvent or protein by MM. Now one might hope to increase the accuracy of the calculations by including, e.g., also neighbouring solvent molecules or amino acid sidechains into the DFT fragment. However, in the few studies in which this approach was chosen [25,75], an enlargement of the DFT fragment apparently deteriorated the results.

Speranskiy and Kurnikova calculated the asymmetric carboxyl stretch frequencies of glutamate in aqueous solution [25]. In two different INMA calculations, the DFT fragment contained either solely the glutamate (DFT/MM) or the glutamate and six water molecules attached to each of its two carboxyl groups (DFT$^+$/MM). Figure 8.4 shows the results and compares with experimental data. According to the figure, at equal 12 Å electrostatics cut-offs the enlargement of the DFT fragment by solvent molecules is unfavourable.

A DFT/MM study ongoing in our group [75] and aiming to compute the IR spectrum of the retinal chromophore in bacteriorhodopsin yielded a similar result. Here, the inclusion of a water molecule attached to the protonated Schiff base group of the chromophore into the DFT fragment caused a red-shift of the C=N stretch frequency that seemed to be incompatible with available spectroscopic evidence. A close look at the data revealed a considerable charge transfer from the water molecule to the Schiff base group as the likely cause for the noted shift.

Such a charge transfer results from the applied Born–Oppenheimer (BO) approximation, within which the chromophore and the water molecule are treated as a supermolecule. However, at room temperature in the condensed phase, the BO approximation may be inapplicable to molecular clusters for the following reason: if the charge transfer coupling between the wavefunctions of neighbouring and saturated molecules is weak, then the noise-induced dephasing can become rapid and effectively localize the wavefunctions

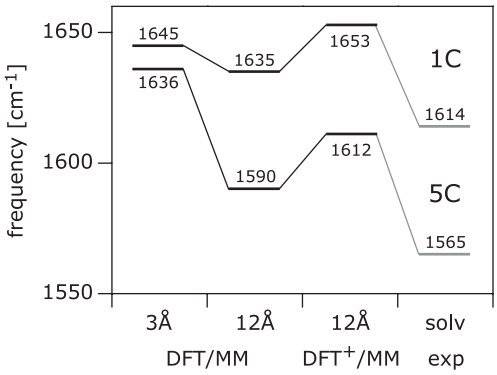


Fig. 8.4. Comparison of experimental data with DFT/MM asymmetric carboxyl stretch frequencies of glutamate in D_2O calculated at different electrostatics cut-offs (3Å, 12Å). In the calculations, the DFT fragment contained the glutamate only (DFT/MM) or the glutamate and six adjacent water molecules (DFT/MM), respectively. Data taken from Ref. [25].

at the individual molecules. If this should be the case, the requirement of localized wavefunctions should be imposed onto joint DFT descriptions of molecular clusters through additional constraints much as is possible, e.g., in the program package Quickstep [76]. By construction, such a localization is guaranteed in DFT/MM simulations, in which only one single molecule is described by DFT.

8.3.2 Modelling of the environment

In the Introduction we stated that the IR spectrum of a molecule in the condensed phase is affected by the charge distribution of its environment and that a vibrational analysis requires a highly accurate force field. Therefore, not only the quantum mechanical description of the solute molecule has to be carefully selected, but also the MM modelling of the environment. Recent studies provide first estimates of how accurate the MM modelling needs to be for precise DFT/MM predictions of condensed phase IR spectra.

Consider first the IR spectra of the two quinone chromophores Q_A and Q_B in the photosynthetic reaction centre (RC) of *Rb. sphaeroides*, investigated by Nonella *et al.* [3]. In this DFT/MM study the BP functional was chosen for the description of the quinones, because, as shown in the left graph of Fig. 8.1, it yields the IR spectrum of a quinone in the gas phase within a few wavenumbers [1]. A three-dimensional crystal structure [77] (PDB code 1PCR) of the protein/chromophore complex was selected to calculate the IR spectra of the quinones in the RC. Particular attention was paid to the frequency range between 1 550 cm^{-1} and 1 700 cm^{-1}, because it covers the spectroscopically well-studied C=O and C=C stretching modes of the quinones [6].

According to Fig. 8.1, the calculated vibrational frequencies of Q_A agree within an RMSD of only 5 cm^{-1} with experimental values. In particular, the unusually low frequency of the C4=O carbonyl stretching mode at 1 601 cm^{-1} is nearly perfectly matched. To identify causes for the low frequency of the C4=O stretch, Nonella *et al*. conducted a second computer experiment, in which they solely changed the charge q of the nearby

iron cofactor from $q = 2e$ to zero. As demonstrated by the column Fe0 in Fig. 8.1, uncharging the iron leads to a 25 cm^{-1} blue-shift of the C4=O frequency and causes sizeable shifts also at the other modes. Thus, the strong electric field generated by the iron cofactor and polarizing Q_A largely explains the low frequency of the C4 = O stretch.

These results indicate that DFT/MM can predict the vibrational spectrum of a molecule in a complex environment within a few wavenumbers, which is comparable to the accuracy of DFT concerning gas-phase spectra. Furthermore, they provide an instructive example for the exploration of certain structural details shaping vibrational spectra of molecules in a complex environment. But by revealing the high sensitivity of vibrational spectra toward changes of local electric fields they also underline the necessity of a careful modelling of the MM environment.

Conversely, because DFT/MM can accurately predict IR spectra and because these spectra probe the structure of the environment, such calculations can even provide evidence for or against the correctness of a structural model. For instance, the good match of the DFT/MM result with spectroscopic data on Q_A as well as other evidences have lead Nonella *et al*. to conclude that the 1PCR crystal structure is correct in the surroundings of Q_A. In contrast, for the IR spectrum of Q_B these authors did not find a good match of DFT/MM results with spectroscopic data [3]. In this case, the absence of a good match was taken as additional support for the concept that the 1PCR structural X-ray model is erroneous near Q_B.

As a result, Nonella *et al*. found a strong correlation between the accuracy of calculated DFT/MM IR spectra and the correctness of given structural models of protein chromophore complexes. In this context, recent reviews [30,31] provide important arguments. It is argued that the neglect of polarizability in standard MM force fields (e.g., CHARMM [78], AMBER [79], or GROMACS [80]) can cause structural artefacts in MD simulations of proteins. By the above results such artefacts will induce errors into IR spectra calculations.

In addition to this hypothesis of an *indirect* structural influence of the usual MM neglect of the polarizability on calculated IR spectra, there is also a *direct* effect of this neglect. This effect was identified in a DFT/MM study [42] of the vibrational spectra of phosphate ions ($H_2PO_4^-$ and $H_2PO_4^{2}$) in aqueous solution, in which the water molecules were modelled by the non-polarizable TIP3P force field [81]. The DFT/MM solvent shifts of vibrational bands turned out to cover only about half the experimental shifts [42]. As can be concluded from the data collected in Fig. 8.5, the calculated frequencies deviate from the experimental data by an RMSD of about 45 cm^{-1}.

The authors explained the DFT/MM failure to reproduce the complete solvent shifts by the following arguments: the parameters of the TIP3P water model have been optimized to reproduce observed properties of bulk water (see, e.g., Ref. [81]). Keeping these parameters in a MM force field at fixed values corresponds to a mean field approach. Such an approach neglects the fluctuations of the dipole moments, force constants and equilibrium structures of the water molecules, which are caused by their fluctuating environments. It is invalid for water molecules close to the singly or doubly charged phosphate ions. The fixed TIP3P charges cannot account for the additional dipole moment, which is induced into these water molecules by the ions. Correspondingly, the polarization of the phosphate ions by the nearby water dipoles is too small and the solvent shifts of the phosphate modes are

underestimated by the DFT/MM calculations [42]. The authors argued that a polarizable force field for the MM description of the surrounding water environment should strongly shift the calculated toward the experimental spectra of the solvated ions [42]. Thus, for ionic molecules, DFT/MM descriptions of vibrational spectra seem to require polarizable force fields, if sufficient accuracy is desired.

This conclusion is supported by a follow-up study [45], which addresses the vibrational spectrum of GTP in the binding pocket of the RAS p. 21 protein. Also here, a DFT/MM approach comprising a non-polarizable MM force field for the protein and the solvent was applied, despite the ultrapolar structure of the GTP binding site, which contains a four-fold charged GTP, and as counter-ions a magnesium ion and a protonated lysine sidechain. According to Fig. 8.5, the frequencies calculated for GTP in Ras p. 21 deviate from the spectroscopic data by about as much as was observed for the phosphate ions in TIP4 water.

Referring again to the reviews of Ponder and Case [30] and of Tavan *et al*. [31], one must additionally suspect that not only the polarizations of the environments of the phosphate and GTP ions, respectively, were underestimated in the quoted studies, but also that the ensembles of solvent shell and protein *structures* sampled in the corresponding MM-MD simulations are contaminated with errors. These potentially erroneous environmental structures may have induced further artefacts into the calculated vibrational spectra of the phosphate and GTP ions.

Summarizing the above studies, we conclude that the DFT force field of the solute is sufficiently accurate that every deficiency in the MM modelling of its environment is relentlessly resolved. In particular, for ionic molecules the use of polarizable force fields appears to be inevitable, if one wants to obtain correct structures in MD simulations and aim at a correct description of the electrostatics shaping the IR spectra. We would like to

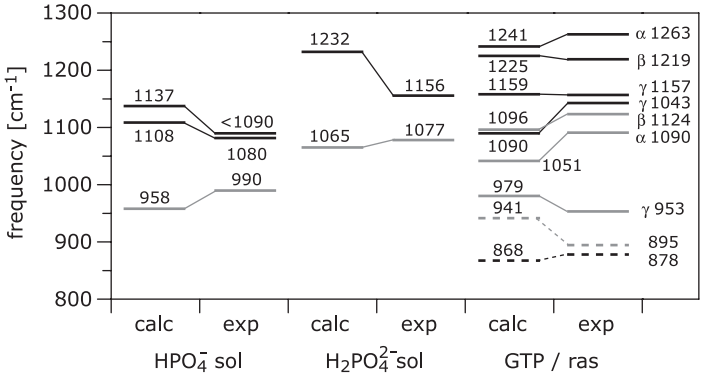


Fig. 8.5. Comparison of the calculated and experimental frequencies of the symmetric (grey) and of the asymmetric (black) PO (solid) and OPO (dashed) stretching modes of the phosphate ions HPO_4^{2-} and $H_2PO_4^-$ in aqueous solution and of GTP in Ras p. 21. The root mean square deviations between calculated and experimental frequencies of the displayed modes are 37 cm^{-1}, 54 cm^{-1}, and 32 cm^{-1}, respectively. The Greek letters α, β, γenumerate the three phosphate groups of the GTP molecule. For better visibility, the POH stretching and bending modes of the phosphate ions are omitted. Data taken from Refs [42,45].

finally note that the phosphate/water system quoted above represents an interesting test case for the evaluation of polarizable force fields.

8.3.3 Non-resonant interactions between DFT- and MM-fragment

Having discussed the accuracy requirements that have to be imposed on the methods dealing with DFT and MM fragments, respectively, we now turn to the interactions between the two fragments, which are non-resonantly mediated by chemical bonds and long-range electrostatic fields.

Watson *et al*. [28] calculated the amide bands of selected peptide units in two different secondary structure motifs and in regions of different solvent accessibility of a soluted protein. In their calculations, they removed the peptide unit of interest together with the two neighbouring peptide groups connecting that unit to the remaining protein and replaced it by an N-Methyl-Acetamide (NMA, $CH_3CONHCH_3$) model. They applied the INMA protocol to calculate the vibrational spectrum of NMA (DFT fragment) in the environment constituted by the remaining protein and the solvent (MM fragment).

As shown in Fig. 8.6, the calculation underestimates the spectroscopic 63 cm^{-1} blue-shift of the amide-II frequency by about 30 cm^{-1}. In agreement with the experimental findings, the calculations showed no significant dependence of the amide-II frequency on the secondary structure. The calculated amide-I band of a peptide unit in an α-helix is calculated to be about 80 cm^{-1} to the red of the NMA vacuum result, whereas the corresponding experimental shift is only about 65 cm^{-1}. For the β-sheet case, the calculation predicts an amide-I frequency quite close to the high-frequency peak of the corresponding IR band. However, because the resonant couplings with neighbouring peptide units are neglected in the applied computational approach, this agreement is accidental, because these couplings will cause a substantial splitting of that band.

It may well be that the neglect of the covalent linkage between the model peptide unit and the protein has substantially contributed to the noted deviations. Here, the use of link atom methods [26] may furnish improvements. Nevertheless it appears that the simple

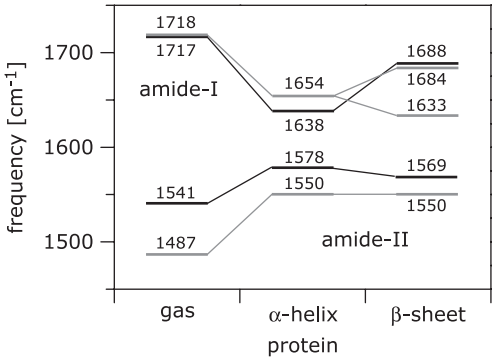


Fig. 8.6. Calculated (black) amide-I and amide-II frequencies of NMA in the gas phase and at the location of a peptide unit in an α-helix and a β-sheet of a protein compared with corresponding experimental data (grey). Data taken from Refs [8,28,82].

model of Watson *et al*. [28] qualitatively accounts for the polarization effects induced by the protein environment on the amide modes.

Watson *et al*. additionally contributed an interesting insight into the effects of long-range electrostatics by conducting their calculations for a protein in aqueous solution and a protein in vacuum. As expected, the amide bands of solvent-exposed peptide units were strongly affected by the neglect of the solvent, as witnessed by shifts of up to 20 cm^{-1}. However, also buried peptide units turned out to be sensitive to solvation and showed shifts of up to 15 cm^{-1} [28]. This strong effect shows that the reaction field generated by the aqueous environment can also significantly change the vibrational spectra of buried groups.

A recent study by Speranskiy and Kurnikova leads us to a similar conclusion [25]. They calculated the vibrational frequencies of the carboxyl modes of glutamate in D_2O with the INMA procedure. Figure 8.4 shows that the vibrational frequencies are overestimated by up to 70 cm^{-1}, if the electrostatic interaction of the glutamate with water molecules is neglected beyond the first hydration shell. Also, by adding a few more solvent molecules upon applying the common 12 Å electrostatics cut-off does not recover the complete solvent shift, pointing towards the necessity of also adequately including the more long-range electrostatic interaction (cf. the Introduction).

The results reviewed above lead to a slight extension of the conclusion stated at the end of the previous section: the DFT description of the vibrational spectra of solute molecules is sufficiently accurate that any inaccuracies in the modelling of the interactions of the solute molecule with its environment are monitored without mercy.

8.4 CONCLUSION

In our attempt to describe the state of the art concerning the computation of IR spectra of molecules in the condensed phase, we first sketched the methods INMA, FTTCF and TF-GV, which allow us to calculate IR spectra from DFT/MM simulations, discussing their basic assumptions, their merits and their deficiencies.

Subsequently, we reviewed selected applications to provide an insight into the intricacies connected with such calculations. By analysing the employed computational set-up and by discussing the achieved accuracy, we have identified four crucial points: (i) the polarizability of the solvent environment has to be included into its MM description, particularly when considering ionic solutes; (ii) including the polarizability of the environment may also become important, because one has to generate realistic environment structures of the DFT fragment in MD simulations; (iii) long-range electrostatic interactions have to be adequately modelled; and (iv) resonant couplings have to be accounted for, if present.

Methods to account for resonant couplings (iv) do already exist [8,68,69,71] and have been applied to frequencies obtained by DFT/MM methods [74]. Here, however, the results are still not satisfactory. Thus, these coupling methods need to be improved, in particular when aiming at the important issues of a prediction and structural decoding of protein amide bands. The requirements (iii) of accurate electrostatics [29,35–38,83] are wellknown and are already met in state-of-the-art technologies, although they are not yet commonly applied.

Polarizable force fields, as required by (i) and (ii), are currently being developed (see, e.g., Refs [30,84] for recent reviews), but have not yet been fully established. In particular, to our knowledge, polarizable force fields have not yet been used in the context of the calculation of IR spectra from DFT/MM simulations. Because a reliable and accurate polarizable force field is needed not only for high accuracy in such calculations, but also for a reliable prediction of structures and dynamics of proteins [30,31], its development will be one of the most important challenges for computational molecular biophysics in the near future.

8.5 ACKNOWLEDGEMENT

Support by the Deutsche Forschungsgemeinschaft (SFB533/C3) and from the Volkswagenstiftung (project I/73224) is gratefully acknowledged.

8.6 REFERENCES

1 M. Nonella and P. Tavan, Chem. Phys., 199 (1995) 19–32.
2 M. Nonella, G. Mathias and P. Tavan, J. Phys. Chem. A, 107 (2003) 8638–8647.
3 M. Nonella, G. Mathias, M. Eichinger and P. Tavan, J. Phys. Chem. B, 107 (2003) 316–322.
4 F. Siebert, Meth. Enzymol., 246 (1995) 501–526.
5 A. Barth and C. Zscherp, FEBS Lett., 477 (2000) 151–156.
6 J. Breton, J.R. Burie, C. Berthomieu, G. Berger and E. Nabedryk, Biochemistry, 33 (1994) 4953–4965.
7 R. Vogel, F. Siebert, G. Mathias, P. Tavan, G. Fan and M. Sheves, Biochemistry, 42 (2003) 9863–9874.
8 C. Allin, M.R. Ahmadian, A. Wittinghofer and K. Gerwert, Proc. Natl. Acad. Sci. USA, 98 (2001) 7754–7759.
9 D.H. Murgida, E. Schleicher, A. Bacher, G. Richter and P. Hildebrandt, J. Ram. Spec., 32 (2001) 551–556.
10 G. Wille, M. Ritter, R. Friedemann, W. Mäntele and G. Hübner, Biochemistry, 42 (2003) 14814–14821.
11 J.P.M. Schelvis, M. Ramsey, O. Sokolova, C. Tavares, C. Cecala, K. Connell, S. Wagner and Y.M. Gindt, J. Phys. Chem. B, 107 (2003) 12352–12362.
12 A. Barth and C. Zscherp, What vibrations tell us about proteins, Quarterly Reviews of Biophysics, 35 (2002) 369–430.
13 K. Hauser, M. Engelhard, N. Friedman, M. Sheves and F. Siebert, J. Phys. Chem. A, 106 (2002) 3553–3559.
14 Z. Xuefeng, S. J. Mole and L. Ruifeng, Vibrational Spectroscopy, 12 (1996) 73–79.
15 J. Neugebauer and B.A. Hess, J. Chem. Phys., 98 (7) (1993) 5648–5652.
16 A.D. Becke, J. Chem. Phys., 98 (7) (1993) 5648–5452.
17 P.J. Stephens, F.J. Devlin, C.F. Chabalowski and M.J. Frisch, J. Phys. Chem., 98 (1994) 11623–11627.
18. A.D. Becke, Phys. Rev. A, 38 (1988) 3098–3100.
19 J.P. Perdew, Phys. Rev. B, 33 (1986) 8822–8824.
20 M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A. Montgomery Jr., R.E. Stratmann, J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, J.L. Andres, C. Gonzalez, M. Head-Gordon, E.S. Replogle and J.A. Pople, GAUSSIAN 98 Revision A.5, Gaussian Inc, Pittsburgh, PA, (1999).

21 R. Ahlrichs, M. Bär, M. Häser, H. Horn and C. Kölmel, Chem. Phys. Lett., 162 (1989) 165.
22 M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, S. Koseki, N. Matsunga, K.A. Nguyen, S. Su, T.L. Windus, M. Dupuis and J.A. Montgomery Jr., J. Comput. Chem., 14 (1993) 1347–1363.
23 C.B. Martin, M.-L. Tsao, C.M. Hadad and M.S. Platz, J. Am. Chem. Soc., 124 (2002) 7226–7234.
24 M. Großjean, P. Tavan and K. Schulten, J. Phys. Chem., 94 (1990) 8059–8069.
25 K. Speranskiy and M. Kurnikova, J. Chem. Phys., 121 (2004) 1516–1524.
26 M. Eichinger, P. Tavan, J. Hutter and M. Parrinello, J. Chem. Phys., 110 (1999) 10452–10467.
27 A. Laio, J.V. de Vondele and U.Rothlisberger, J. Chem. Phys., 116 (2002) 6941–6947.
28 M. Watson and J.D. Hirst, J. Phys. Chem. A, 107 (2003) 6843–6849.
29 G. Mathias, B. Egwolf, M. Nonella and P. Tavan, J. Chem. Phys., 118 (2003) 10847–10860.
30 W. Ponder and D.A. Case, Advances in Protein Chemistry, 66 (2003) 27–85.
31 P. Tavan, H. Carstens and G. Mathias, Molecular dynamics simulations of proteins and peptides: problems, achievements, and perspectives, in J. Buchner and T. Kiefhaber (Eds.), Protein folding handbook, Wiley-VCH, Weinheim, 2005, pp. 1170–1195.
32 T. Darden, D. York and L. Pedersen, J. Chem. Phys., 98 (1993) 10089–10092.
33 B.A. Luty, I.G. Tironi and W.F. van Gunsteren, J. Chem. Phys., 103 (1995) 3014–3021.
34 U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee and L.G. Pedersen, J. Chem. Phys., 103 (1995) 8577–8593.
35 C. Niedermeier and P. Tavan, J. Chem. Phys., 101 (1994) 734–748.
36 C. Niedermeier and P. Tavan, Molecular Simulation, 17 (1996) 57–66.
37 M. Eichinger, H. Grubmüller, H. Heller and P. Tavan, J. Comp. Chem., 18 (1997) 1729–1749.
38 G. Mathias and P. Tavan, J. Chem. Phys., 120 (2004) 4393–4403.
39 C. Rovira, B. Schulze, M. Eichinger, J.D. Evanseck and M. Parrinello, Biophys. J., 81 (2001) 435–445.
40 R.A. Wheeler, H. Dong and S.E. Boesch, Chem. Phys., Chem., 4 (2003) 382–384.
41 R. Rousseau, V. Kleinschmidt, U. W.Schmitt and D. Marx, Angew. Chem. Int. Ed., 43 (2004) 4804–4807.
42 M. Klähn, G. Mathias, C. Kötting, M. Nonella, J. Schlitter, K. Gerwert and P. Tavan, J. Phys. Chem. A, 108 (2004) 6186–6194.
43 M. Schmitz and P. Tavan, J. Chem. Phys., 121 (2004) 12233–12246.
44 M. Schmitz and P. Tavan, J. Chem. Phys., 121 (2004) 12247–12258.
45 M. Klähn, J. Schlitter and K. Gerwert, Biophys. J., 88, 3829–3844.
46 M. Garcia-Viloca, K. Nam, C. bal Alhambra and J. Gao, J. Phys., Chem. B, 108 (2004) 13501–13512.
47 Q. Cui and M. Karplus, J. Chem. Phys., 112 (2000) 1133–1149.
48 S. Hayashi and I. Ohmine, J. Phys. Chem. B, 104 (2000) 10678–10691.
49 D.A. McQuarrie, Statistical mechanics, Harper and Row, New York, 1976.
50 W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, Numerical Recipes in C, 2nd Edition, Cambridge University Press, Cambridge, 1992.
51 L. Verlet, Phys. Rev. Lett., 159 (1967) 98–103.
52 P.A. Egelstaff, Adv. Phys., 11 (1962) 203–232.
53 J. Borysow, M. Moraldi and L. Frommhold, J. Mol. Phys., 56 (1985) 913–922.
54 P.L. Silvestrelli, M. Bernasconi and M. Parrinello, Chem. Phys. Lett., 277 (1997) 478–482.
55 R. Ramirez, T. Lopez-Ciudad, P. Kumar and D. Marx, J. Chem. Phys., 121 (2004) 3973–3983.
56 H. Ahlborn, X. Ji, B. Space and P.B. Moore, J. Chem. Phys., 111 (1999) 10622–10632.
57 H. Ahlborn, B. Space and P.B. Moore, J. Chem. Phys., 112 (2000) 8083–8088.
58 M.P. Gaigeot and M. Sprik, J. Phys. Chem. B., 107 (2003) 10344–10358.
59 H.H. Nielsen, Rev. Mod. Phys., 23 (1951) 90–136.
60 A. Willets, N.C. Handy, W.H. Green Jr. and D. Jayatilaka, J. Phys., Chem., 94 (1990) 5608–5616.
61 R. Kubo, Adv. Chem. Phys., 15 (1969) 101–127.
62 S. Mukamel, Principles of nonlinear optical spectroscopy, Oxford University Press, Oxford, 1995.
63 R. Kubo, M. Toda and N. Hashitsume, Statistical physics II – nonequilibrium statistical mechanics, Springer, Heidelberg, 1985.
64 D.W. Oxtoby, Adv. Chem. Phys., 40 (1979) 2–48.
65 R.M. Stratt, Acc. Chem. Res., 28 (1995) 201–207.
66 T. Keyes, J. Chem. Phys., 101 (1997) 2921–2930.

67 T. Kalbfleisch and T. Keyes, J. Chem. Phys., 108 (1998) 7375–7383.
68 S. Krimm and Y. Abe, Proc. Natl. Acad. Sci. USA, 69 (1972) 2788–2792.
69 H. Torii and M. Tasumi, J. Chem. Phys., 97 (1992) 86–91.
70 H. Torii and M. Tasumi, J. Chem. Phys., 97 (1992) 92–98.
71 J.W. Brauner, C. Dugan and R. Mendelsohn, J. Am. Chem. Soc., 122 (2000) 677–683.
72 P. Bour and T.A. Keiderling, J. Phys. Chem. B, 109 (2005) 5348–5357.
73 H. Lee, S.-S. Kim, J.H. Choi and M. Cho, Theoretical study of internal field effects on peptide amide i modes, J. Phys. Chem. B, 109 (2005) 5331–5340.
74 T.M. Watson and J.D. Hirst, Phys. Chem. Chem. Phys., 6 (2004) 998–1005.
75 G. Babitska and P. Tavan, unpublished.
76 J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing and J. Hutter, Comput. Chem. Comm., 167 (2005) 103–128.
77 U. Ermler, G. Fritzsch, S.K. Buchanan and H. Michel, Structure, 2 (1994) 925–936.
78 B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan and M. Karplus, J. Comput. Chem., 4 (1983) 187–217.
79 D. Pearlman, D. Case, J. Caldwell, W. Ross, T. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel and P. Kollman, Comp. Phys. Comm., 91 (1995) 1–41.
80 E. Lindahl, B. Hess and D. van der Spoel, J. Mol. Mod., 7 (2001) 306–317.
81 W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey and M.L. Klein, J. Chem. Phys., 79 (1983) 926–935.
82 T. Miyazawa, T. Shimanouchi and S.-I. Mizushima, J. Chem. Phys., 24 (1956) 408–418.
83 K. Nam, J. Gao and D.M. York, J. Chem. Theory Comput., 1 (2005) 2–13.
84 B. Guillot, J. Mol. Liq., 101 (1–3) (2002) 219–260.

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 9

High throughput in-silico screening of large ligand databases for rational drug design

H. Merlitz, B. Fischer and W. Wenzel

Forschungszentrum Karlsruhe GmbH, Institut für Nanotechnologie, Postfach 3640, D-76021 Karlsruhe, Germany

Abstract

We review recent improvements in the FlexScreen docking approach which permit the investigation of large databases and receptor flexibility in high-throughput in-silico screening. Using a database of 10 000 random compounds and known inhibitors for two popular benchmark systems (DHFR and TK), it is shown that the 'classical' approach, i.e., selecting a specific target conformation, which is kept rigid during the database screen, leads to rankings of high specificity but low diversity. We then demonstrate that the introduction of moderate target flexibility into the simulation procedure increases both the diversity of the screen and the overall docking performance of the method.

9.1 DRUG ACTIVITY AND KEY-LOCK PRINCIPLE

In 1941, the drug Methotrexate (MTX) was introduced as a new treatment against leukemia. At that time it was already known that MTX was acting as an antagonist of the vitamin dihydrofolate (DHF), an essential constituent of cell metabolism [1]. It is interesting to note the close similarity of both molecules (see Fig. 9.1, MTX). In MTX, the carbonyl (oxygen) on the pteridine ring is replaced with an NH_2 group and the bridge features an extra methylgroup; otherwise both molecules are identical. At that time it was not foreseeable that the underlying mechanisms which made MTX a successful antimetabolite drug would later serve as a paradigm in rational drug design.

As we know today, there exists an enzymatic process in which a proton is attached to the pteridine ring of the substrate DHF, which is thereby reduced to tetrahydrofolate (THF). The corresponding enzyme is dihydrofolate reductase (DHFR), a protein with two active docking sites. To perform the proton transfer, DHF is docking into one of these pockets, whereas in the second pocket another substrate (NADPH) is waiting to transfer

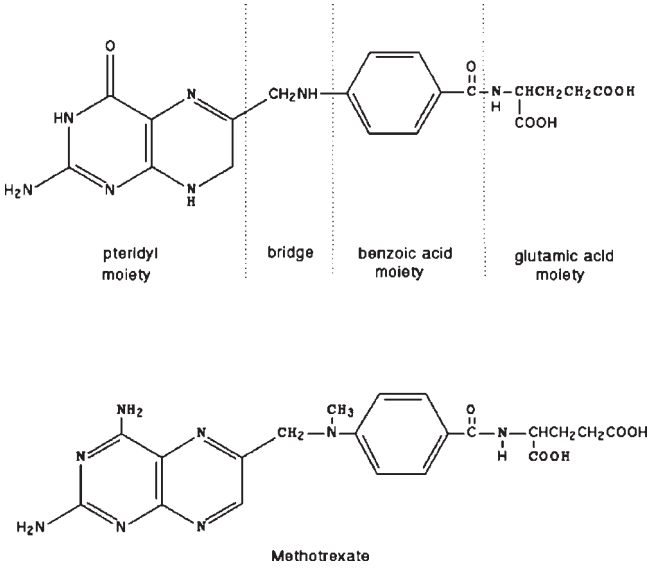


Fig. 9.1. The substrate DHF (upper plot) and its inhibitor MTX (lower plot).

its proton. Afterwards THF leaves the pocket to take part in further biochemical reactions, which ultimately lead to the synthesis of the nucleic acid thymine. MTX with its sterical and chemical similarity to DHF can also dock into the pocket of DHFR. Its additional NH_2 group, however, modifies the basicity of the ring and prevents the reduction of the adjacent nitrogen. Additionally, the binding energy of MTX to the DHFR's receptor is larger than that of DHF, allowing it to win the competition for binding sites. As a result, the THF production ceases, the cell is unable to produce DNA and cell division comes to an end.

Because of its property to inhibit the synthesis of THF, MTX is called an *inhibitor*. Figure 9.2, docked, displays the three-dimensional structure of the receptor pocket of DHFR including MTX in its bound state. Such structures of atomistic resolution can now be obtained with the help of X-ray crystallography or NMR spectroscopy for a large variety of ligand-receptor complexes. The figure illustrates the tight fit of the inhibitor into its receptor, several hydrogen bonds lead to a binding mode of high affinity. This close fit of a molecule into its binding pocket is often referred to as the *key-lock principle*. Minor modifications of either the inhibitor (drug design) or the receptor (mutation) can have a drastic impact on the binding affinity or even sterically hinder the ligand from entering the binding pocket at all.

9.2 RATIONAL DRUG DESIGN AND DATABASE SCREENING

The key-lock principle, primarily focused only on geometric criteria [2], is the starting point for rational drug design. If either one ligand of the enzymatic process is known, or the X-ray crystallographic structure of a binding site of the protein has been determined,

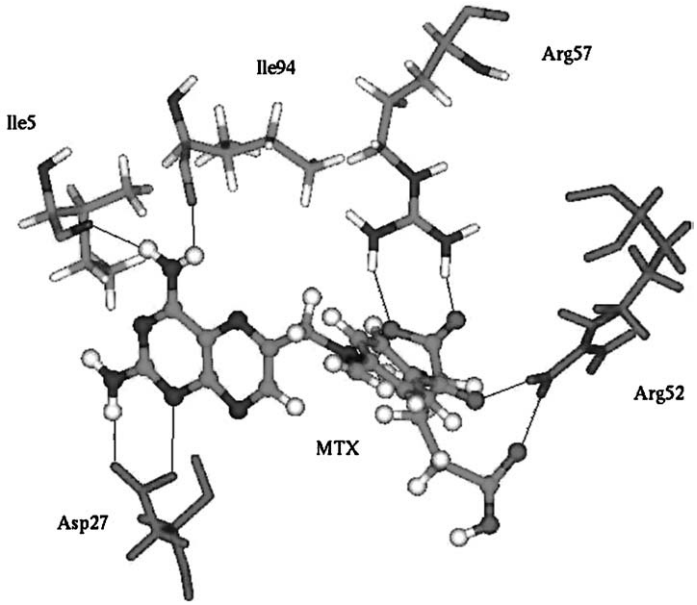


Fig. 9.2. The inhibitor MTX in docked position inside the DHFR receptor. The black lines indicate hydrogen bonds, the red coloured residues were replaced with valine during the second screen (see Color Plate 18).

then a blueprint of a potential drug candidate, a pharmacophore model, can be constructed and molecules designed that share a certain similarity with that blueprint. This strategy has been applied during the last two decades in many successful drug design projects [3–5]. Present effort focuses on increasing the reliablity of the predictions, which to a large degree depend on the viability of the underlying pharmacophore model.

With increasing computational power, a new strategy for computer-aided drug design has become the focus of interest [6–8] in recent years: the screening of large virtual compound databases to a receptor of known structure. The chemical and pharmaceutical industry have compiled large libraries of chemical compounds which have been synthesized over the years, and it was realized that it would save time and costs to start a drug design process from molecules that can be readily purchased instead of being designed from scratch. In the following sections we describe the ingredients and some applications of *FlexScreen*, a tool for virtual database screening that has been developed in recent years at the Karlsruhe Research Center [9,10]. FlexScreen, like many other programs of its class (Autodock [11], GOLD [12], Glide [13], FlexX [14]), aims to select the best possible ligand from a database for a specific receptor.

9.3 ANATOMY OF A VIRTUAL SCREENING TOOL

There are two major ingredients to an *insilico* screening method: (1) a *scoring function* that approximates the binding energy (ideally the affinity) of the receptor–ligand

complex as a function of the conformation of this complex, and (2) an efficient *optimization method* that is able to locate the binding mode of a given ligand to the receptor as the global optimum of the scoring function. In a database screen, all ligands are assigned an optimal score which is then used to sort the database to select suitable ligands for further investigations.

9.3.1 The scoring function

For virtual database screening it is mandatory to model the receptor–ligand system as accurately as possible in order to obtain a precise representation of the key-lock principle. This is best achieved using a scoring function of atomistic resolution. There are two principally different approaches to construct such a scoring function [15]: *Knowledge-based* and *physics-based* scoring functions. The former approach is based on the statistical analysis of a large number of receptor–ligand complexes to determine the most probable distances between various pairs of atom types [16–24]. With an increasing amount and accuracy of available input data, these scoring functions can be 'trained' to find the most likely binding modes of receptor–ligand complexes. This approach has proved to be quite successful, but as an obvious disadvantage it does not offer much information about the fundamental mechanisms of the binding. It therefore provides little help for systematic improvement. In contrast, a physics-based scoring function can be constructed from first principles, although major approximations are presently required to keep the computations feasible. FlexScreen employs the following simple, first-principle scoring function:

$$S = \sum_{\text{Protein}} \sum_{\text{Ligand}} \left(\frac{R_{ij}}{r_{ij}^{12}} - \frac{A_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{r_{ij}} \right) + \sum_{\text{H–bonds}} \cos \Theta_{ij} \left(\frac{\tilde{R}_{ij}}{r_{ij}^{12}} - \frac{\tilde{A}_{ij}}{r_{ij}^{10}} \right) \quad (1)$$

It contains the empirical Pauli repulsion (first term), the Van de Waals attraction (second term), the electrostatic potential (third term) and an angular-dependent hydrogen bond potential (term four and five). The Lennard–Jones parameters (R_{ij} and A_{ij}) were taken from OPLSAA [25], the partial charges q_i were computed with InsightII and esff force field, and the hydrogen bond parameters ($\tilde{R}_{ij}$, $\tilde{A}_{ij}$) were taken from AutoDock [11]. This force field lacks solvation terms to model entropic or hydrophobic contributions. The omission of such terms has been argued to be appropriate for constricted receptor pockets in which all ligands with high affinity displace essentially all water molecules. In a database screen, such a scoring function may therefore preserve the correct *ranking* of the compounds, whereas the absolute scale of their binding energy is far from being accurate.

9.3.2 The global optimization engine

As another important constituent of a screening tool the global minimum of the receptor–ligand scoring function has to be determined. There exist different strategies to address

this problem [26]: methods based on a *deterministic approach* discretize the conformational space of the system and probe each of the possible positions for its score. This approach is fast and exhaustive as long as the dimension of the conformational space remains low. With increasing flexibility of the ligand (or receptor), however, this method suffers from a *combinatorial explosion* caused by the exponential increase of possible conformations.

As an alternative, the *stochastic approach* to global optimization can be applied for high-dimensional systems. The conformation is modified in a random walk [27] and new conformations are accepted using a statistical method based on thermodynamic principles. Consequently, these methods are not exact and their results exhibit a statistical noise, the error of which has to be analysed in order to judge the final accuracy of the database screen [28]. However, they have been shown to obtain good approximate solutions for many problems not presently amenable to enumerative techniques [29]. Among this class of methods, the stochastic tunneling (STUN) technique has been proven to be especially suited to the receptor-ligand problem [9]. This technique was proposed as a generic global optimization method for complex rugged potential energy surfaces (PES). In STUN the dynamic process explores not the original, but a transformed PES, which dynamically adapts and simplifies during the simulation. For the simulations reported here we replace the original transformation [30] with:

$$E_{\mathrm{STUN}} = \ln\left(x + \sqrt{x^2 + 1}\right) \tag{2}$$

where $x = \gamma(E - E_0)$. E is the energy of the present conformation and E_0 the best energy found so far. The problem-dependent transformation parameter γ controls the steepness of the transformation [30]. The general idea of this approach is to flatten the potential energy surface in all regions that lie significantly above the best estimate for the minimal energy (E_0). Even at low temperatures the dynamics of the system becomes diffusive at energies $E >> E_0$ independent of the relative energy differences of the high-energy conformations involved. The dynamics of the conformation on the untransformed PES then appears to 'tunnel' through energy barriers of arbitrary height, while low metastable conformations are still well resolved. Applied to receptor-ligand docking this mechanism ensures that the ligand can reorientate through sterically forbidden regions in the receptor pocket.

9.4 A PRACTICAL EXAMPLE: DATABASE SCREEN TO A RIGID MTX RECEPTOR

In this application [10], the drug MTX was screened together with 10 000 chemical compounds randomly chosen from the open part of the database of the National Cancer Institute (NCI) [31]. The receptor was taken from the X-ray structure of escherichia coli dihydrofolate reductase with MTX in its docked position (pdb entry: 4dfr, monomer B) [32]. The docked ligand was removed from the receptor pocket and the conformation of the target remained fixed during the simulation runs.

Figure 9.3 displays the distribution of binding energies of MTX and those database ligands that had docked into the pocket (left panel). MTX scored best among the 10 000 compounds, which underlines its exceptionally optimized fit to the target (Fig. 9.2, docked).

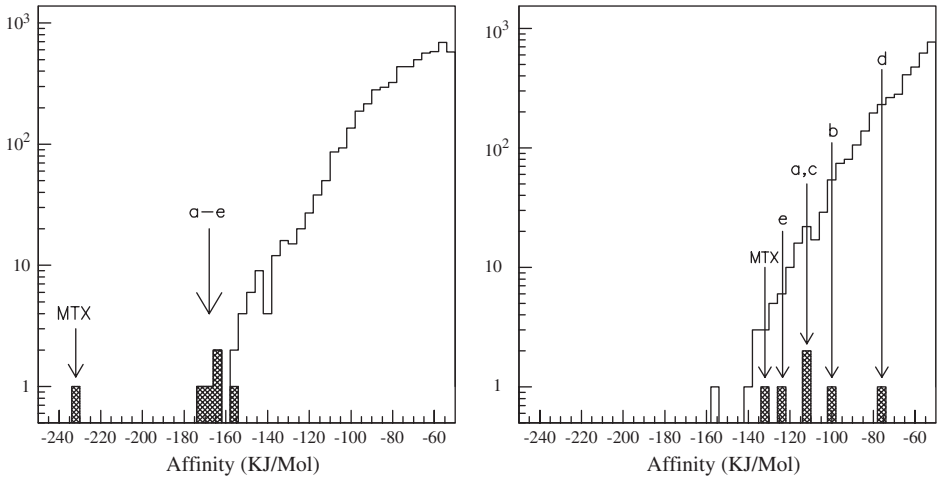


Fig. 9.3. The binding energies of MTX and the docked compounds of the database in the original target (left) and the modified target (right). MTX and the five top ranking ligands of the first run (left) are highlighted in grey.

The right panel displays the distribution of binding energies after a modification of the receptor: Two of the polar residues inside the receptor pocket had been replaced with valine, which is a non-polar residue. Such modifications have been observed after bacteria have mutated and thereby created a resistance against drugs which previously were of high activity. The figure shows that MTX had now lost its exceptionally high binding energy, and the five top ranking compounds of the first run (compounds ‘a’ to ‘e’) had also been replaced with other database compounds.

This result is an impressive demonstration of the key-lock principle. MTX docked into its receptor and scored higher than any other of the compounds. Once the receptor (the ‘lock’) was modified, MTX lost most of its advantage against the competitors. This screen proves a high level of specificity; It selected the correct inhibitor for this specific target among thousands of compounds.

The impressive level of specificity of such a database screen, however, also contains a weakness of this approach: MTX was fitting tightly into the pocket because the pocket’s conformation was taken from the X-ray structure of MTX docked into that pocket and kept rigid during the simulation. In reality the positions of the residues inside the pocket are not always conserved and may depend on the interaction with its ligand. The neglection of these receptor degrees-of-freedom may reduce the chance of alternative compounds to create an individual tight binding mode, thereby reducing the diversity of the database screen. In other words: MTX may have dominated the competition so clearly because the target was biased towards its optimum binding mode. In a real-life application, the screening tool should be able to select good ligands without the existence of such a biased X-ray structure. It was therefore decided to develop FlexScreen further and include target flexibility.

9.5 TARGET DEGREES OF FREEDOM: SPECIFICITY VERSUS DIVERSITY

While ligand flexibility is now routinely considered in many atomistic in-silico screening methods, accounting for receptor flexibility still poses significant challenges [33–36]. This is not surprising because the introduction of additional receptor degrees of freedom adds to the complexity of the conformational space to an extent that most of the existing docking tools are unable to deal with.

The following investigations were made using the thymidine kinase (TK) receptor. This enzyme has long since been a focus of pharmaceutical research because of its role in the reproduction of the herpes simplex virus. It has since then emerged as a useful benchmark system in rational drug design, because not just one, but several active inhibitors are known and the X-ray structures of their binding modes have been measured [37]. As has been discussed elsewhere, they differ significantly in their target conformations [38]. In the following sections we will first compare how database screens to several of these rigid receptor conformations were leading to different and even contradictive results, and then we repeat the same database screen to a receptor with certain selected degrees of freedom.

9.5.1 Screens using various rigid TK receptor conformations

We investigated the degree of database enrichment of 10 000 compounds, randomly chosen from the NCI database, and 10 known inhibitors when docked to the X-ray TK receptor structure, which was experimentally determined in complex with one of the inhibitors, dt (deoxythymidine, pdb entry 1kim [39]). In this screen 5 353 of the 10 000 database ligands attained a stable conformation with negative affinity within the receptor pocket. The resulting ranks for the inhibitors during this screen are summarized in Table 9.1 (second column). Three structurally similar inhibitors, including the ligand (dt) associated with the receptor conformation, were ranked with very high affinity. This result demonstrates that the docking method and scoring function were adequate to approximate the affinity of these ligands to the receptor, i.e., the specificity of the selection method was high. Four further ligands (idu, acv, gvc, pcv, for a detailed description of TK and its inhibitors we refer to [37]) docked badly, and three further ligands did not dock at all according to the criteria above. Repeating the docking simulations for these ligands did not substantially improve their rank in the database, eliminating statistical fluctuations of the docking algorithm as the source for this difficulty. This enrichment rate is comparable to the results of other scoring functions that were previously investigated for this system, but the overall performance is disappointing [37].

The inspection of the crystal structures of the different receptor-ligand complexes reveals differences in the conformation of some side groups inside the receptor pocket, depending on the docked inhibitor. This is a well-known fact, but it is often assumed that the impact of these conformational variations on the ranking accuracy is moderate. We therefore repeated the screening with the X-ray structure of TK in a complex with the inhibitor gcv (ganciclovir, pdb entry: 1ki2 [39]), which had scored particularly badly in the original screen. The results are shown in Table 9.1 (third column). Now, gcv was ranked within the leading 1 percent of the database, but dt, formerly ranked on position 5, dropped to 1 310. Figure 9.4 (left panel) shows the number of docked ligands as a function of binding energy and highlights the ranks

Table 9.1 Ranking of the TK inhibitors in a screen of 10 000 randomly chosen ligands of the NCI database. The top row designates the crystal structure of the receptor that was used in the screen, the last column indicates the results of a flexible receptor screen. (nd = not docked)

Inhibitor	1kim	1ki2	1ki3	1e2h	flex
acv	719	9	22	2 048	199
ahiu	nd	nd	nd	nd	
dhbt	4	104	118	38	13
dt	5	1 310	2 576	2 779	681
gcv	3 351	78	15	4 516	57
hmtt	nd	nd	nd	nd	656
hpt	6	152	266	36	148
idu	515	2 436	3 272	2 913	1 365
idu	515	2 436	3 272	2 913	1 365
mct	nd	6 074	nd	nd	247
pcv	4 845	952	4	4 739	1 656
Score:	3 751	3705	4 575	1 926	4 999

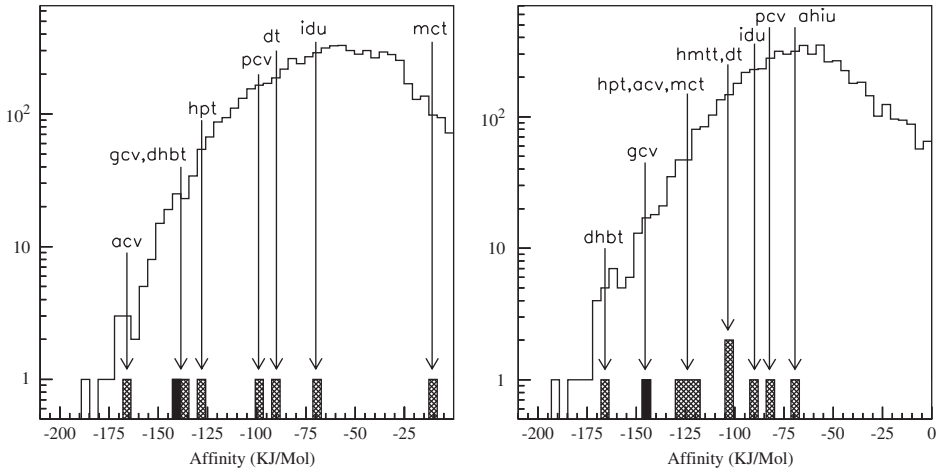


Fig. 9.4. Histogram of the binding energies of the docked ligands after a screen to the TK receptor (positions of the known inhibitors are highlighted). Left: Screen to the rigid target conformation of gcv. Note that only 8 of the 10 inhibitors have docked at all. Right: Screen using a flexible target: All 10 inhibitors have docked.

of the known TK inhibitors in the screen. The same procedure was then repeated with TK in complex with pcv (penciclovir, pdb entry: 1ki3), which raised its rank from 4 845 to 4 (fourth column of Table 9.1).

For comparison purposes, we also performed a screen of the ligand free X-ray structure of TK (pdb entry: 1e2h [40]), which would most likely be used in a screen if no inhibitor was known. In this screen the receptor is unbiased to any of the inhibitors, which results in a dramatic loss of screening performance. As shown in column 7 only two ligands scored reasonably well (within the upper 10 percent of

the database), all others would be discarded by any rational criterion as possible lead candidates.

These screening runs once more demonstrate the impact of the key-lock principle: Inhibitors were correctly identified if the target conformation was taken from X-ray structures of the receptor in a complex with this particular inhibitor, proving a high degree of specificity of the screening technique. However, other inhibitors, known to be active in experiment, were ranked badly if the receptor conformation was not biased towards their individual binding modes. The diversity of these runs was poor. Especially disappointing is the fact that while using the unbiased receptor conformation (1e2h) most of the inhibitors failed to be selected as good drug candidates. In pharmaceutical applications, a good inhibitor may not be readily available and therefore no obvious target conformation that offers the 'lock' to which the 'key' is to be found, may exist. It is obvious that any rigid target conformation would reject the majority of good drug candidates due to the high degree of specificity of the database screen, so that the introduction of target degrees of freedom becomes an important tool to increase the diversity of the method.

9.5.2 Flexible receptor screens

We therefore performed a flexible receptor screen against the same database. We identified the critical amino acid sidechains and introduced 23 receptor degrees of freedom into the structure 1ki2, i.e., dihedral rotations of the amino acids His13(2), Gln76(3), Arg173(4), Glu176(4), Tyr52(3), Tyr123(3) and Glu34(4). The numbers in brackets indicate the degrees of freedom for each sidechain. Each step in the stochastic search now consisted of an additional random rotation for one receptor degree of freedom. The results of this screen are summarized in Fig. 9.4 (right panel), the scores of the individual inhibitors are listed in the column labelled 'flex' in the table. The figure demonstrates that in contrast to all rigid receptor screens now all inhibitors dock to the receptor. As expected, the number of database compounds that achieved a high binding energy increases as well, because a flexible conformation of the receptor reduces the bias of the screen against the known inhibitors. The diversity of the docking tool has increased, but the specificity has decreased because the 'lock' now allows for a broader class of 'keys' to fit. It was also observed that the accuracy of the flexible receptor screen is lower than that of the rigid receptor screens (with the same number of function evaluations) because the number of degrees of freedom has increased. The increased fluctuations in the flexible screen can be best seen for acv, where the optimization method failed to locate the global optimum of the affinity (as independently obtained in a longer screen for just this ligand). We are presently developing algorithms to move only the sidechains close to the present ligand position to reduce the computational effort in the flexible receptor screen.

9.5.3 Comparison: Rigid screen versus flexible screen

To quantitatively compare different screens against the same ligand database, which used different receptor geometries, scoring functions or docking methods, it is sensible to assign

an overall score to each screen that rates its performance [41]. We computed such a 'score' for the entire screen from the ranks of the docked known inhibitors among the $N = 1\,000$ best ligands. This score is computed as the sum of $N - P$ where P is the rank of the known inhibitor and shown in the bottom row of Table 1. An inhibitor ranking in the top of the screen contributes a score of 1 000 to the sum, a badly ranked inhibitor comparatively little. Because the best ligands are evaluated, screens which dock many known inhibitors with moderate rank may have comparable scores with screens that perform perfectly for one inhibitor, but fail for all others. For the rigid receptor screens performed here the scores for the entire screen ranged from between 1926 for the screen against 1e2h, the ligand free X-ray structure of TK, to 4 575 (1ki3, X-ray structure of TK in complex with pcv), which was arguably the best performing screen of all receptor conformations. Despite the increase in the total number of docked compounds the overall score of the flexible receptor screen (4 999) was better than that of any rigid receptor screens. Seven of the ten inhibitors scored within the first 10 percent of the database. This result indicates that the increase in diversity of the technique had out-balanced the decrease in specificity, leading to an overall better docking performance of the screening tool.

9.6 CONCLUSIONS

Our results offer a good demonstration that the ranking of known inhibitors can strongly depend on the particular receptor structure used for the screen. Binding energy and rank of a given inhibitor differ significantly depending on the receptor conformations. Our data demonstrates that this variability in rank is not, in general, a shortcoming of either scoring function nor docking methodology. Using a fixed three-dimensional structure of the receptor that is suitable for a single ligand introduces a significant bias in the overall scoring of the entire database. As a consequence, differences in the enrichment ratio for different scoring functions [37] may depend more on the suitability of the receptor conformation and environment than on the quality of the scoring function.

These findings suggest the importance of the consideration of a flexible binding pocket to obtain a better unbiased scoring of high-affinity ligands. The results of the flexible receptor screen reported here suggest that better accuracy of the scoring process can be achieved when receptor flexibility is considered. Ultimately, only the routine use of accurate scoring techniques for flexible receptors, such as *FlexScreen*, will ameliorate this problem. The results presented here demonstrate that such screens will become feasible with present day computational resources in the near future.

9.7 ACKNOWLEDGEMENTS

We thank the Fond der Chemischen Industrie, the BMBF, the Deutsche Forschungsgemein-schaft (grant WE 1863/11-1) and the Kurt Eberhard Bode Stiftung for financial support.

9.8 REFERENCES

1 S.H. Fleischman and C.L. Brooks III, PROTEINS, 7 (1990) 52.
2 E. Fischer, Ber. Dtsch. Chem. Ges., 27 (1894) 2985.
3 P. Gund, W.T. Wiebke and R. Langridge, Comput. Chem. Res. Educ. Technol., 46 (1974) 528.
4 Z. Simon, A.T. Balaban, D. Ciubotariu and T.S. Balaban, Revue Roumaine de Chimie, 30 (1985) 985.
5 R.D. Cramer III, D.E. Patterson and J.D. Bunce, J. Am. Chem. Soc., 110 (1988) 5959.
6 W.P. Walters, M.T. Stahl and M.A. Murcko, Drug Discovery Today, 3 (1998) 160.
7 J. Drews, Science, 287 (2000) 1960.
8 R. Abagyan and M. Totrov, Curr. Opin. Chem. Biol., 5 (2001) 375.
9 H. Merlitz and W. Wenzel, Chem. Phys. Lett., 362 (2002) 271–277.
10 H. Merlitz, B. Burghardt and W. Wenzel, Chem. Phys. Lett., 370 (2003) 68.
11 G.M. Morris, D.S. Goodsell, R. Halliday, R. Huey, W.E. Hart, R.K. Belew and A.J. Olson, J. Comput. Chem., 19 (1998) 1639.
12 G. Jones, P. Wilett, R.C. Glen, A.R. Leach and R. Taylor, J. Mol. Biol., 267 (1997) 721.
13 R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Reparsky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis and P.S. Shenkin, J. Med. Chem., 47 (2004) 1739.
14 M. Rarey, B. Kramer, T. Lengauer and G.A. Klebe, J. Mol. Biol., 261 (1996) 470.
15 G. Klebe and H. Gohlke, Angew. Chemie (Intl. Ed.), 41 (2002) 2644.
16 G. Verkhivker, K. Appelt, S.T. Freer, and J.E. Villafranca. Protein Eng., 8 (1995) 677.
17 R.S. DeWitte and E.I. Shakhnovich, J. Am. Chem. Soc., 118 (1996) 11733.
18 R.D. Head, M.L. Smythe, T.I. Oprea, C.L. Waller, S.M. Green and G.R. Marshall, J. Am. Chem. Soc., 118 (1996) 3959.
19 H.J. Böhm, J. Comput. Aided Mol. Des., 12 (1998) 309.
20 J.B.O. Mitchell, R.A. Laskowski, A. Alex and J.M. Thornton, J. Comput. Chem., 20 (1999) 1165.
21 I. Muegge and Y.C. Martin, J. Med. Chem., 42 (1999) 791.
22 D. Hoffmann, B. Kramer, T. Washio, T. Steinmetzer, M. Rarey and T. Lengauer, J. Med. Chem., 42 (1999) 4422.
23 H. Gohlke, M. Hendlich and G. Klebe, J. Mol. Biol., 295 (2000) 337.
24 C. M. Venkatachalam, X. Jiang, T. Oldfield and M.J. Waldman, J. Mol. Graphics Modell., 21 (2003) 289.
25 W.L. Jorgensen and N.A. McDonald, J. Mol. Struct., 424 (1997) 145.
26 G. Schneider and H.J. Boehm, Drug Discovery Today, 7 (2003) 64.
27 N. Metropolis and U. Stanislaw, JASA, 44 (1949) 335.
28 H. Merlitz, T. Herges and W. Wenzel, J. Comp. Chem., 25 (2004) 1568–1575.
29 S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi, Science, 220 (1983) 671–680.
30 W. Wenzel and K. Hamacher, Phys. Rev. Lett., 82 (1999) 3003–3007.
31 G.W.A. Milne, M.C. Nicklaus, J.S. Driscoll, D. Zaharevitz and S. Wang, J. Chem. Inf. Comput. Sci., 34 (1994) 1219.
32 J.T. Bolin, D.J. Filman, A. Matthews, R.C. Hamlin and J. Kraut, J. Biol. Chem., 257 (1982) 13650.
33 V. Schnecke and L.A. Kuhn, Persp. Drug Des. Discovery, 20 (2000) 171.
34 H. Claußen, C. Buning, M. Rarey and T. Lengbauer, J. Mol. Biol., 308 (2001) 377–395.
35 F. Osterberg, Proteins, 46 (2002) 34.
36 H. A. Carlson, Curr. Opin. Chem. Biol., 6 (2002) 447.
37 C. Bissantz, G. Folkerts and D. Rognan, J. Med. Chem., 43 (2000) 4759.
38 H. Merlitz and W. Wenzel, Chem. Phys. Lett., 390 (2004) 500.
39 J.N. Champness, M.S. Bennett, F. Wien, R. Visse, W.C. Summers, P. Herdewijn, E. de Clerq, T. Ostrowski, R.L. Jarvestand and M.R. Sanderson, Proteins, 32 (1998) 350.
40 J. Vogt, R. Perozzo, A. Pautsch, A. Prota, P. Schelling, B. Pilger, G. Folkerts, L. Scapozza and G.E. Schulz, Proteins, 42 (2000) 545.
41 R.M.A. Knegtel and M. Wagnet, Proteins, 37 (1999) 334.

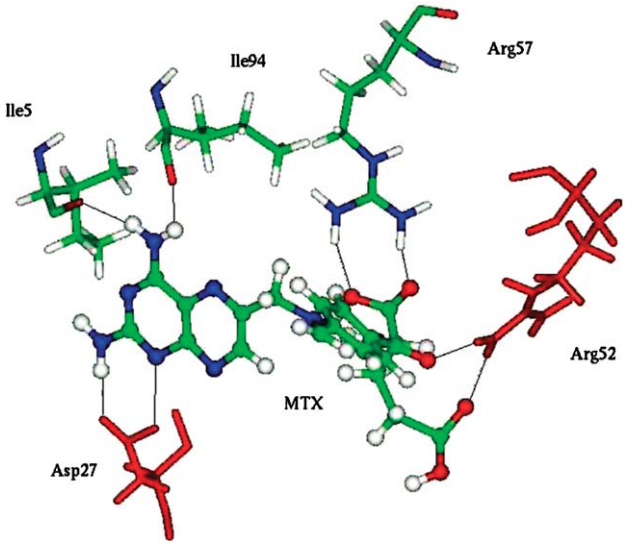


Plate 18. The inhibitor MTX in docked position inside the DHFR receptor. The black lines indicate hydrogen bonds, the red coloured residues were replaced with valine during the second screen (illustration appears on page 181 of this volume).

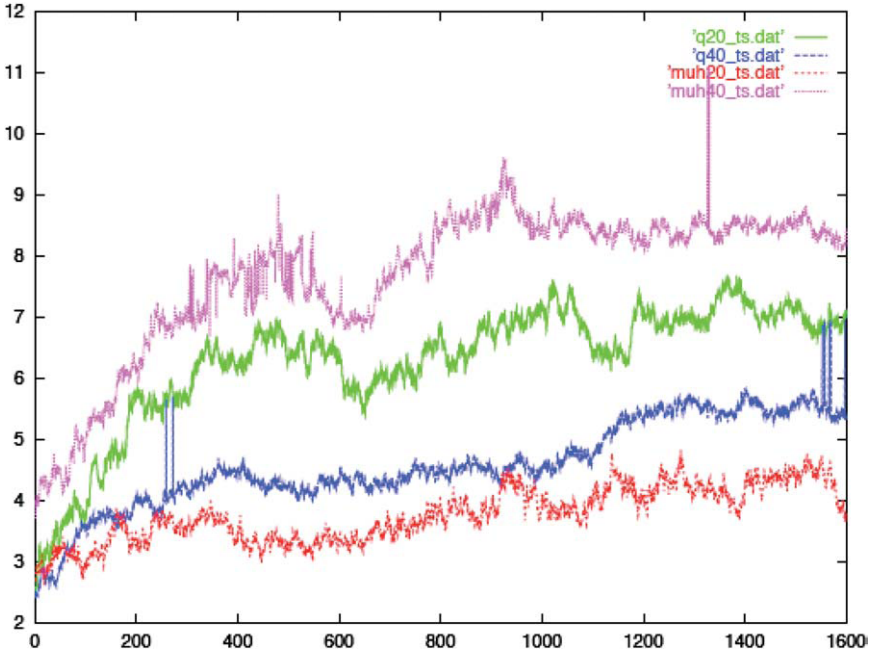


Plate 19. Root-mean-square deviation (Rmsd) plot for Q20 and Q40 in the μ-helix and Perutz helix conformations, showing RMSD between initial structures described in Section 11.1.3.2 and those produced in the MD trajectory as a function of simulation time (illustration appears on page 219 of this volume).

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 10

Enzymatic recognition of radiation-produced oxidative DNA lesion. Molecular dynamics approach

Miroslav Pinak

Japan Atomic Energy Agency, Shirakata, Shirane 2-4, 319-1195 Tokai, Japan

Abstract

A Review of molecular dynamics (MD) studies of several DNA lesions on DNA is presented. The main focus is to describe structural and energy changes in DNA molecules with respect to proper recognition of the lesion by respective repair enzymes. Pyrimidine lesions (cytosinyl radical, thymine dimer, thymine glycol) and purine lesion (8-oxoguanine) were subjected to MD simulations for several hundreds of picoseconds using MD simulation code Amber and its respective force field modified for each lesion. In most cases the significant structural changes in DNA are observed: breaking of the hydrogen bond network opening and bending the DNA double helix. These changes are related to overall collapsing of the double helical structure around the lesion and are considered to facilitate docking of repair enzymes into DNA and the formation of a DNA–enzyme complex. In addition to structural changes, specific values of electrostatic interaction energy are detected at several lesion sites. This specific electrostatic energy is considered as a factor that enables repair enzymes to discriminate between lesion and undamaged DNA site.

10.1 INTRODUCTION

All organisms that use molecular oxygen need to defend themselves from reactive byproducts of oxygen metabolism. Reactive oxygen species such as superoxide, hydrogen peroxide, and hydroxyl radical can be produced by the incomplete reduction of oxygen during aerobic metabolism [1]. Intensive studies have determined that the deoxyribonucleic acid (DNA) macromolecule is the most important component in the study of oxidative damage to cells. Its changes may persist over an indefinite number of generations. It is generally considered to be the target for endogenous as well exogenous action. Oxidative damage to deoxyribonucleic acid has been estimated at 10^9

lesions per cell per day in humans. Although organisms have developed systems to control active oxygen species, those that escape the primary defenses can damage nucleic acids and other cellular macromolecules.

Several metabolic pathways generate oxidative radicals within cells, and these radicals can attack DNA to give both DNA molecular structure alteration or strand break. The significance of effective repair of these damages comes from the facts that DNA is a repository of hereditary information and it is the blueprint for the operation of individual cells. Considering these important features, it can be concluded that nearly all DNA damage is harmful. Not surprisingly, organisms have developed second-line defense systems to repair the oxidative damage to DNA in order to reduce this damage to a tolerable level. DNA repair is very important since DNA is the only biomolecule that is specifically repaired, all the others are replaced. The repair process is very complex and more than 100 genes have been found to participate in various aspects of DNA repair, even in organisms with very small genomes.

Cancer, as one of the most serious consequences of DNA damage is caused by mutations at the DNA level. In most cases, the 'genomic instability' (elevated mutation rate) is required to permit accumulation of sufficient mutations to generate cancer during a human lifetime. DNA repair mechanisms promote genomic stability and prevent cancer. Many, perhaps most, cancers are thus at least partially attributable to defects in DNA repair [2–5].

In addition to experimental assays, theoretical investigations of lesioned subcellular biological structures play an important role. Computational simulation is one of the theoretical methods which allows one to describe the stochastic nature of events that occur at the micro- or nanometer level in a DNA molecule. Molecular dynamics (MD) simulation is a computational technique allowing detail conformational and energy studies of molecular changes in very short time segments of the order of femto- or picoseconds that may correspond to the chemically important time intervals of formation and breakage of chemical bonds.

10.2 COMPUTATIONAL STUDIES AND RADIATION RISK

Among other DNA damaging agents, ionizing radiation represents a significant factor producing damage to DNA by ionization or excitation of atoms and molecules surrounding DNA as well as DNA subunits themselves. Radiation risk estimates of the effects on living material are primarily based on epidemiological studies. These studies are mostly relevant for high doses and dose rates and may lack sufficient information at low doses, which are more relevant to human exposure to radiation. Extrapolation of high-dose results to low-dose regions requires a detailed understanding of the mechanisms by which radiation induces molecular alterations, cell mutations, cancer and other forms of genetic disorders. The biochemical and biological mechanisms of the cell response to the radiation and oxidative damage are controlled by a specific set of genes that encodes enzymes catalyzing cellular response to radiation damage. These enzymes (e.g. hRAD51, XRCC4, hRAD52, hREV3, hNTH, hOGG1, etc.) attempt to repair the damage when it occurs. Repair enzymes are involved in DNA repair via several pathways and are functioning in certain phases of the complex repair process.

The following are samples of several radiation-originated human disorders characterized by defects in DNA repair.

(a) Patients with xeroderma pigmentosum (XP) have clinical sun an sensitivity, extensive freckle-like lesions in sun-exposed parts of the skin, and have an increase in incidence of skin cancer (basal cell carcinoma, squamos cell carcinoma and melanoma). All XP cells have been detected to be deficient in DNA repair.
(b) Patients with another human disorder–Cockayne syndrome (CS) have increased sun sensitivity, short stature and progressive neurological degeneration. In contrary to the xeroderma pigmentosum, Cockayne syndrome does not result in cancer-proneness. Cultured cells from CS patients have defective DNA repair.
(c) Patients with trichodiostrophy have short stature, mental retardation and brittle hair. Their cells also have defective DNA repair.

For the study of qualification of radiation risk, it is necessary to determine the specific pathways and features that are typical for radiation-originated DNA damage. In addition, the quantification of radiation risk would need to involve the study of relative efficiency of the repair processes with respect to the increased incidence of DNA damage above the endogenous level. The ultimate goal of radiation risk-related computational studies is to investigate the sequence of events that leads from energy deposition of radiation in the material of a living cell to the significant changes that occur at the level of cellular DNA.

10.3 DNA REPAIR

All the above named disorders have similar DNA repair pathways at the molecular level–base or nucleotide excision repair. Specifically, isolated DNA base damages are removed by a base excision repair pathway while bulky damages are repaired by a nucleotide repair pathway in which a damaged segment of DNA strand is removed. Detailed descriptions of DNA radiation damages' repair pathways can be found elsewhere [e.g., 6]. The excision repair pathway is composed of the following steps with several enzymes acting in each of them:

(a) DNA damage recognition; repair enzyme specifically searches for and binds to DNA lesion, and eventually attracts one or more repair enzymes to the damaged site;
(b) DNA helix opening; repair enzymes unwind the DNA double helix around the lesion;
(c) dual incision; enzyme cuts the damaged DNA strand on both sides of the lesion;
(d) excision; damaged DNA fragment containing the lesion is removed;
(e) gap-filling; the remaining gap is filled in by polymerization and ligation.

The complexity of these repair pathways and the large number of enzymes acting in each step do not allow a simple approach to their solutions. MD simulations of molecules acting on the first step of repair, that is, recognition of lesion, may provide the stepwise

description of biomolecular reactions. In combination with a quantum chemical approach, which allows the study of qualitative changes at the atomic and molecular level, the MD simulation determines specific structural conformation and energetic properties of the lesion that guide a repair enzyme to discriminate a lesion from an undamaged DNA part.

10.4 METHOD

Since the DNA molecule is not rigid, static structure, the X-ray diffraction and NMR results usually show average structural parameters. In reality, every DNA molecule is under constant thermal fluctuations, which result in local twisting, stretching, bending and unwinding of the double helix. The classical MD is based on solving Newton's equations of motion for each atom in the system. Solving of these equations produces new atomic coordinates that can be used to calculate a new set of forces. Static and dynamic properties of the system are then obtained as time averages over the trajectory. For simulations reported in this section the molecular dynamics program package Amber (versions 5 and 7) was used [7,8].

The module Nucgen of the Amber software package was used to prepare an undamaged DNA segment. In the case of damaged DNA, the standard force field parameters included in Amber need to be modified to satisfy new parameters of a particular lesion. These parameters are calculated using the quantum mechanical approach employing Hartree–Fock theory implemented in Gaussian [9] or other software packages. In Hartree–Fock theory, the variational principle is used to determine the best possible state of the molecule when its energy is at a minimum.

Total molecular-mechanical energies and internal energy (bonds, angles and dihedrals), van der Waals and electrostatic components discussed in this section were determined using module Anal of the Amber software package with the same force field that was used in the MD simulations with the inclusion of solute–solute pairwise interactions. Amber also supports free energy calculations, where one chemical species is transformed into another one and the (free) energy change associated with the transformation is measured. For all simulations presented in this section the free energy calculation was not used and the potential energy of the molecular function was calculated as contributions from bonds, single angles, torsional and electrostatic functions (Eq. 1):

$$\begin{aligned} E_{\text{pot}} = {} & 1/2\sum_{\text{bonds}} K_r(r-r_{\text{eq}})^2 + 1/2\sum_{\text{angles}} K_\theta(\theta-\theta_{\text{eq}})^2 \\ & + 1/2\sum_{\text{dihedrals}} K_\varphi[1+\cos(n\varphi-\chi)] \\ & + \sum_{j=1}^{\text{atoms}}\sum_{i>j}^{\text{atoms}}\left[\left(\frac{R_{ij}^*}{r_{ij}}\right)^{12} - \left(\frac{R_{ij}^*}{r_{ij}}\right)^{6} + \frac{q_i q_j}{\varepsilon r_{ij}}\right] \end{aligned} \tag{1}$$

Here r, θ and φ are bond lengths, planar angles and dihedral angles, respectively; r_{eq} and θ_{eq} are their equilibrium values; r_{ij} is the distance between atoms i and j; q_i is the partial charge on atom i; ε is dielectric constant; K_r, K_θ, K_φ, and R_{ij} are empirical parameters

depending on atom types. The point charges q_i and q_j in a calculation of electrostatic function are centered on each atom and are derived by fitting to quantum mechanical electrostatic potential.

The general strategy used in this work involves the calculation of energies for molecular configurations taken from the MD trajectories of the simulated molecules and then averaging these values.

The MD protocol used in all simulations consisted of several preparatory steps–minimization, stepwise heating up to 310K, density stabilization and production dynamics. Details of the MD protocol have been described elsewhere [e.g., 10,11].

To satisfy the electronegativity of the system, the appropriate number of sodium counterions (Na^+) were placed at the initial positions bisecting the O–P–O angle at the initial distance of 5 Å from phosphorus atoms of DNA (version Amber 5) or using module Xleap (version Amber 7). During the MD simulation no restrictions were applied on the positions of counterions.

Solvating of solute molecules (DNA and enzyme), i.e., immersing solute molecules into the water box consisting of several thousands water molecules, resulted in the system consisting of several tens of thousands of atoms in total, depending on the size of the solute molecule(s).

10.4.1 Analysis of electrostatic energy

Electrostatic interactions are found to be a dominant factor in determining conformation of biomolecules, are significant in ligand–receptor binding and play an important role in the stability and function of biomolecules, e.g., [12,13]. For example, chemical reactivity, molecular recognition, and biological activity are known to correlate with electrostatic potentials, and may be critical to achieving a physical allocation of solvation effects such as solvation energies and molecular properties. The accurate modeling of molecules in solution requires realistic models for the interaction of solvent with molecules [14–16]. To treat such a medium in a molecular calculation, both explicit and implicit models have been developed [17]. Explicit solvent models rely on using hundreds of thousands of discrete solvent molecules [18]. These are the most widely used methods for performing simulations in solvent. Such calculations converge relatively slowly because of the large number of atoms involved and thus require much more CPU time. As well as explicit models there are also implicit models that treat the solvent as a continuous medium having the average properties of the real solvent, and surrounding the solute beginning at the van der Waals surface. A variety of continuum models have been described, e.g., [19–22]. Among these, the generalized born approximation (GBA) and analytical continuum electrostatic model (ACE) are the most popular.

In the presented simulations, the explicit model involving a large number of water molecules was used. To conform to the Amber force field, the solute molecules were surrounded by explicit solvent and a dielectric multiplicative constant equal to one for electrostatic interactions was used. One to four van der Waals interactions and one to four electrostatic interactions (interactions separated by only three bonds) were divided by 2.0 and 1.2, respectively. These are the recommended values for the Amber force field. In the performed MD simulations a constant dielectric function was used.

Electrostatic energy (included in the fourth term of Eq. (1)) was calculated as the sum over all atoms of a solute molecule (or its part). This calculation is based on atom-centered point charges and the sum runs over all atoms of the computed system. This way calculated electrostatic interaction includes factors originated by the entire solute molecular composition as well as those from the solvent environment. In addition, several electrostatic properties were calculated using the DelPhi program in order to evaluate the impact of DNA lesions on the DNA structure and the interaction of the lesions with repair enzymes and the surrounding environment. DelPhi is the program which uses the finite difference method to solve the Poisson-Boltzmann (P–B) equation for molecules of arbitrary shape and charge distribution [23]. The P–B equation is solved for the system containing solute molecule(s) and ions. The overall charge density in the system is obtained from the charges of the DNA adding the contributions from dissolved ions, Eq. (2):

$$\Delta\varepsilon(r)\Delta\varphi(r) = \rho_{\text{macro}}(r)\varphi + \sum_i q_i \eta_i^0 \exp^{-q_i\varphi(r)/kT} \tag{2}$$

where ρ_{macro} r is the charge density due to the DNA and the sum of the exponentials is the charge density due to the dissolved ions (φ is the electrostatic potential in the region of space, ε is dielectric constant, q_i is the charge on dissolved ions, η_i^0 is the density of ions i in the bulk solution, k is the Boltzman's constant and T is the temperature). The electrostatic energy is the only long-range non-bonded interaction in the potential energy function (Eq. (1)), and its value is calculated for pairs of atoms that are separated by three or more bonds. Thus it was assumed that the well-separated atoms were interacting mainly electrostatically, since no cut-off distance was used in the calculation of the electrostatic interactions. The electrostatic potential $\varphi(r)$ at a derived position r includes the contributions from all atoms of the molecule as well as from the ions in the environment.

The particle mesh ewald (PME) method [24], with no cut-off distance was used for the calculation of electrostatic interactions. In this method the Gaussian charge distribution of opposite sign is superimposed upon the original point charges, producing a screened charge distribution. The electrostatic interaction between the screened charges is then short ranged. The original distribution is recovered by adding a second Gaussian charge distribution identical to the first, but of opposite sign. In the calculation of electrostatic interactions no cut-off distance was applied and thus all water molecules in the system were included. The electronegativity assumed in the PME method was satisfied by adding an appropriate number of counterions as described in the previous paragraph.

The van der Waals interactions were calculated within a defined cut-off distance (around 10Å, which is well beyond the van der Waals radius; e.g., 1.7Å for C). Periodic boundary conditions were applied throughout all simulations to eliminate undesirable edge effects.

10.5 RESULTS AND DISCUSSIONS

10.5.1 Cytosinyl radical (5-hydroxy-6-cytosinyl radical)

The cytosinyl radical is the lesion produced by indirect radiation action, i.e., in the interaction of the active OH water radical produced in water radiolysis with the cytosine base.

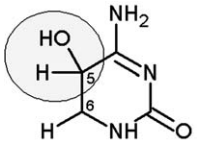


Fig. 10.1. The molecule of 5-hydroxy-6-cytosinyl radical with the axial position of OH in the C5 atom of the pyrimidine ring (marked by shadow). The structure was obtained by quantum mechanical structure optimization.

Formation of pyrimidine radicals by OH addition is kinetically controlled with relative energies of the radicals with the OH group on C6 lower than those with the OH on C5. The structure of the 5-hydroxy-6-cytosinyl radical with the axial position of OH in the C6 atom of the pyrimidine ring is shown in Fig. 10.1.

This lesion is important in the study of strand break formation through the intramolecular process of H-abstraction from pentose and it emphasizes the importance of initial base damage in connection and strand breaks. The 200 picoseconds (ps) of MD simulation of DNA dodecamer d(CGCGAATTC*GCG)$_2$ with cytosinyl radical C*(9) revealed a strong bending at the A(6) and T(7) DNA segment (numbers in parentheses indicate position of nucleotides in the simulated DNA segment) [25]. Since this large bending was observed at A(6) and T(7) and not at the damaged site C*(9), it suggests the existence of intramolecular interactions among C*(9) and A(6) and T(7). In addition to the bending, large distortions and disruptions of the hydrogen bonding network between bases of neighboring pairs were also observed. The results indicate that even small and local damage in DNA may cause a large and global conformational change. This conformational change may satisfy the favorable environment for processing of the lesion by repair enzymes or eventually inhibit its proliferation.

10.5.2 Thymine dimer (5,6 cis.sin cyclobuthane thymine dimer)

Cyclobuthane thymine dimer is a photolesion produced by UV radiation in sunlight and is considered as a potential factor causing skin cancer. It is formed as a covalently bonded complex of two adjacent thymines on a single strand of DNA. This damage is very frequent but almost 90 percent of TDs are repaired within a short time of the order of minutes, and only a few are experimentally observable and originate future changes at the cell level.

This study was conducted with DNA dodecamer d(TCGCG′TD′GCGCT)$_2$, where TD refers to thymine dimer [10]. The results of 600 ps of MD simulation indicate that this lesion does not disrupt the double helical structure and the hydrogen bonds are well preserved throughout the simulation. Instead, a thymine dimer lesioned DNA, if compared with a non-lesioned one, has sharp bending at the dimer site which is originated by two covalent bonds C(5)-C(5) and C(6)-C(6) between adjacent thymines forming thymine dimer (Fig. 10.2). This bending is supposed to discriminate lesion from undamaged DNA segment and to originate conformation that facilitates the formation of a DNA–enzyme complex by generating complementary structural shapes of repair enzymes and bent DNA.

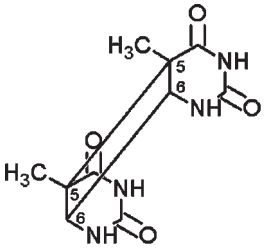


Fig. 10.2. Thymine dimer as a composition of two adjacent thymine bases covalently joined between C(5)–C(5) and C(6)–C(6) atoms of adjacent thymine bases.

Thymine dimer excision repair is initiated by *E. coli* endonuclease V of bacteriophage T4 that slides on non-target sequences and progressively incises at all dimers within the DNA molecule. This enzyme binds to the DNA double strand in a two-step process: at first it scans non-target DNA by electrostatic interaction to search for damaged sites, and secondly it sequentially specifically recognizes the dimer sites. The process of binding of T4 endonuclease V to thymine dimer lesioned DNA was simulated using the MD method. Considering the limitations arising from the simulations of large systems and requirements for CPU time, instead of the whole enzyme a small isolated part that included a catalytic center was subjected to the simulations. The part of the enzyme was selected considering the structure of T4 endonuclease V. This enzyme consists of three *a* helices (H1: amino acids 14–38, H2: 64–82 and H3: 108–124) standing side by side, several reverse turns and several loops [26]. Glu-23, of which a carboxyl chain plays a crucial role in the cleavage of the *N*-glycosyl bond in DNA during enzymatic repair process [27,28], is surrounded by amino acids Arg-3, Arg-22 and Arg-26 belonging to helix H1. The sidechain of Glu-23 also forms hydrogen bonds with the sidechains of helices H1 and H2, and lies on the molecular surface. Considering these properties, eight amino acids of H1, Glu-20, Tyr-21, Arg-22, Glu-23, Leu-24, Pro-25, Arg-26, Val-27, and two amino acids at the NH_2 terminus, Thr-2 and Arg-3, were selected to form the simulated part of the enzyme. Amino acids Thr-2, Arg-22, Glu-23 and Arg-26 form the catalytic center that is active in the incision of the thymine dimmer during the repair process and together with other six selected amino acids: Arg-3, Glu-20, Tyr-21, Leu-24, Pro-25 and Val-27, are located at the central part of a concave site of the enzyme and may be easily exposed to the DNA surface.

MD simulation showed that after nearly 100 ps of the MD simulation, the catalytic part of the enzyme approached the DNA at the thymine dimer site, docked into it, and this complex remained stable afterwards (the simulation was performed for 500 ps).

Examining the contact area between DNA and the enzyme it can be observed that there is an intensive interaction arising from the close proximity of Arg-22 and Tyr-21 to C3′, and Arg-26 to C5′ atoms of thymine dimer. In addition, Arg-26 comes very close to the C5′ atom. This specific position of Arg-26 determines the orientation of the enzyme toward the thymine dimer. It can also be noted that Arg-22 and Arg-26 contribute to the recognition and stability of the complex structure. Glu-23, cleaving the *N*-glycosyl bond in the repair process, is located close to the C5′ atom. These observations are in good

agreement with crystal structure, where side chains of Arg-22 and Arg-26 form stacking contact with the DNA [29].

When the control simulation was performed with a non-lesioned DNA molecule, the selected part of the enzyme did not dock into the DNA molecule and the DNA–enzyme complex was not formed.

In this simulation only a relatively small part of the enzyme was used and thus the analysis of results is mainly focused on the description of the structural properties at the contact area and on the possible role of the electrostatic energy in the recognition process, as will be discussed later in the text.

10.5.3 Thymine glycol (5,6-dihydroxy-5,6-dihydro-pyrimidine)

Thymine glycol is observed in DNA after irradiation *in vitro* as well *in vivo* and after oxidation by chemicals (Fig. 10.3). This lesion is known to cause Cockayne syndrome–an inherited disorder in which people are sensitive to sunlight, have short stature and have the appearance of premature aging. It is repaired with endonuclease III glycosylase, which removes a number of damaged pyrimidines from DNA via a its *N*-glycosylase activity and also cleaves the phosphodiester backbone at apurinic/apyrimidinic sites via a ß-elimination mechanism (AP-lyase activity) [30]. The crystal structure of this repair enzyme has been solved to 2.0Å resolution [31]. To study the time evolution of the recognition processes of thymine glycol-lesioned DNA by repair enzyme endonuclease III MD simulation of molecular systems was performed: DNA 30-mer d(CCAGCGCACGACGCA 'TG' GCACGACGACCGGG)$_2$ where TG refers to thymine glycol; and the identical DNA 30-mer complexed with repair enzyme endonuclease III.

Analysis of the results of a 1 ns MD simulation of DNA 30-mer indicates that the double helical structure and hydrogen bonding are well kept through the simulation (except the end base pair of cytosine the C5′–guanine C3′, in which hydrogen bond pairing collapsed after 850ps) [32]. The DNA began to bend at the thymine glycol site after 500 ps of MD and bending continued until simulation was terminated at 1 ns. At the lesion site a kink was observed, a change that relocated thymine glycol closer to the DNA surface. Bending associated with the kink at the lesion site dislocated the phosphodiester bond at C5 atom closer to the DNA surface, enabling it to be eventually processed by the repair enzyme.

In the second system with endonuclease III the simulation lasted for 2 ns [33]. It was also observed that the enzyme was moving toward the surface of the DNA and was also rotating around its axis. These movements caused the surface area surrounding the

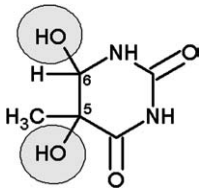


Fig. 10.3. Molecule of thymine glycol–5,6-dihydroxy-5,6-dihydro-pyrimidine (OH groups marked by shadow).

glutamic acid tobe stabilized close to the C3′ and C2′ atoms of TG (carbon atom of the –COO– group at 1.61Å to C3 and at 1.68Å to C2; the situation at the 950th ps upon formation of the complex). It can be estimated that this structural configuration would enable cleavage of the phosphodiester backbone and removal of the damaged nucleotide [34]. Glutamic acid is one of only two amino acids (another one is aspartic acid) that have a net negative charge at physiological pH that is caused by the negative –COO– group making it a very polar molecule.

10.5.4 8-oxoguanine (7,8-dihydro-8-oxoguanine)

8-oxoguanine (8-oxoG) is formed by the oxidation of a guanine base in DNA (Fig. 10.4). It is considered to be one of the major endogenous mutagens contributing broadly to spontaneous cell transformation. Its frequent miss-pairing with adenine during replication increases the number of G–C → T–A transversion mutations. This mutation is among the most common of somatic mutations in human cancers.

In human cells, the 8-oxoguanine is recognized and repaired with specific repair enzyme hOGG1 (human oxoguanine glycosylase 1) that is located in a region on the short arm of chromosome 3 (3p26). Loss of one hOGG1 allele increases the mutagenic burden imposed by guanine oxidation. The biochemical activity of this enzyme has been extensively studied revealing that the enzyme recognizes ′8-oxoG′-cytosine base pairs, catalyzes expulsion of the aberrant base and cleaves the DNA backbone [30, 35]. The crystal structure of the hOGG1 bound to 8-oxoG containing a duplex nucleotide has been reported [36]. In the performed MD simulation of 8-oxoG lesioned DNA 15-mer, d(GCGTCCA′8-oxoG′GTCTACC)$_2$ it was found that this lesion causes a local instability and originates structural deformation by disruption of base stacking in the DNA duplex in the close vicinity of the lesion. This local instability at the lesion site may favor recognition of lesion and aggregation of DNA and enzyme. In the control simulation of undamaged DNA, the B-DNA structure around undamaged guanine at the equivalent position was well kept.

During 1 ns of MD simulation of the system composed of the enzyme hOGG1 and lesioned DNA, the enzyme approached DNA and formed close contacts between several atoms. The direct contact between amino acids and 8-oxoguanine was not established and instead, the aggregation of both molecules was caused by an extensive van der Waals interaction between enzyme and DNA strand at the 8-oxoguanine site. Favorable electrostatic interaction between amino acids and DNA backbone was enhanced by the presence of the lesion. This feature is common for most DNA-binding proteins which

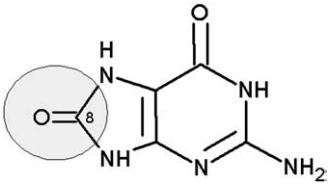


Fig. 10.4. Molecule of 8-oxoguanine (oxidized C8 is marked by shadow).

contact backbone phosphates. There were three regions forming van der Waals contacts of amino acids and nucleotides with respective bases (Fig. 10.5):

(a) Arg-277–guanine G5 (guanine at the C5′ DNA end);
(b) Arg-324–guanine, 8-oxoguanine, adenine;
(c) Ser-54, Glu-53–adenine, cytosine.

Interaction between van der Waals surfaces affects the binding affinity directly and also deters the exclusion of water and counterions from the interface. Figure 10.6 shows a detailed view of the contact region (b) (in Fig. 10.5), with overlapping van der Waals surfaces between respective parts of the sugar–phosphate backbone of the DNA and the amino acids of the enzyme. The van der Waals contacts between the DNA and the enzyme were established after 900ps of MD simulation.

The entire mechanism of how hOGG1 specifically recognizes the 8-oxoG:C pair and binds the 8-oxoG into its active site pocket is not yet clarified. There are several hypotheses based on the crystal as well as theoretical explorations. One suggestion is that the hOGG1 is sliding along the DNA in the search for 8-oxoguanine [37]. An alternative suggestion is that the 8-oxoG can spontaneously flip out [38] of the helix and the hOGG1 will recognize it in this state. The structures of the hOGG1–DNA complex, obtained from X-ray experiments [36,39,40] provide important information on how this DNA repair

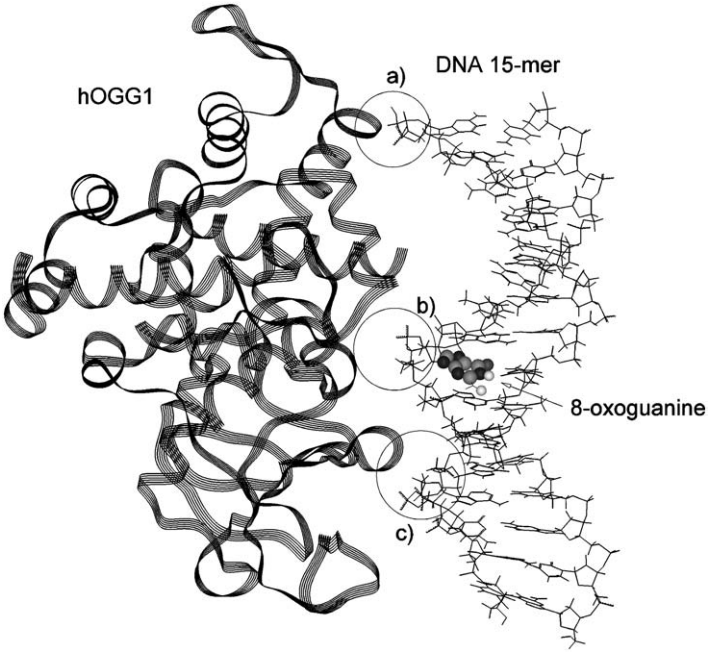


Fig. 10.5. Molecular configuration of the lesioned DNA and the enzyme hOGG1 at 1 ns of MD simulation. The enzyme and the DNA formed close van der Waals contacts between several atoms in three regions (encircled): (a) Arg-277–guanine G5, (b) Arg-324 – guanine, 8-oxoG, adenine, and (c) Ser-54, Glu-53–adenine, cytosine.

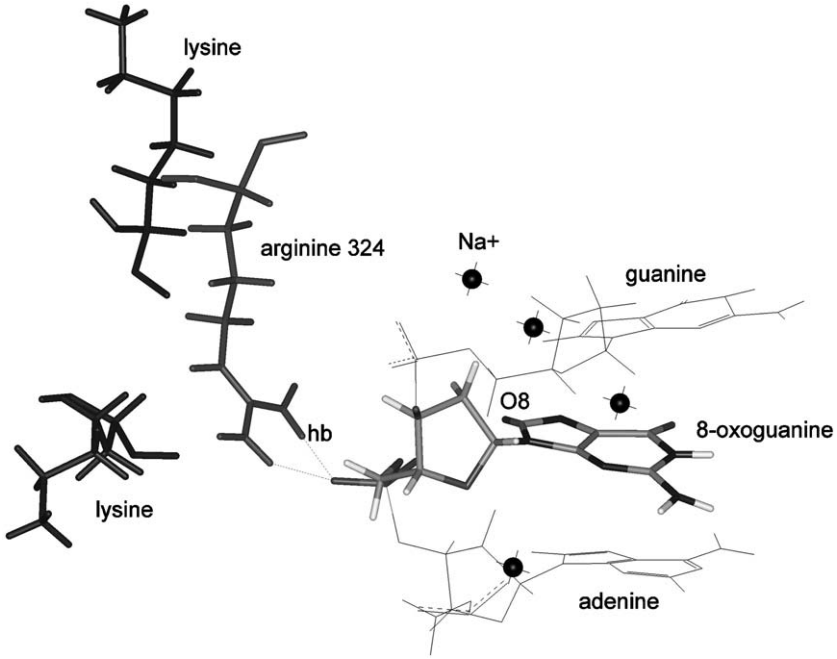


Fig. 10.6. Detailed view of the contact site between Arg-324 and the nucleotide with 8-oxoG. The amino group of Arg-324 is located close to the C59 atom of 8-oxoG, forming two hydrogen bonds between O1P (8-oxoguanine) and N atoms (arginine 324). The attractive electrostatic interaction between $^{+}H_2N$ and the DNA phosphate group contributes to the stability of the complex. (hb – hydrogen bonds between O1P of 8-oxoG and N atoms of Arg-324; Na^{+}-sodium counterions; O8 – oxygen at C8 atom)

protein binds 8-oxoG into the active site. In the presented simulation, the entire docking of hOGG1 into DNA at the 8-oxoG site was not achieved and instead the intensive van der Waals interaction was established. This requires further QM/MM studies on the DNA–enzyme complex.

10.5.5 Role of electrostatic energy in recognition

Even if the structural changes presented in Sections 10.5.1 through 10.5.4 are very important in the formation of the DNA–enzyme complex, the other factors seem to be needed in order to signal to the enzyme the presence of the lesion. One candidate is electrostatic interaction. It has an important contribution to the binding energy of association and is common for many intermolecular interactions (e.g., hydrogen bonds, salt bridges, van der Waals interactions) e.g., [11,12,41]. Thus it is important to consider the electrostatic component of the potential energy when examining the role of electrostatics in the formation of the complex. As has been demonstrated in many studies, e.g., [42–44], electrostatics generally disfavor the docking of ligand and receptor molecules because the unfavorable change in the electrostatics of solvation is mostly compensated for by the favorable electrostatics

within the resulting enzyme–DNA complex. It has also been suggested that molecular recognition is mediated by direct contacts between amino acids of enzyme and DNA bases [45]. The extent to which such contacts function as discrete elements in enzyme–DNA recognition is not known, but a prerequisite for proper recognition is that matching of electrostatic patterns is attained [46]. This is in coincidence with the 'lock-and-key' concept, which emphasizes that an amino acid as an agent is interacting individually and effectively with the lesioned part of the DNA as a target [47]. In the design of an ideal model, one must consider the electrostatic and van der Waals interactions as well as the specific arrangements of hydrogen bonds between enzyme and DNA. The affinity and specificity of an enzyme to DNA is accomplished through the optimization of electrostatic interactions at the binding site.

The important factor in the process of recognition is how the lesion is perceived by the DNA molecule itself, and by atoms and molecules in the surrounding environment. The former is crucial for the overall impact the lesion has on the structural properties and functionality of DNA, the latter plays a role in the recognition of a lesion by a respective repair enzyme and the distribution of charged ions in the vicinity.

On consideration of the factors that caused the fusion on the thymine dimer-lesioned DNA and simulated part of the repair enzyme T4 endonuclease V, the electrostatic interaction energy between the dimer lesion and catalytic center was calculated. It has been found that while the electrostatic energy of the thymine dimer is negative at around –9 kcal/mol, the electrostatic energy of Glu-23 (the closest amino acid to the C5′ atom of the phosphodiester bond of dimer) is around +10 kcal/mol. Since the electrostatic energy of the undamaged thymine is nearly 0 kcal/mol, this specific value of electrostatic energy of the lesion represents a factor discriminating the thymine dimer lesion from the undamaged thymine [10,48] (Fig. 10.7).

In the case of thymine glycol, the intermolecular electrostatic interaction was calculated between enzyme and DNA 30-mer. At the beginning of the MD simulation, energy had a positive value (~ +10 kcal/mol), representing local electrostatic repulsion between enzyme and DNA. After molecules approached each other (interatomic distance ~2 nm), the electrostatic interaction between them became negative (~ –20 kcal/mol), representing electrostatic attraction. Upon formation of the complex (after 950 ps of the MD simulation), the electrostatic interaction changed to a value of around –40 kcal/mol (interatomic distance of the closest atoms was ~1.6 Å), contributing further to the stability of the complex.

The presented study on thymine glycol reveals that the electrostatic energy of a nucleotide with thymine glycol is negative at value of –26 kcal/mol. This negative electrostatic energy is probably one factor contributing to proper recognition of a thymine glycol lesion by an enzyme, and discriminates it from undamaged thymine, which has a value of ~ 0 kcal/mol [10,48].

In the case of the 8-oxoG and hOGG1, the DNA–enzyme complex was formed at the level of van der Waals interactions. Since this form of aggregation was not observed in the system with an undamaged DNA molecule, the electrostatic interaction caused by the presence of 8-oxoguanine is supposed to contribute to the formation of the complex and to the recognition. This electrostatic interaction may further enhance the existing

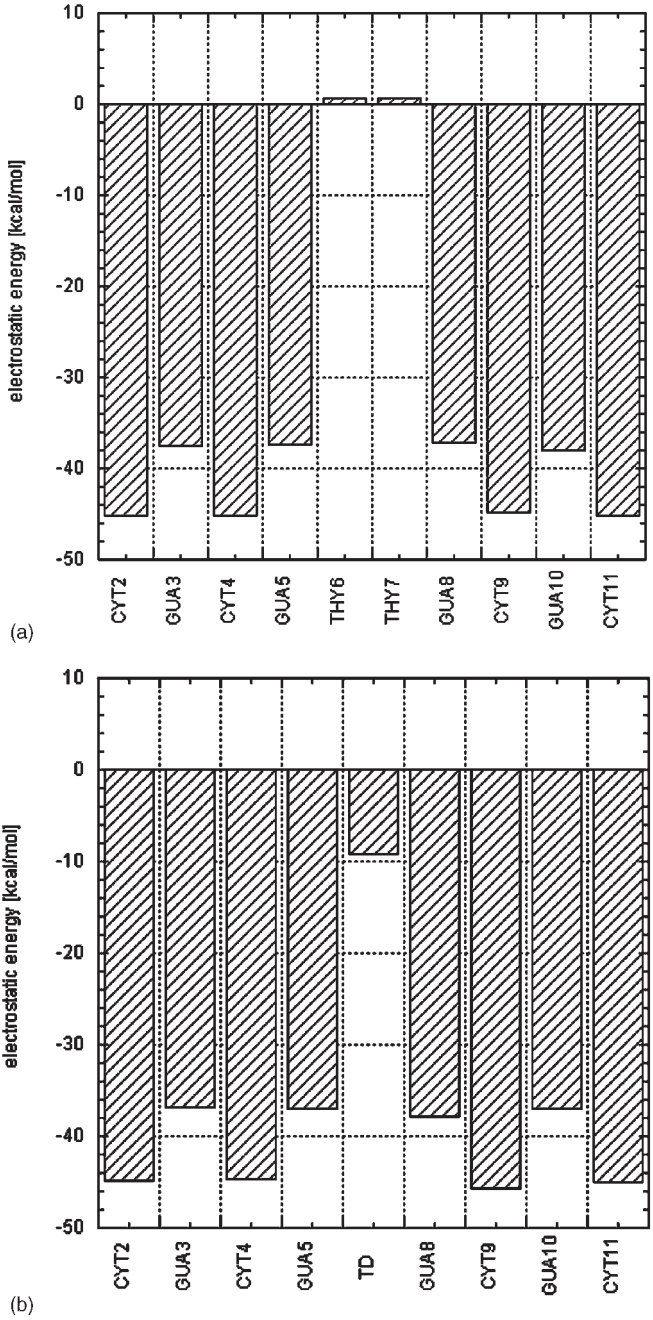


Fig. 10.7. Electrostatic energy of DNA decamer bases; (a) undamaged DNA (b) thymine dimer-lesioned DNA. The electrostatic energy of thymine dimer (TD) is negative at around −9 kcal/mol and that of undamaged thymines is nearly 0 kcal/mol.

van der Waals interaction between both molecules. The amino group of arginine 324 is located close to the C5′ atom of 8-oxoguanine, forming two hydrogen bonds between O1P (8-oxoguanine) and N atoms (Arg-324). The amino group of the arginine is very basic due to its positive charge $+H_2N$. This positive charge establishes an attractive force to the negative charge of the DNA–phosphate group. In addition to the observed van der Waals interaction, the electrostatic attractive interaction between the amino group and the negative charge of the DNA–phosphate group contributes to the stability of the DNA–enzyme complex. The amino group of arginine 324 is also located close to the phosphodiester bond of the nucleotide with 8-oxoguanine that is supposed to be sequentially cleaved during the repair process.

Further analysis of electrostatic energy showed that it is also a significant factor in causing the disruption of DNA base stacking in the DNA duplex. In order to get more detailed information on the interaction of single nucleotides with in the DNA environment and its impact on the stability of the DNA double helix, the electrostatic interaction energies between complementary nucleotides were calculated for 8-oxoguanine-lesioned DNA. Electrostatic energy was calculated as an average value for the stable period of root mean square deviation of the respective simulated system (root mean square deviation (rmsd) represents an average deviation calculated for all heavy atoms of the solute molecule with respect to its structure that was reached after preparatory steps–minimization, stepwise heating and density stabilization). Figure 10.8 shows the average electrostatic interaction energy for each base pair (data

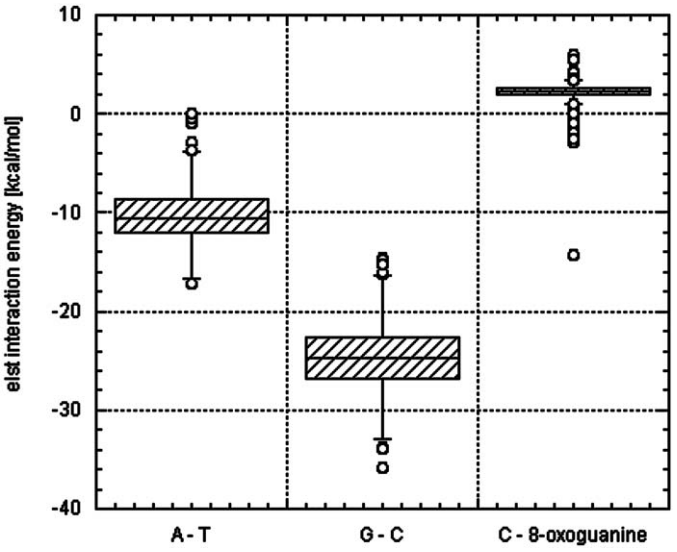


Fig. 10.8. Electrostatic interaction energies between complementary nucleotides. Stable negative energy between undamaged complementary DNA base pairs (A–T and G–C) represents electrostatic attraction. The C–′8-oxoG′ pair has positive electrostatic interaction, energy, (average values for rmsd stable period within 1 ns of the MD simulation) (A–adenine, T–thymine, C–cytosine, G–guanine).

Table 10.1 Average electrostatic energy of nucleotides and selected lesions (the total electrostatic interaction energy including both Coulombic interactions and reaction field interactions from all atoms in the solute using particle mesh ewald method) (average values calculated for all nucleotides of the same type of the simulated DNA segment and for the rmsd stable period of the respective MD simulation)

	electrostatic energy [kcal/mol]		electrostatic energy [kcal/mol]
adenine	−15		
guanine	−37	8-oxoguanine	−48
cytosine	−45	cytosinyl radical	not calculated
thymine	0	thymine dimer	−10
		thymine glycol	−26

averaged for rmsd stable period within ns of MD simulation). The guanine–cytosine and adenine–thymine undamaged pairs were stable, having an average electrostatic energy of ~ −25 kcal/mol and ~ –10 kcal/mol, respectively. This negative energy well represents attractive electrostatic attraction between complementary bases. In the case of 8-oxoguanine-lesioned DNA, the cytosine-8-oxoG pair had positive electrostatic interaction energy (~2 kcal/mol), which indicated local repulsion between complementary strands [49]. Repulsive electrostatic interaction caused by this positive interaction energy may have also contributed to the observed local instability of the lesioned DNA region.

A summary of electrostatic energy of damaged and undamaged nucleotides is shown in Table 10.1.

These findings suggest that a specific value of electrostatic energy at the lesion is recognized by the enzyme during electrostatic scanning of the DNA surface. The specific value is a physically meaningful representation of how the lesion is perceived by the enzyme in its vicinity. It is estimated that specific energy in combination with structural characteristics form a favorable physico-chemical environment for the recognition and for the formation of the stable DNA–enzyme complex. Since the electrostatic energies are usually calculated for a large number of atoms (calculated values are per 1 mol), the real difference at the interaction site may be smaller and its local significance would need further examination by theoretical as well experimental work.

10.6 CONCLUSION

This work comprises the results of MD simulations of several lesions on DNA molecules and their recognition by respective repair enzymes. The subjects of the presented studies were pyrimidine base lesions – cytosinyl radical, thymine dimer, thymine glycol, and purine lesion – 8–oxoguanine. Some of these lesions are considered either to originate neoplastic transformation of the cell (thymine dimer) or are found in cancer cells (8-oxoguanine). The common features observed for all lesions are the specific conformation originated at the lesion site, like disruption of hydrogen bonding networks (cytosinyl

radical, 8-oxoguanine), bending at the lesion site (thymine dimer, thymine glycol), and specific values of the electrostatic interaction energy at all lesions. The bending is considered as an important structural deformation since it forms a complementary shape with respect to the repair enzyme and further facilitates formation of the DNA–enzyme complex. The specific electrostatic energy at the lesion discriminates the lesion from the undamaged DNA part and is considered as contributing to the proper recognition during electrostatic scanning of the DNA surface by the repair enzyme.

The results of the MD simulations, in addition to the existing crystallographic and molecular biological techniques, contribute to the studies of radiation carcinogenesis and DNA radiation damage repair by the dynamic description of the structural and chemical processes that are being undergame at the lesioned DNA molecules. These studies may also contribute to the determination and enhancement of the key factors in the process of recognition of a DNA lesion by a respective repair enzyme in order to satisfy its correct repair.

10.7 ACKNOWLEDGEMENTS

The author wishes to thank Mr. Toshiyuki Nemoto of the Research Organization for Information Science and Technology for the installation simulation codes and for maintenance and adjustment of the computing environment on a FUJITSU VPP 5000 or vector/parallel type, and HITACHI SR 8000 and FUJITSU PRIME POWER parallel supercomputers. The valuable support from all members of the Radiation Risk Analysis Laboratory, Japan Atomic Energy Research Institute is also highly acknowledged.

10.8 APPENDIX

10.8.1 Computational details

The module Sander performs production MD simulation in Amber codes. Since this module requires a long calculation time for completing a simulation for large systems, it has sequential and parallel versions included in the software package. The parallel version is coded by MPI (message passing interface), [http://www.mpi-forum.org/], designed for various parallel computers. This coding can be selected from the original code to match the existing computing environment.

10.8.2 Supercomputers used in simulations

Amber codes were installed and executed on the following three supercomputers (Table 10.2): Fujitsu parallel computer PRIME POWER; Hitachi parallel computer SR 8000; Fujitsu vector/parallel computer VPP 5000.

The PRIME POWER computer has 128 CPUs in one node. It is of SMP (symmetric multi-processing) type, one node has a memory of 250GB with four nodes available. When the MPI program is executed, communication within one node is satisfied via a

Table 10.2 Specifications of PRIME POWER, SR 8000 and VPP 5000

Computer	Type	Max performace /cpu (Gflops)	Clock speed (Mhz)	Memory	CPU/Node	Network speed
PRIME POWER	parallel	2.0	675	250GB/ Node	128cpu	4GB×2
SR 8000	parallel	1.5	375	16GB/ Node	8cpu	1.0GB×2
VPP 5000	vector/ parallel	9.6 (vector unit) 1.2 (scalar unit)	non-public	8GB/PE	1PE	1.6GB×2

Table 10.3 Execution time of parallel processing per 1 ps of simulated time (for a system of 80 000 atoms)

	Number of CPU	1	2	4	8	16	24	32
PRIMEPOWER	Calculation time	1:37:17	0:58:51	0:34:43	0:23:51	0:20:36	0:19:45	0:20:19
	Speed ratio	1.00	1.65	2.80	4.08	4.72	4.92	4.79
SR8000	Calculation time	3:26:04	1:49:33	1:00:08	0:35:35	0:23:36	0:19:16	0:17:37
	Speed ratio	1.00	1.88	3.43	5.79	8.73	10.70	11.69
VPP5000	Calculation time	4:46:01	2:38:55	1:27:54	0:50:58	0:31:58	0:25:13	0:22:51
	Speed ratio	1.00	1.80	3.25	5.61	8.95	11.35	12.51

DTU (data transfer unit) (maximal communication speed 4GB/s × 2). Parallel processing of the Sander module uses 32 CPUs.

The SR 8000 is a scalar computer with a pseudo vector processing facility. This allows data to be fetched from the main memory in a pipelined manner without stalling the succeeding instructions. The SR 8000 has eight CPUs in one node with 20 nodes available in total. one node of parallel processing has 16GB of memory (SMP). Parallel processing of the Sander module uses eight CPUs, which is an optimal number for the SR 8000.

The VPP 5000 is vector/parallel computer. Calculation is done using PE (processing element) with 64 PEs available in total. Each PE contains vector and scalar units with 8GB of memory.

10.8.3 Performance parameters

Parallel MD simulations were executed using various numbers of CPUs using the above computers. The effect of parallel processing is shown in Table 10.3 and in Fig. 10.9.

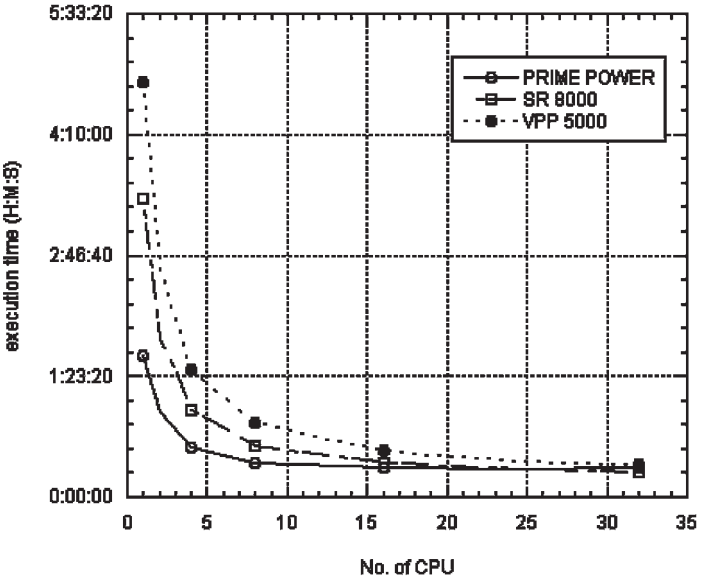


Fig. 10.9. The effect of parallel processing, (for a system of 80 000 atoms).

10.9 REFERENCES

1 B. Halliwell and J.M.C. Gutteridge, Free radicals in biology and medicine, 2nd edition, Clarendon, Oxford, 1989.
2 S. Harrison and A. Aggarwal, Annu. Rev. Biochem., 59 (1990) 933.
3 P.C. Hanawalt, Mutat Res., 400 (1–2) (1998) 117.
4 N.G. Jaspers, Cytokines Mol. Ther., 2 (2) (1996) 115.
5 C.R. Bartram, Eur. J. Pediatr., 135 (2) (1980) 121.
6 UNSCEAR 2000 Report, UN, New York, 2000.
7 D.A.Case, D.A. Pearlman, J.W. Caldwell, T.E. Cheathman III, W.S. Ross, C.L. Simmerling, T.A. Darden, K.M. Merz, R.V. Stanton, A.L. Cheng, J.J. Vincent, M. Crowley, D.M. Ferguson, R.J. Radmer, G.L. Seibel, P.K. Weiner and P.A. Kollman, AMBER 5.0, University of California, San Francisco, 1997.
8 D.A. Case, D.A. Pearlman, J.W. Caldwell, T.E. Cheatham III, J. Wang, W.S. Ross, C.L. Simmerling, T.A. Darden, K.M. Merz, R.V. Stanton, A.L. Cheng, J.J. Vincent, M. Crowley, V. Tsui, H. Gohlke, R.J. Radmer, Y. Duan, J. Pitera, I. Massova, G.L. Seibel, U.C. Singh, P.K. Weiner and P.A. Kollman, AMBER 7, University of California, San Francisco, 2002.
9 M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery Jr, T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez and J.A. Pople, GAUSSIAN 03, Revision C.02, Gaussian, Inc., Wallingford, CT, 2004.
10 M. Pinak, J. Mol. Struct. Theochem., 466 (1999) 219.

11 M. Pinak, Molecular Simulations, 23 (2000) 307.
12 A.S. Yang and B. Honig, Curr. Opin. Struct. Biol., 2 (1992) 40.
13 M.K. Gilson and B.H. Honig, Nature, 330 (1987) 84.
14 T.E. Cheatham and P.A. Kollman, Ann. Rev. Phys. Chem., 51 (2000) 435.
15 V. Makarov, B.M. Pettit and M. Feig, Accounts of Chem. Res., 35 (2002) 376.
16 M. Karplus, Accounts of Chem. Res., 35 (2002) 376.
17 B. Xia, V. Tsui, D.A. Case, H.J. Dyson and P.E. Wright, J. Biol. NMR, 22 (2002) 317.
18 A.R. Bizzari and S. Cannistraro, J. Phys. Chem. B, 106 (2002) 6617.
19 T. Simonson, Curr. Opin. Struct. Biol., 11 (2001) 243.
20 S.A. Hassan and E.L. Mehler, Proteins: Struct. Funct. Genet., 47 (2002) 45.
21 M.S. Lee, F.R. Salsbury and C.L. Brooks, J. Chem. Phys., 116 (2002) 10606.
22 M. Orzo and F.J. Luque, Chem. Rev., 100 (2000) 4187.
23 B. Honig and A. Nicholls, Science, 268 (1995) 1144.
24 P.E. Smith and B.M. Petit, J. Chem. Phys., 105 (1996) 4289.
25 M. Pinak, H. Yamaguchi and R. Osman, J. Radiat. Res., 37 (1996) 20.
26 K. Morikawa, O. Matsumoto, M. Tsujimoto, K. Katayanagi, M. Ariyoshi, T. Doi, M. Ikehara, T. Inaoka and E. Ohtsuka, Science, 256 (1992) 523.
27 T. Doi, A. Recktenwald, Y. Karaki, M. Kikuchi, K. Morikawa, M. Ikehara, T. Inaoka, N. Hori and E. Ohtsuka, Proc. Natl. Acad. Sci. USA, 89 (1992) 420.
28 K. Morikawa, M. Ariyoshi, D.G. Vassylyev, K. Katayanagi, H, Nakamura, T. Doi, N. Hori and E. Ohtsuka, in S. Wallace, B.V. Houten and Y.W. Kow (Eds.), DNA damage, New York Academy of Sciences, New York, 1994, p. 198.
29 D.G. Vassylyev, T. Kashiwanagi, Y. Mikami, M. Ariyoshi, S. Iwai, E. Ohtsuka and K. Morikawa, Cell, 83 (1995) 773.
30 J.A. Tainer, M.M. Thayer, R.P. Cunningham, Curr. Opin. Struct. Biol., 5 (1995) 20.
31 C.F. Kuo, D.E. McRee, C.L. Fisher, S.F. O′Handley, R.P. Cunningham and J.A. Tainer, Science, 258 (1992) 434.
32 M. Pinak, J. Comput. Chem., 22 (15) (2001) 1723.
33 M. Pinak, JAERI-research, 2001-038 (2001).
34 S. Chevillard, J.P. Radicella, C. Levalois, J. Lebeau, M.F. Poupon, S. Oudard, B. Dutrillaux and S. Boiteux, Oncogene, 16 (1998) 3083.
35 C. Dherin, J.P. Radicella, M. Dizdaroglu and S. Boiteus, Nucl. Acid Res., 27 (1999) 4001.
36 S.D. Bruner, D.P. Norman and G.L. Verdine, Nature, 403 (2000) 859.
37 P.H. von Hippel and O.H. Berg, J. Biol. Chem., 264 (1989) 675.
38 G. Panayotou, T. Brown, L.H. Pearl and R. Savva, J. Biol. Chem., 273 (1998) 45.
39 D.P. Norman, S.J. Chung and G.L. Verdine, Biochemistry, 42 (2003) 1564.
40 D.P.G. Norman, S.D. Bruner and G.L. Verdine, J. Am. Chem. Soc., 123 (2001) 359.
41 D.Q. Shortle, Rev. Biophys., 25 (1992) 205.
42 J. Novotny and K. Sharp, Prog. Biophys. Mol. Biol., 58 (1992) 203.
43 J. Novotny, R.E. Bruccoleri, M. Davis and K.A. Sharp, J. Mol. Biol., 268 (1997) 401.
44 J. Shen and J. Wendoloski, J. Comput. Chem., 17 (1996) 350.
45 J.M. Benevides, M.A. Weiss and G.J. Thomas, J. Biol. Chem., 269 (1994) 10869.
46 G. Naray-Szabo and P. Nagy, Enzyme, 36 (1986) 44.
47 S. Rose, J. Theor. Biol., 195 (1998) 111.
48 M. Pinak, J. Mol. Struct., 499 (2000) 57.
49 M. Pinak, Comput. Biol. Chemi., 27 (2003) 431.

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 11

Nucleation of polyglutamine amyloid fibres modelling using molecular dynamics

James Elliott[1], Evgeni B. Starikov[2,3], Jane Crawshaw[1], Peter Claiden[1], Lennart Nilsson[3] and Alan Windle[1]

[1]*Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK*
[2]*Department of Chemistry, Freie Universität Berlin, Germany*
[3]*Department of Structural Biology at NOVUM, Karolinkska Institute, Huddinge, Sweden*

Abstract

It has recently been suggested by Perutz and Windle [1] that the formation of amyloid plaques – insoluble, fibrillar deposits which are implicated in the pathology of a wide range of neurodegenerative disorders – may be related to nucleation of protein aggregates. This scenario is supported by recent research in which aggregation was stimulated by inoculating cells with misfolded proteins [2]. In one particular case, Huntington's disease (HD), it is known that the age of onset of clinical symptoms decreases exponentially with the length of a sequence of glutamine amino acids in the huntingtin (Ht) protein. This observation can be explained simply if the rate-determining step in the process of developing HD is the formation of a stable nucleus for an amyloid fibre. Recently, Perutz *et al*. [3] proposed a new structural model for polyglutamine amyloid fibres, based on X-ray diffraction evidence, in the form of a helically coiled β-sheet. In this paper, we examine the thermodynamic stability of such structures, in comparison with a number of alternative models proposed in the literature, using canonical (constant NVT) and isothermal–isobaric (constant NpT) molecular dynamics simulations of putative amyloid nuclei in an aqueous environment. We analyse the conformational changes in the models as a function of time and glutamine repeat length, and speculate on how the nucleation mechanism might help to explain the pathologies of a range of protein aggregation diseases, such as Huntington's, Alzheimer's, and prion-related disorders.

11.1 INTRODUCTION

11.1.1 Polyglutamine sequences in nature

Poly(L-glutamine), or polyQ, whose structural repeat unit is shown in Fig. 11.1, is a naturally occurring peptide sequence that is attracting increasing attention from materials scientists for two main reasons. First, sufficiently long polyQ sequences in biological proteins are thought to be linked to the cause of several inherited neurodegenerative diseases, including Huntington's disease (HD) [4], spinobulbar muscular atrophy, dentatorubral pallidoluysian atrophy and certain types of spinocerebellar ataxia [5]. Second, the self-assembling properties of pertinent polyQ derivatives have been recognised as providing the potential to form molecular wires to be used in microprocessors and other nanoscale devices (e.g., [6,7]).

In fact, it has recently been suggested that the propensity to form amyloid fibres is a fundamental property of polypeptide chains [8], and that many other amino acid sequences in addition to polyQ can be made to form amyloid fibres. There is also a possible link between the process of amyloid aggregation and prion diseases such as Creutzfeldt–Jacob disease (CJD), scrapie and Bovine Spongiform Encephalopathy (BSE).

Furthermore, similarities have been observed between the way in which molecular chaperone proteins affecting protein folding influence the propagation of both protein-based and inherited structural elements [9].

However, systematic biophysical studies of polyQ are rather scarce, and information on its physical and chemical properties still remains elusive. The main difficulty with the experimental work on polyQ lies in its insolubility in water, which is usually overcome by splicing it into water-soluble peptide frameworks [10–15]. In general, biophysical and chemical data on the resulting systems show that, irrespective of their length, polyQ sequences flanked by sequences of other amino acid residues tend to adopt a random coil conformation in an aqueous environment [11–13], although there is some debate as to the possibility of β-sheet formation under these same conditions [14,15]. Furthermore, longer polyQ sequences, containing approximately 40 or more glutamine residues, exhibit a drastically increased propensity for self-aggregation compared with shorter sequences [13,16], and mediate formation of β-fibrils in the

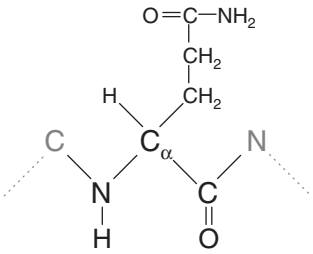


Fig. 11.1. Structural repeat of polyglutamine (polyQ), showing the side group connected to the α-carbon, with connections along the protein backbone between the amino group of one glutamine and the carboxyl group of another.

solid state [17]. An original proposal that polyQ stretches play the role of 'polar zippers' came from the late Max Perutz and his collaborators [18], and seems to be the general principle governing the solid-state structure of polyQ and related systems [19,20].

Although it is not yet clear whether polyQ self-aggregation is the principal pathogenic factor in neurodegenerative disorders, it looks like a physically plausible hypothesis [1] due to the close correlation between polyQ sequence length and the onset of the Huntington's disease. It is clear that long polyQ sequences cause gross conformational changes or misfolding of their protein carriers [21] and in this way trigger a definite succession of pathological processes leading ultimately to biological cell death [22,23]. Thus, the problem of deriving the correct structural model or models for polyQ aggregated sequences is of crucial importance. During the last decade, Max Perutz has addressed this problem and made several proposals which are based on a coiled β-helix. An alternative polyQ structural model was also advocated, consisting of a μ-helix [24] which was then used to rationalize the hypothesis that polyQ forms cation channels [25] that cause depolarization of the cell membrane [26]. The aim of the present article is to critically evaluate the proposed structures by means of molecular modelling. The utility of this approach applied to the polyQ aggregation problem has been previously demonstrated elsewhere by Starikov *et al*. [27].

11.1.2 Huntington's disease

A study by Nancy Wexler and others first established the genetic nature of HD (see [28], and subsequently described in [29]). A mutated gene on chromosome 4, common to all sufferers, was eventually isolated in 1993 [30] and found to be the template for a glutamine rich-protein called huntingtin (Ht). Glutamine repeats in healthy individuals range from 6 to 36, while the repeats range from 36 to 180 in HD sufferers [31,32], as a result of CAG triplet repeat expansions which code for glutamine. Although the precise function of Ht is not yet clear, it is known to be essential during embryonic development as mice in which the Ht gene has been deleted cannot survive [33]. The incorporation of extra glutamine causes Ht to fold incorrectly, and subsequently to aggregate. These aggregates result in highly insoluble fibrils that accumulate in cell tissue of the HD sufferer, with particularly high concentrations being found in neural cells [34]. The age at which the symptoms of HD become noticeable ('onset') has an inverse exponential relation with the number of glutamine repeats in the Ht protein.

It is tempting to look for a direct causal link between glutamine repeats and progress of the disease. However, research shows the importance of nucleation in the aggregation process. Local concentrations of Ht are likely to be a prominent factor because recent studies show that seeding with the aggregated protein accelerates development of the disease. Although no high resolution structural data are available for the precise conformation of the fibres, electron microscopy, X-ray diffraction, nuclear magnetic resonance and a variety of modelling techniques have been used to define their morphology [35,36].

Based on the results obtained from these studies, the possible structures of polyQ sequences are discussed below, in Section 11.1.3.2.

11.1.3 Protein conformations

11.1.3.1 Ramachandran diagrams

A useful tool in the quantitative structural analysis of protein conformations is the Ramachandran diagram (RD), developed by Ramachandran and Sasisekharan [37]. Each amino acid residue in a sequence is represented as a point in two-dimensional space in which the axes correspond to the values of two dihedral angles defined between neighbouring peptide bonds. As shown in Fig. 11.2, the angle ϕ is defined between the planes containing the leading carbon, α-carbon and amine nitrogen, and the amine nitrogen, α-carbon, and carbonyl carbon along the backbone. The angle ψ is defined similarly between the planes defined by the amine nitrogen, α-carbon, and carbonyl carbon, and the α-carbon, carbonyl carbon and trailing nitrogen. All these planes contain the α-carbon, and thus the ordered pair (ϕ,ψ) over the interval $-180° < \phi \leqslant 180°$, $-180° < \psi \leqslant 180°$ uniquely defines the backbone conformation of each residue, with the side group constrained only by its attachment to the α-carbon. A sequence of ordered pairs (ϕ_i,ψ_i) thus defines a protein conformation, which can be represented by a locus of points in the two dimensional RD. Although a third angle ω (not shown) can be defined between the α-carbon, carbonyl carbon and trailing nitrogen and the carbonyl carbon, trailing nitrogen and leading carbon of the next amino acid, it is invariably very close to 180° due to the rigid planar *trans* configuration of the peptide bond. For this reason, and to maintain the simplicity of a two-dimensional representation, ω is usually omitted when quoting Ramachandran angles. However, it should be noted here that the μ-helix structure described in Section 11.1.3.2 requires significant deviations from planarity of the peptide bond.

Ramachandran and Sasisekharan (1968) showed that virtually all naturally occurring proteins reside in certain rather specific regions of the RD, which generally correspond to areas of low steric hindrance and favourable intramolecular hydrogen bonding. In particular, three main regions can be identified, corresponding to two important classes of secondary structure: the right-handed α-helix and the β-sheet. These regions are shown in Fig. 11.3.

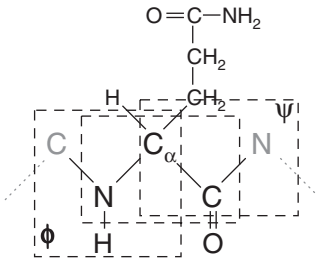


Fig. 11.2. Structural repeat of polyglutamine (polyQ), showing the Ramachandran dihedral angles ϕ and ψ associated with each residue. The atoms in grey belong to neighbouring amino acids.

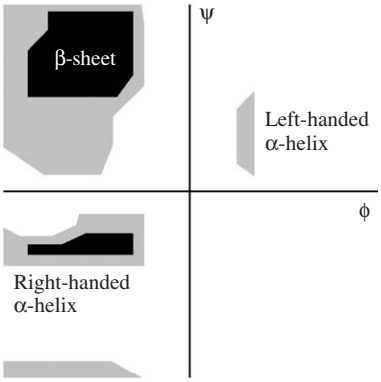


Fig. 11.3. Ramachandran diagram, showing the three main conformational regions in which naturally occurring proteins reside. Darker areas correspond to lower conformational energies. The Ramachandran angles ϕ and ψ are defined in Fig. 11.2.

Of course, not all residues are constrained to reside in the darker low energy regions, but the fact that the majority of them do, regardless of amino acid type, accounts for the widespread occurrence of the α-helix and the β-sheet structural motifs in proteins. Changes in protein conformation can be visualized by examining the dynamic behaviour of points in the RD, and in Section 11.3.2 this method is used to quantify the degree of stability of the polyQ nuclei.

11.1.3.2 Polyglutamine amyloid nuclei

We will now present a selection of the proposed structural models for polyQ amyloid fibres, with their corresponding RDs, with a view to assessing their thermodynamic stability via molecular dynamics. The nomenclature Q_n will be used to denote a polyQ sequence containing n repeat units, the repeat unit having being defined in Fig. 11.1. Figure 11.4(a) shows Q_{20} in the coiled β-sheet conformation proposed by Perutz *et al*. [3], with atomic coordinates kindly supplied by Dr. Arthur Lesk, together with its RD in Fig. 11.4(b). The RD shows that there are only two distinct conformations of Q residues in the Perutz structure: $\phi = -160.7^{\circ}$, $\psi = 166.7^{\circ}$ (upper left quadrant) and $\phi = 159.8^{\circ}$, $\psi = -170.2^{\circ}$ (lower right quadrant). Even-numbered residues are of the latter type, and odd-numbered residues are of the former type, corresponding to the alternating positions of side-groups inside and outside of the helix along the backbone. It can be seen that 20 Q residues comprise one complete turn of the coil, giving a 20_1 helix with a pitch of 4.8 Å. Although helical in form, the Perutz structure has more in common with a β-sheet than an α-helix due to the nature of the hydrogen bonding between the protein backbone and the side groups.

Q_{40}, containing two turns of the coil, is shown in Fig. 11.5(a). The corresponding RD, Fig. 11.5(b) appears superficially identical to Fig. 11.4(b), although there are actually twice the number of points. It is interesting to note at this stage that while the odd-numbered residues fall clearly in the β-sheet region of the RD, the even-numbered residues lie just outside (the RD being periodic) in a higher energy region. Hence, the precise zero

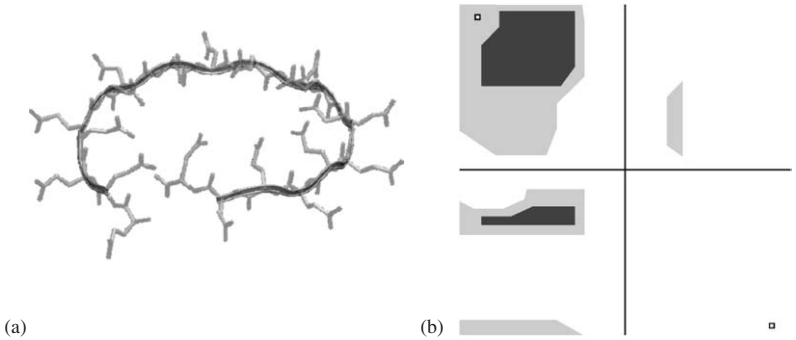


Fig. 11.4. (a) Q_{20} Perutz helix, with (b) corresponding RD. Hydrogen atoms have been omitted from the atomistic structure for clarity, and the polyQ backbone has been highlighted with a dark ribbon. The small squares in (b) denote the exact peptide backbone configuration.

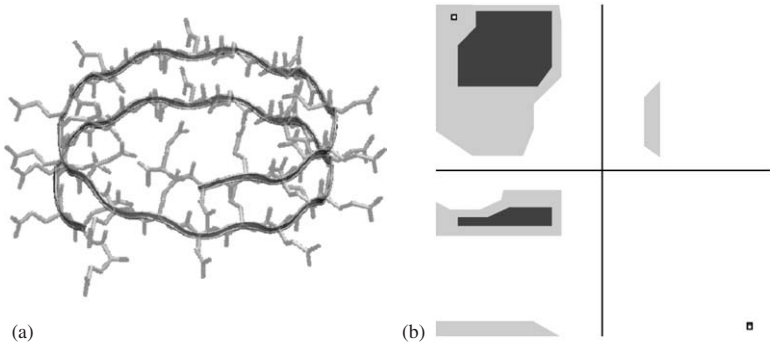


Fig. 11.5. (a) Q_{40} Perutz helix, with (b) corresponding RD. Hydrogen atoms have been omitted from the atomistic structure for clarity, and the polyQ backbone has been highlighted with a dark ribbon. The small squares in (b) denote the exact peptide backbone configuration.

temperature structure originally proposed by Perutz *et al.* [3] is unlikely to be the stable equilibrium conformation in vivo.

Q_{20} in the form of a μ-helix, as proposed by Monoi [24], is shown in Fig. 11.6(a), along with the corresponding RD in Fig. 11.6(b). Each residue has an identical conformation, with principal Ramachandran angles $\phi = 98.0^{\circ}$, $\psi = 81.0^{\circ}$ (upper right quadrant). These conformations lie outside the normal range found in naturally occurring proteins, although they most closely resemble those found in a left-handed α-helix (although the μ-helix shown is right-handed, the most stable enantiomorph according to [24]). As noted in Section 1.3.1, the structure must accommodate significant deviations from planarity of the peptide bond, with $\omega \sim 170^{\circ}$ for each residue. Comparing Figs 11.6(a) and 11.4(a), it is clear that the μ-helix is much more tightly coiled than the Perutz model. It can be seen that 20 Q residues comprise just over three complete turns of the coil, giving a $20_{3,2}$ helix with a pitch of 5.0 Å.

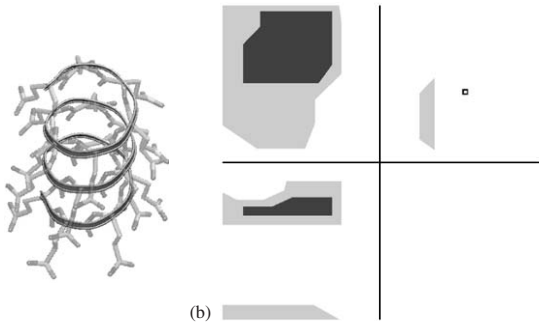


Fig. 11.6. (a) Q_{20} μ-helix, with (b) corresponding RD. Hydrogen atoms have been omitted from the atomistic structure for clarity, and the polyQ backbone has been highlighted with a dark ribbon. The small squares in (b) denote the exact peptide backbone configuration.

11.2 SIMULATION METHODOLOGY

The molecular dynamics simulations were carried out with the software package DL PROTEIN, developed by Melchionna and Cozzini [38] using the standard CHARMm22 force field parameter set described by Brooks *et al*. [39]. Electrostatic forces were calculated using a smooth particle mesh Ewald algorithm, with a 12 Å cut-off (also applied to the van der Waals forces). Initial polyQ structures were generated as described in Section 11.1.3.2 and placed into a cubic periodic cell, of side length 50 Å together with TIP3P water molecules [40] to a give a total density of 1 g cc^{-3}. This gives approximately 4 000 water molecules for each polyQ molecule. The total number of atoms was approximately between 12 000 and 13 000, depending on the length of glutamine and number of waters used to build the models. The equilibration phase of the molecular dynamics consisted of a short (0.2 ps, time step $\tau = 1$ fs) constant volume (NVT) run to remove any atomic overlaps from the model building stage, followed by a longer (20 ps, time step $\tau = 2$ fs) constant pressure (NpT) simulation to equilibrate the cell box density, which rose slightly due to presence of the polyQ. During the equilibration phase, the positions of the backbone atoms in the polyQ (i.e., the atoms contained in the planes which define the Ramachandran angles, see Section 11.1.3.1) were constrained, allowing only motion of the side chains and water. This was in order to permit the side chains and water molecules associated with the protein to relax and equilibrate without disturbing the protein conformation. It necessarily follows that there were no changes in the RD during equilibration. It was found that the dynamics trajectories produced from the models equilibrated in such a fashion were more stable than those started from an energy minimized state.

After equilibration, the production phase of the simulations consisted of an extended (up to 2 ns, time step $\tau = 2$ fs) constant NpT run during which the polyQ molecule was free to relax under conditions of $T = 310$ K, $p = 1$ atm. The resulting trajectories were visualized using the VMD software package from the University of Illinois [41] and the polyQ conformations analysed in terms of dynamic RDs, as described in Section 11.1.3.1. Temperature and pressure were regulated using the Nosé–Hoover method [42], giving fluctuations of the order or ± 3 per cent in these quantities.

The simulated X-ray fibre diffraction patterns shown in Fig. 11.8 were calculated from isolated, infinitely long helices using the Cerius^2 software package from Accelrys Inc. [43]. Cylindrically-averaged structure factors were calculated from the model structures, with the fibre axis running parallel to the meridian on the diffraction patterns. A broadening factor for the finite crystallite size of 500\AA in each direction was used, along with an orientation half-width of 15 degrees to simulate the effects of misalignment to the fibre axis.

11.3 RESULTS

We begin by discussing the relative stability of the μ-helix, Perutz helix and other structures for polyQ nuclei in Section 11.1.3.2, using sequence lengths Q_{20} (below the critical limit for aggregation *in vivo*) and Q_{40} (above the critical limit). We then discuss the stability of the Perutz helix model as a function of glutamine repeat length in more detail using Ramachandran analysis.

11.3.1 Relative stability of different polyQ nuclei

During the production phase described in Section 11.2, the root-mean-square deviations (RMSD) averaged over all atoms of the polyQ nuclei with respect to their initial equilibrated structures were plotted as a function of time in Fig. 11.7 for Q_{20} and Q_{40} sequence lengths in the μ-helix and Perutz helix conformations. It can be seen that the RMSD saturation levels for Q_{20} Perutz helix and Q_{40} μ-helix are between 8 and 9 Å, whereas those for the Q_{40} Perutz helix and Q_{20} μ-helix are approximately half. This would imply that the Perutz helix structure is increasingly stabilized as the glutamine repeat length becomes longer, whereas the μ-helix structure is destabilized. Therefore, only the Perutz model is consistent with the hypothesis that intracellular aggregation of polyQ is due to glutamine repeat length expansion. It would seem that the μ-helix is not sufficiently stable to persist for extended periods of time. A number of other structural models for the polyQ nuclei, including planar β-sheets and amorphous aggregates, were evaluated using the same methods, and none was able to demonstrate the systematic increase in stability with glutamine repeat length shown by the Perutz helix.

Since the Perutz helix was originally derived from consideration of X-ray diffraction data [3], we have calculated the X-ray fibre patterns, averaged parallel to the helical axis, from the model as presented in the original paper in Fig. 11.8(a), and 1 ps after the start of the MD production run described in Section 11.2 in Fig. 11.8(b). Both fibre patterns show strong, sharp reflections on the meridian at around 4.8 Å and 8.40 Å, and strong, broad reflections on the equator at 9.31 Å, respectively, which are characteristic of the cross-β-pattern [44] commonly obtained from amyloid structures. The equatorial reflections at 3.75 Å and 3.16 Å in Fig. 11.8(a) correspond to the repeats of backbone amino acids and sidechain amides, respectively, in the original Perutz model. However, these appear much weaker as the model is allowed to relax in an aqueous environment, Fig. 11.8(b). Nevertheless, both patterns appear broadly consistent with published experimental patterns from synthetic polyQ and Ht amyloid structures [45].

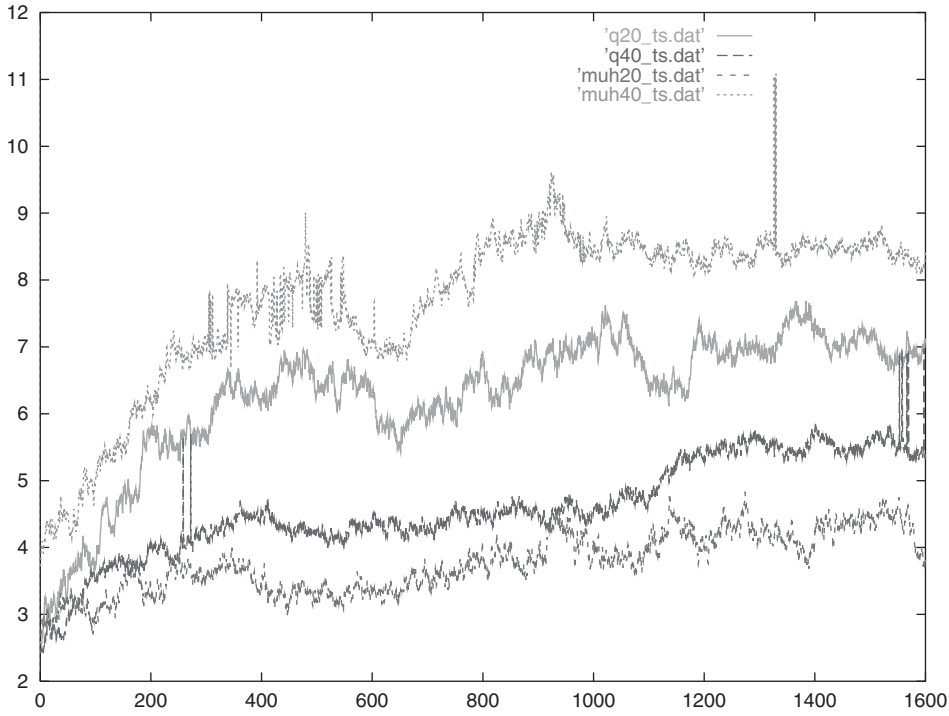


Fig. 11.7. Root-mean-square deviation (RMSD) plot for Q_{20} and Q_{40} in the μ-helix and Perutz helix conformations, showing RMSD between initial structures described in Section 11.1.3.2 and those produced in the MD trajectory as a function of simulation time (see Color Plate 19).

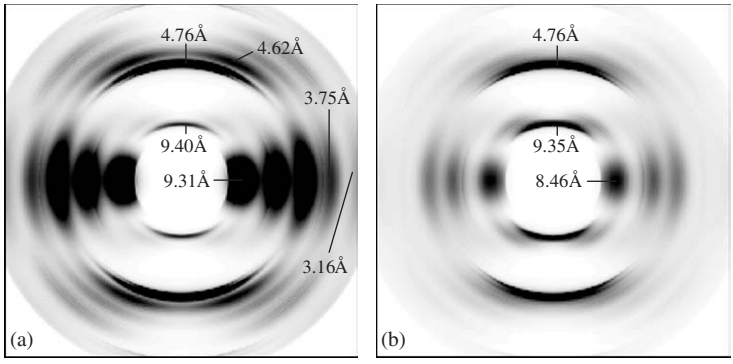


Fig. 11.8. Simulated X-ray fibre diffraction patterns of a isolated, infinitely long Perutz helix structure: (a) model as presented in [3] and (b) 1 ps after the start of MD production run.

11.3.2 Stability of Perutz model with glutamine repeat length

The Perutz model was established in the previous section as being a more viable model for a putative polyQ nucleus than the μ-helix, as its stability increased with the number of glutamine repeats. However, the question remains as to exactly how the lifetime of the nucleus varies with the number of glutamine repeats, and whether there is an abrupt transition at a certain critical value. In order to investigate this in more detail, MD simulations were run on Q_{30} and Q_{35} oligomers, corresponding to one-and-a-half and one-and-three-quarter turns of the Perutz helix, respectively, in addition to Q_{20} and Q_{40} results presented in the previous section. Figure 11.9 shows the backbone configurations of these structures, in which all the atoms have been hidden for clarity, at time $t = 1$ ps after the start of the production phase. It can be clearly seen that all four oligomers are stable over this time period, and this is demonstrated by their corresponding RDs, shown in Fig. 11.10.

Common to all RDs in Fig. 11.10 is a characteristic diagonal trend $\phi \sim -\psi$ in the upper left quadrant of the RD, which is a signature of the Perutz helix. However, compared with Figs 11.4(b) or 11.5(b), there is already a considerable spread in the positions of the residues in the RD, which seems to increase with the number of glutamine repeats (in fact, the mean-squared deviation of the Q_{40} residues from the diagonal $\phi \sim -\psi$ is lower than for Q_{35} and Q_{30}).

In particular, there are no residues left in the lower right quadrant, with the majority occurring in the upper left β-sheet region. This would seem to indicate that all Q_n Perutz helices have relaxed to a common structure, which is close to that originally proposed by Perutz *et al*. [45].

However, we have found that the stability of the polyQ nuclei with time is not independent of repeat length. Figure 11.11 shows the backbone configurations for the Q_n Perutz helices at time $t = 600$ ps after the start of the production phase. Q_{20}, shown on the upper left, has clearly unravelled to a large extent, whereas Q_{30} and Q_{35} retain vestigial loops of the original helix. Only Q_{40} has preserved the motif of the Perutz helix

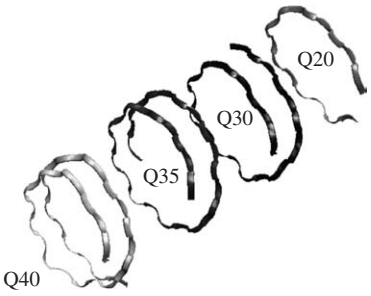


Fig. 11.9. Backbone configurations for Q_n Perutz helices at time $t = 1$ ps after the start of the production phase, shown for Q_{20}, Q_{30}, Q_{35} and Q_{40}. The backbone has been highlighted by a ribbon, with atomistic detail and solvent suppressed for clarity. The relative arrangement of the Q_n molecules is for the purposes of display only; each molecule was simulated in isolation. Corresponding Ramachandran diagrams are given in Fig. 11.10.

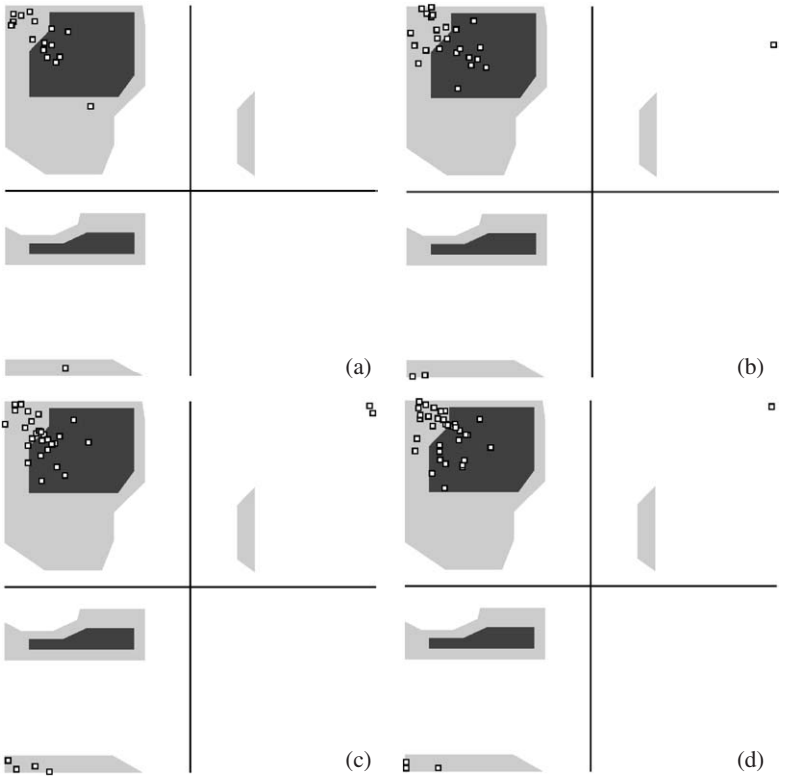


Fig. 11.10. Ramachandran diagrams for Q_n Perutz helices at time $t = 1$ ps after the start of the production phase: (a) Q_{20}, (b) Q_{30}, (c) Q_{35} and (d) Q_{40}. The corresponding backbone configurations are shown in Fig. 11.9.

model, albeit containing some considerable degree of disorder. Despite this internal rearrangement, the RDs in Fig. 11.12 show that Q_{40} has approximately 82 per cent of its residues in the β-sheet region, with only 18 per cent in the lower left quadrant.

Meanwhile, Q_{20}, Q_{30} and Q_{35} each have significantly higher proportions of residues in the latter region, corresponding to structural defects of a right-handed α-helix type. The Q_{40} RD also preserves some of the diagonal character of the equilibrated Perutz helix shown in Fig. 11.10.

Whilst it is difficult to construct a single parameter for characterizing the stability of the Perutz helix as a function of glutamine repeat length, it is clear from Figs 11.11 and 11.12 that the trend is one of increasing stability. A tentative attempt to quantify the degree of disorder was made by calculating the mean-squared deviation of the Ramachandran angles from the diagonal defined by $\phi = -\psi$ (zero for a perfect Perutz helix) for the RDs shown in Figs 11.10 and 11.12. The results showed that, after 600 ps, Q_{40} possessed the lowest overall mean-squared deviation from the Perutz helix structure of all the sequences considered. Furthermore, shorter glutamine sequences showed greater percentage increases in mean-squared deviation over the production period. Of course, this measure does not take into account the relative ordering of the residues giving rise to the

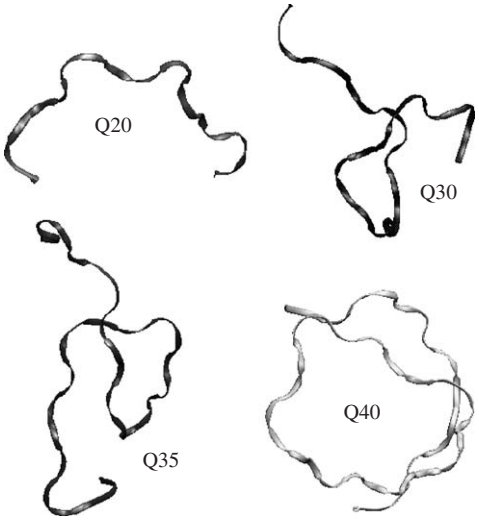


Fig. 11.11. Backbone configurations for Q_n Perutz helices at time $t = 600$ ps after the start of the production phase, shown for Q_{20}, Q_{30}, Q_{35} and Q_{40}. The backbone has been highlighted by a ribbon, with atomistic detail and solvent suppressed for clarity. The relative arrangement of the Q_n molecules is for the purposes of display only; each molecule was simulated in isolation. Corresponding Ramachandran diagrams are given in Fig. 11.12.

Ramachandran angles along the backbone, but nevertheless the results described in this section lend considerable weight to the hypothesis that the Perutz helix structure for polyQ is increasingly stabilized by the addition of glutamine units. There was no evidence for an abrup transition in stability over a time period of 1 ns, although a direct measurement of lifetime as a function of the number of glutamine repeats was not attempted.

11.4 DISCUSSION AND CONCLUSIONS

The results so far show that the Perutz helix, as opposed to a number of other putative nuclei structures, demonstrates increased thermodynamic stability of polyQ sequences as the number of glutamine repeats is increased. The implications of this fact for the pathology of Huntington's disease is that the formation of such stable nuclei in the polyQ portion of the Ht protein would lead to rapid growth of amyloid structures and the onset of symptoms. Clearly, direct computer simulation of nucleus formation is impossible, due to the extreme unlikelihood of such an event (which requires a timescale of the order of one human lifetime for this event to occur), and so we must appeal to indirect evidence from the measured stability of fully formed nuclei. A possible mechanism for the aggregation process is illustrated schematically in Fig. 11.13, which shows the polyQ sequences attached to the Ht protein nucleating and coming together to form a protofibril, which then aggregates further to form a superhelical fibre. Since the nucleation step is rate-determining, the onset of pathological symptoms due to amyloid formation is linked directly to the polyQ sequence length. The advantage of the Perutz helix structure is that

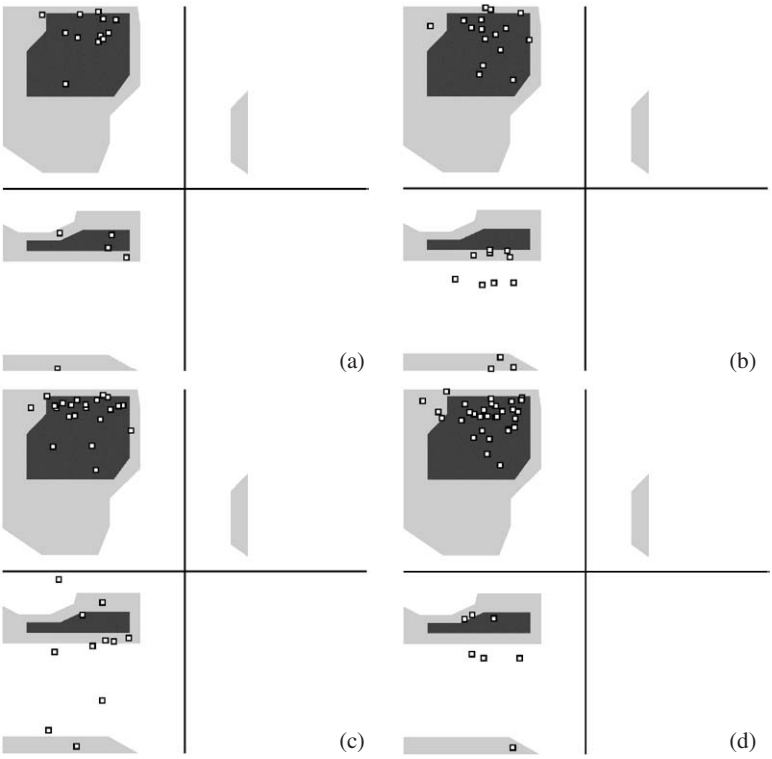


Fig. 11.12. Ramachandran diagrams for Q_n Perutz helices at time t = 600 ps after the start of the production phase: (a) Q_{20}, (b) Q_{30}, (c) Q_{35} and (d) Q_{40}. The corresponding backbone configurations are shown in Fig. 11.11.

the globular non-polyQ portion of the Ht protein is free to hang off the side of the main helix, which is only slightly perturbed by its presence. The adoption of superhelical structures could help to minimize the steric interactions between globules attached to adjacent polyQ sequences, and evidence for such configurations containing two or three protofibrils can be seen in electron micrographs from aggregated polyQ sequences [3].

Such a picture of the Ht amyloid formation process may still be considered highly speculative even in the light of the results contained in this paper. However, there is growing experimental evidence to support the idea that polyQ aggregation is a nucleated growth polymerization reaction. In particular, Chen *et al*. [46] have recently shown that, based on the *in vitro* aggregation kinetics of polyQ peptides, the critical nucleus size (i.e., the number of monomeric units comprising the nucleus) is equal to one. They used parameters from a mathematical model of the aggregation kinetics to determine the difference in free energy between the nucleation of benign and pathological repeat lengths of polyQ, and found that the predicted lag times corresponding to the observed age of onset of symptoms were in very good agreement with clinical data. Their picture of the aggregation process (see Fig. 5 in [46]) is very similar indeed to our Fig. 11.13, although they were not able to detect or structurally characterize the nuclei due to the

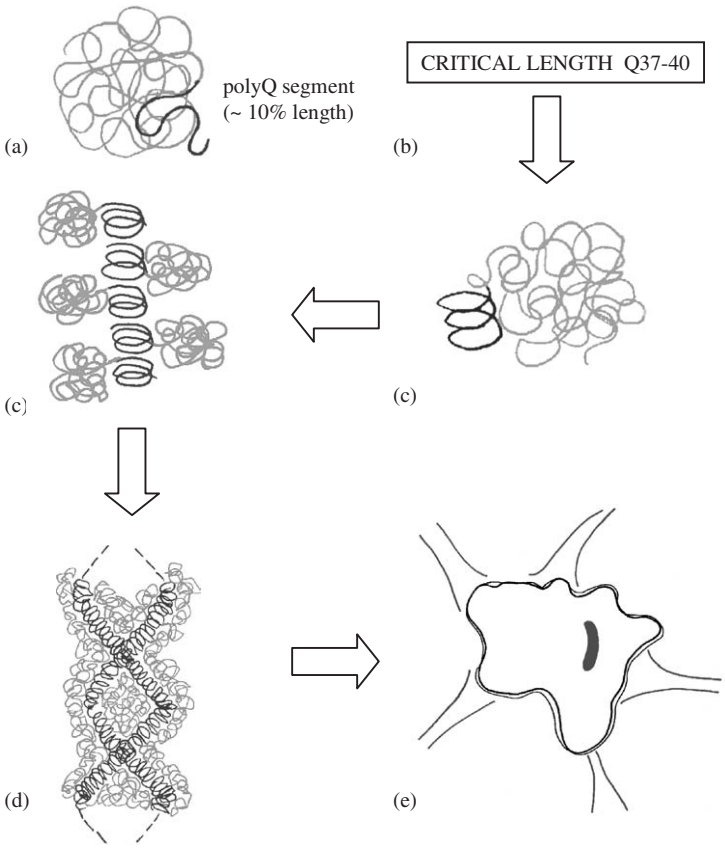


Fig. 11.13. Schematic illustration of a possible mechanism for the formation of Ht amyloid fibres: (a) Ht protein containing polyQ (b) formation of stable nucleus by polyQ, which is the rate-determining step in the (c) aggregation of stable nuclei to form a protofibril, which can in turn form (d) pleated superhelices that precipitate as (e) amyloid plaques in neural cells.

very low apparent concentration of monomer in solution and the speed with which it either becomes incorporated into a protofibril or reverts to the bulk phase structure.

This brings us finally to the wider implications of the work, which relate particularly to the autocatalytic nature of amyloid fibril formation. Since the formation of individual stable nuclei is such a slow process, the aggregation step shown in Fig. 11.13(c) is very unlikely to occur by the chance coming together of independent nuclei. Rather, the process would involve the molecular templating of bulk phase polyQ into the helical motif of the protofibril. As Chen *et al.* [46] remark, this is very similar to the dock-and-lock mechanism postulated for Alzheimer β-fibril extension [47]. Further investigation into the mechanisms of molecular self-organization and assembly by computer simulation and other techniques will be vitally important in order to elucidate the driving forces behind these complex transformations. The degree to which such knowledge obtained from basic

biophysical studies carried out *in vitro* and *in silico* can be applied to real biological systems *in vivo*, in which there are known to be many complicating factors, will determine their utility in deriving potential therapeutic treatments. For instance, it could be envisaged that there would be certain amino acid sequences that could bind to a nascent polyQ protofibril and prevent further growth. Also, there may be many, as yet unknown, biological regulatory systems in which structural information is transmitted directly from protein to protein, rather than being encoded by the DNA.

11.5 ACKNOWLEDGEMENTS

JAE would like to acknowledge the Nuffield Foundation (NUF-NAL 02) for financial support and Dr Arthur Lesk for providing atomic coordinates of the Perutz helix structure described in Section 11.1.3.2. The authors would like to acknowledge stimulating discussions with the late Max F. Perutz, whose ideas inspired the work described in this paper.

11.6 REFERENCES

1 M.F. Perutz and A.H. Windle Nature, 412 (2001) 143–144.
2 Y. Narain, A. Wyttenbach, J. Rankin, R.A. Furlong and D.C. Rubinsztein, J. Med. Genet., 36 (1999) 739–746.
3 M.F. Perutz, J.T. Finch, J. Berriman and A. Lesk, Proc. Natl. Acad Sci., 99 (2002) 5591–5595.
4 J.F. Gusella and M. E. MacDonald, Nature Reviews Neuroscience, 1 (2000) 109–115.
5 H.L. Paulson, and K.H. Fischbeck, Annu. Rev. Neurosci., 19 (1996) 79–107.
6 O. Shoji, M. Okumura, H. Kuwata, T. Sumida, R. Kato, M. Annaka, M. Yoshikuni and T. Nakahira, Macromolecules, 34 (2001) 4270–4276.
7 O. Shoji, Y. Higashi, S. Hishinuma, M. Sato, M. Annaka, M. Yoshikuni, and T. Nakahira, Macromolecules, 35 (2002) 2116–2121.
8 C.E. MacPhee and C.M. Dobson, JACS, 122 (2000) 12707–12713.
9 S. Krobitsch and S. Lindquist, PNAS, 97 (2000) 1589–1594.
10 E.E. Wanker, Bio. Chem., 381 (2000) 937–942.
11 D. Sharma, S. Sharma S. Pasha and S.K. Brahmachari, FEBS Lett., 456 (1999) 181–185.
12 S. Chen, V. Berthelier, W. Yang and R. Wetzel, J. Molec. Bio., 311 (2001) 173–182.
13 L. Masino, G. Kelly, K. Leonard, Y. Trottier and A. Pastore, FEBS Lett., 513 (2002) 267–272.
14 E.L. Altschuler, N.V. Hud, J.A. Mazrimas and B. Rupp, FEBS Lett., 472 (2000) 166–167.
15 S.K. Brahmachari, D. Sharma, S. Sharma, S. Pasha, S. Sen and Q. Saleem, FEBS Lett., 472 (2000) 167–168.
16 Y. Georgalis, E.B. Starikov, B. Hollenbach, R. Lurz, E. Scherzinger, W. Saenger, H. Lehrach and E.E. Wanker, Proc. Natl. Acad. Sci., 95 (1998) 6118–6121.
17 A.E. Bevivino and P.J. Loll, Proc. Natl. Acad. Sci., 98 (2001) 11955–11960.
18 M.F. Perutz Trends Biochem. Sci., 24 (1999) 58–63.
19 J.A. Subirana and J. Palau, FEBS Lett., 448 (1999) 1–3.
20 M. Balbirnie, R. Grothe and D.S. Eisenberg, Proc. Natl. Acad. Sci., 98 (2001) 2375–2380.
21 C. Soto, FEBS Lett., 498 (2001) 204–207.
22 J.S. Steffan, L. Bodai, J. Pallos, M. Poelman, A. McCampbell, B.L. Apostol, A. Kazantsev, E. Schmidt, Y.Z. Zhu, M. Greenwald, R. Kurokawa, D.E. Housman, G.R. Jackson, J.L. Marsh and L.M. Thompson, Nature, 413 (2001) 739–743.
23 A. McCampbell, A.A. Taye, L. Whitty, E. Penney, J.S. Steffan and K.H. Fischbeck, Proc. Natl. Acad. Sci., 98 (2001) 15179–15184

24 H. Monoi, Biophys. J., 69 (1995) 1130–1141.
25 H. Monoi, S. Futaki, S. Kugimiya, H. Minakata and K. Yoshihara, Biophys. J., 78 (2000) 2892–2899.
26 J.D. Lear, Biophys. J., 78 (2000) 2733–2734.
27 E.B. Starikov, H. Lehrach and E.E. Wanker, J. Biomol. Struc. Dyn., 17 (1999) 409.
28 J.F. Gusella, N.S. Wexler and P.M. Conneally, Nature, 306 (1983) 234–238.
29 N.S. Wexler, Clairvoyance and caution: repercussions from the Human Genome Project, in D.J. Kevles and L. Hood, The Code of Codes: Scientific and Social Issues in the Human Genome Project, Harvard University Press, Cambridge, 1992, 211–243.
30 M.E. MacDonald, C.M. Ambrose, M.P. Duyao, R.H. Myers, C. Lin, L. Srinidhi, G. Barnes, S.A. Taylor, M. James, N. Groot, H. Macfarlane, B. Jenkins, M.A. Anderson, N.S. Wexler, J.F. Gusella, G.P. Bates, S. Baxendale, H. Hummerich, S. Kirby, M. North, S. Youngman, R. Mott, G. Zehetner, Z. Sedlacek, A. Poustka, A.M. Frischauf, H. Lehrach, A.J. Buckler, D. Church, L. Doucettestamm, M.C. Odonovan, L. Ribaramirez, M. Shah, V.P. Stanton, S.A. Strobel, K.M. Draths, J.L. Wales, P. Dervan, D.E. Housman, M. Altherr, R. Shiang, L. Thompson, T. Fielder, J.J. Wasmuth, D. Tagle, J. Valdes, L. Elmer, M. Allard, L. Castilla, M. Swaroop, K. Blanchard, F.S. Collins, R. Snell, T. Holloway, K. Gillespie, N. Datson, D. Shaw and P.S. Harper, Cell, 72 (1993) 971–983.
31 D.C. Rubinsztein, J. Leggo, R. Coles, E. Almqvist, V. Biancalana, J.J. Cassiman, K. Chotai, M. Connarty, D. Craufurd, A. Curtis, D. Curtis, M.J. Davidson, A.M. Differ, C. Dode, A. Dodge, M. Frontali, N.G. Ranen, O.C. Stine, M. Sherr, M.H. Abbott, M.L. Franz, C.A. Graham, P.S. Harper, J.C. Hedreen, A. Jackson, J.C. Kaplan, M. Losekoot, J.C. MacMillan, P. Morrison, Y. Trottier, A. Novelletto, S.A. Simpson, J. Theilmann, J.L. Whittaker, S.E. Folstein, C.A. Ross and M.R. Hayden, Am. J. Hum. Genet., 59 (1996) 16–22.
32 K. Sathasivam, I. Amaechi, L. Mangiarini and G. Bates, Hum. Genet., 99 (1997) 692–695.
33 M.P. Duyao, A.B. Auerbach, A. Ryan, F. Persichetti, G.T. Barnes, S.M. McNeil, P. Ge, J.P. Vonsattel, J.F. Gusella, A.L. Joyner and M.E. MacDonald, Science, 269 (1995) 407–410.
34 M. DiFiglia, E. Sapp, K.O. Chase, S.W. Davies, G.P. Bates, J.P. Vonsattel and N. Aronin, Science, 277 (1997) 1990–1993.
35 J.L. Jimenez, J.L. Guijarro, E. Orlova, J. Zurdo, C. M. Dobson, M. Sunde and H.R. Saibil, EMBO J., 18 (1999) 815–821.
36 M. Sunde and C. Blake, Adv. Prot. Chem., 50 (1997) 123–159.
37 G.N. Ramachandran and V. Sasisekharan, Adv. Prot. Chem., 23 (1968) 283–437.
38 S. Melchionna and S. Cozzini, DL PROTEIN molecular dynamics code, Supported by CCP5 of the EPSRC, Daresbury Laboratory, UK, 1998.
39 B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan and M. Karplus, J. Comput. Chem., 4(1983) 187–217.
40 W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R. W. Impey and M.L. Klein, J. Chem. Phys., 79 (1983) 926–935.
41 W. Humphrey, A. Dalke and K. Schulten, J. Molec. Graphics, 14 (1996) 33–38.
42 S. Melchionna, G. Ciccotti and B.L. Holian, Molec. Phys., 78 (1993) 533–544.
43 Accelrys Inc., Cerius2 Modeling Environment, Release 4.6, Accelrys Inc., San Diego, 1999.
44 W.T. Astbury, S. Dickinson and K. Bailey, J. Biochem., 10 (1935) 2354–2365.
45 M.F. Perutz, B.J. Pope, D. Owen, E.E. Wanker and E. Scherzinger, Proc. Natl. Acad. Sci., 99 (2002) 5596–5600.
46 S. Chen, F.A. Ferrone and R. Wetzel Proc. Natl. Acad. Sci., 99 (2002) 11884–11889.
47 W.P. Esler, E.R. Stimson, J.M. Jennings, H.V. Vinters, J.R. Ghilardi, J.P. Lee, P.W. Mantyh and J.E. Maggio, Biochemistry, 39 (2000) 6288–6295.

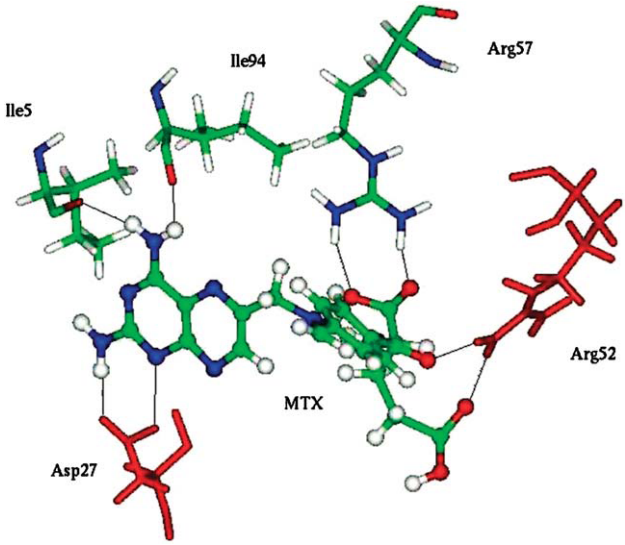


Plate 18. The inhibitor MTX in docked position inside the DHFR receptor. The black lines indicate hydrogen bonds, the red coloured residues were replaced with valine during the second screen (illustration appears on page 181 of this volume).

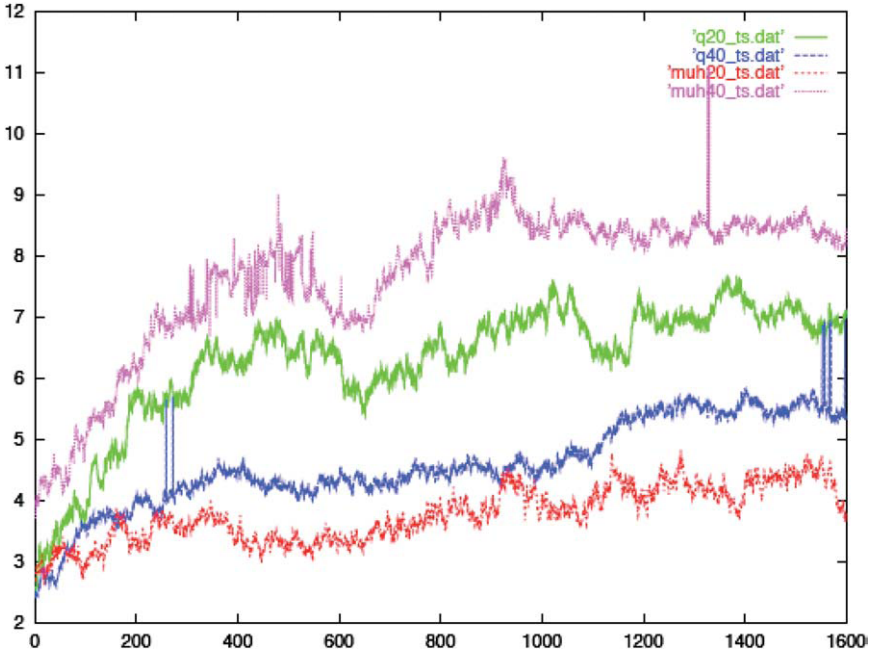


Plate 19. Root-mean-square deviation (Rmsd) plot for Q20 and Q40 in the μ-helix and Perutz helix conformations, showing RMSD between initial structures described in Section 11.1.3.2 and those produced in the MD trajectory as a function of simulation time (illustration appears on page 219 of this volume).

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 12

Drug discovery using grid technology

Hitoshi Goto[1], Sigeaki Obata[1], Toshiyuki Kamakura[1], Naofumi Nakayama[1], Mitsuhisa Sato[2], Yoshihiro Nakajima[2], Umpei Nagashima[3], Toshio Watanabe[3], Yuichi Inadomi[3], Masakatsu Ito[3], Takeshi Nishikawa[3], Tatsuya Nakano[4], Lennart Nilsson[5], Shigenori Tanaka[6], Kaori Fukuzawa[7], Yuichiro Inagaki[7], Michiaki Hamada[7] and Hiroshi Chuman[8]

[1]*Department of Knowledge-based Information Engineering, Toyohashi University of Technology, Toyohashi, 441-8580, Japan*
[2]*Center for Computational Sciences, University of Tsukuba, Tsukuba, 305-8577, Japan*
[3]*Grid Technology Research Center, National Institute of Advanced Industrial Science and Technology, Tsukuba, 305-8568, Japan*
[4]*National Institute of Health Sciences, Tokyo 158-8501, Japan*
[5]*Department of Biosciences at NOVUM, Center for Structural Biochemistry, Karolinska Institutet, S-141 57 Huddinge, Sweden*
[6]*Graduate School of Science and Technology, Kobe University, Kobe 657-8501, Japan*
[7]*Mizuho Information and Research Institute, Inc., Tokyo, 101-8443, Japan*
[8]*Institute of Health Biosciences, University of Tokushima, Tokushima, 770-8505, Japan*

Abstract

Demands for molecular calculations and simulations of huge biomolecular systems such as a drug-receptor complex have been increasing in structural biology and pharmaceutical sciences. A number of computer resources, such as CPUs and storages, can be connected over networks to construct a huge virtual computing environment using grid technology. Our project "Drug Discovery on Grid" aims at developing a platform for drug discovery with the Grid technology. Several types of molecular calculation/simulation for the electronic and conformational structures of huge biomolecular systems can be executed efficiently on the platform. We have also developed Drug Markup Language (DrugML) by extending Chemical Markup Language for storage, reuse and integration of the calculated results. The three-dimensional quantitative structure-activity relationship analyses can be performed based on the DrugML database. "Drug Discovery on Grid" is now a powerful approach in the computer aided drug discovery and is useful for a deeper understanding of the biomolecular phenomena.

12.1 INTRODUCTION

The challenge of the quantitative prediction of the biological potency of a new drug molecule has not yet been met. However, in recent years there have been major advances in our understanding of ligand–receptor interactions, molecular similarity, pharmacophores and macromolecular structures by use of modern molecular calculations and simulations. With the recent progress of technology such as combinatorial chemistry and high throughput screening, and the rapid increase of genomic information, development of a practical and integrated platform for extracting useful information from these results for drug discovery has been increasingly expected. We have developed a novel platform system for rational drug design based on a high performance computing environment. The huge number of molecules poses a serious challenge in executing the molecular computations. To overcome this problem, we introduced grid technology to a workbench of parallel software libraries and database systems specialized for drug discovery. A number of computer resources, such as CPUs and storages, can be connected over networks to construct a huge virtual computing environment using grid technology. Such attempts other than ours have started at nearly the same time; United Device Inc. in the USA have started to provide a network computing service for drug discovery with the grid solution [1]. Not for drug discovery, but Combechem [2] is known to be one of the UK e-Science projects and has been working on grid-enabled combinatorial chemistry, concentrating on crystallography, and laser and surface chemistry.

Our 'grid drug discovery' project is aimed at developing a platform for drug discovery, on which various calculations and simulations are effectively executed at high speed. For storing and retrieving calculated and experimental information, we have developed the 'drug, markup language database' (Drug-ML DB) as an extension of the 'chemical markup language' (CML) database [3]. Figure12.1 schematically shows the concept of

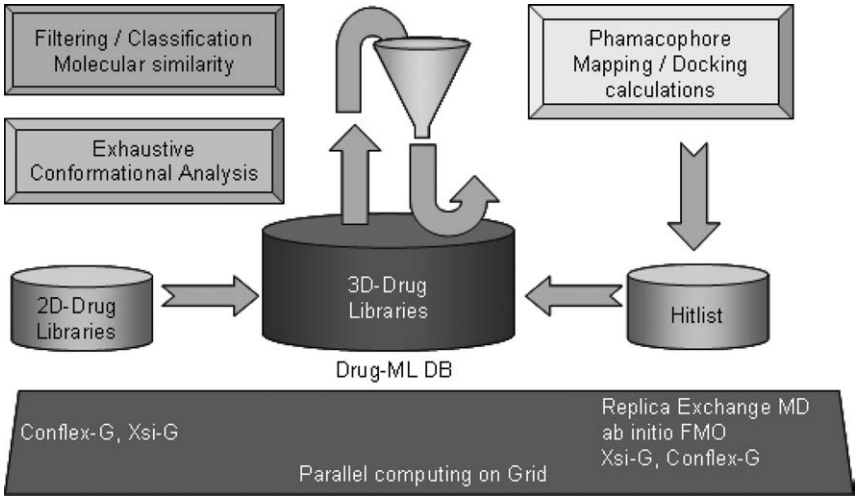


Fig. 12.1. The grid-drug discovery platform.

the grid drug discovery platform. Prior to docking of a drug molecule with its target receptor, an extensive conformational search of the drug molecule is carried out for defining the bioactive conformation. The conformational search is carried out for a series of molecules selected from the database, where information on the two-dimensional structure of molecules is stored. Then, all the low energy conformers obtained in the previous procedure are subjected to the docking analyses. We have developed three methods for evaluating the docking energy; the fragment molecular orbital (FMO), the replica exchange molecular dynamics (REMD), and the empirical scoring function based on the energy obtained from the force field calculation and indexes of three-dimensional complementary structure between a drug and its target protein. All the procedures are executable on the grid environment and Drug-ML DB is implemented as a data grid.

In this chapter, we will describe (1) a grid-based exhaustive conformational search of a drug molecule, (2) a replica exchange molecular dynamic toolkit, (3) a fragment molecular orbital calculation, and (4) a database system written in drug markup language and an example of virtual screening in the grid environment.

12.2 GRID-BASED EXHAUSTIVE CONFORMATIONAL SEARCH

For the computational chemistry tools of drug discovery, the first difficulty to overcome is the conformational flexibility. Fast and exhaustive conformation search is one of the key technologies for prediction of bioactive conformations, development of a three-dimensional molecular conformation database, and docking analysis with some target receptors. To resolve these problems, we have started on improvements to our original conformational space search method 'CONFLEX' using parallel computing and the grid environment which is provided by the 'OmniRPC' system.

12.2.1 OmniRPC system

OmniRPC is a grid RPC system which allows seamless parallel programming from a PC cluster to the grid environment. OmniRPC inherits its API and basic architecture from Ninf. A client and the remote computational hosts, which execute the remote procedures, may be connected via a network. The remote libraries are implemented as an executable program which contains a network stub routine as its main routine. We call this executable program a remote executable program. OmniRPC's worker process is invoked on demand when the program makes a remote procedure call, and all workers are killed when the client program is finished. During the execution of the client program the workers are reused for the subsequent calls. The client does not always poll the workers. The endian conversion is done in OmniRPC automatically by checking the endian of the worker's machine. The conversion between 32 bits and 64 bits is not supported because it may loose the accuracy. This matter is up to the users.

When the OmniRPC client program starts, the initialization function of the OmniRPC system invokes the OmniRPC agent program omrpc-agent in the remote hosts listed in the host file. To invoke the agent, the user can use the remote shell command *rsh* in a local-area network, the GRAM (globus resource allocation manager) API of the Globus

toolkit in the grid environment, or the secure remote shell command *ssh*. The user can switch the configurations only by changing the host file.

OmniRPC Call is a simple client programming interface for calling remote functions. When OmniRPC Call makes a remote procedure call, the call is allocated to an appropriate remote host. When the client issues the RPC request, it requests that the agent in the selected host submit the job of the remote executable with the local job scheduler specified in the host file. The client sends the data of the input arguments to the invoked remote executable, and receives the results upon return of the remote function. Once a remote executable is invoked, the client attempts to use the invoked remote executable for subsequent RPC calls to eliminate the cost of invoking the same remote executable again.

Herein, a typical grid resource is regarded as a cluster of geographically distributed PC clusters. For PC clusters on a private network, an OmniRPC agent process on the server host functions as a proxy to relay communications between the client and the remote executables by multiplexing the communications using a single connection. This feature, called multiplex IO (MXIO), allows a single client to use up to 1 000 remote computing hosts. When the PC cluster is inside a firewall, the port forwarding of SSH enables the node to communicate with the outside with MXIO. Figure 12.2 shows the overview of the OmniRPC system for a remote cluster with a private IP address.

For parallel programming, the programmer can use asynchronous remote procedure calls, allowing the client to issue several requests while continuing with other computations. The requests are dispatched to different remote hosts to be executed in parallel, and the client waits or polls the completed request. In such a programming model with asynchronous remote procedure calls, the programmer should handle outstanding requests explicitly.

OmniRPC efficiently supports typical master/worker parallel applications such as parametric execution programs. For parametric search applications, which often require large amounts of identical data for each call, OmniRPC supports a limited persistence model, which is implemented by the automatic initializable module. The user can define an initialization procedure in the remote executable in order to send and store data automatically in advance of actual remote procedure calls. Since the remote executable may

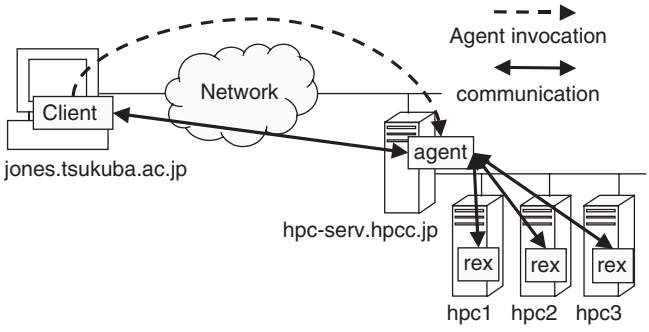


Fig. 12.2. Overview of the OmniRPC system for the remote cluster having a private IP address.

accept requests for subsequent calls, the data set which has been set by the initialization procedure can be reused. As a result, the worker program can execute efficiently and reduce the amount of data transmitted for initialization.

Once a remote executable is invoked, the client attempts to use the invoked remote executable for subsequent RPC calls. However, OmniRPC does not guarantee persistence of the remote executable, so that the data set by the previous call cannot be used by subsequent calls. This is because a remote call by OmniRPC Call may be scheduled to any remote host dynamically, and remote executables may be terminated accidentally due to dynamic rescheduling or host faults. However, persistence of the remote executable can be exploited in certain applications. An example is a parametric search application: in such an application, it would be efficient if a large set of data could be pre-loaded by the first call, and subsequent calls could be performed on the same data, but with different parameters. This is the case for CONFLEX.

OmniRPC provides a restricted persistence model through the *automatic initializable module (AIM)* in order to support this type of application. If the initialization procedure is defined in the module, the module is automatically initialized at invocation by calling the initialization procedure. When the remote executable is rescheduled in different hosts, the initialization is called to initialize the newly allocated remote module. This can eliminate unnecessary communications when RPC calls use the same data.

12.2.2 CONFLEX as an exhaustive conformational search algorithm

Since its first development in 1988, CONFLEX has been recognized as one of the most efficient conformational space search programs that can predominately and exhaustively search the conformers existing in the lower energy regions, and has been applied to the elucidation of the reactivity and selectivity of drug-related materials [4]. Figure 12.3 shows a flow diagram of the original CONFLEX search algorithm. There are four simple procedures: (1) the selection of an initial structure, (2) the generation of many trial structures by using local perturbations that are applied to several parts of the initial structure, (3) all trial structures are then subjected to optimizations, and (4) those successively optimized structures are compared with the other conformers stored in the conformation database. If a new conformer is found, it is added to the conformation database. When this search cycle is finished, the next new initial structure is selected from the conformation database. Although all of these procedures are almost same as for other reliable conformational space search methods, CONFLEX is characterized by two unique strategies: the thermal perturbations (corner flap, edge flip and stepwise rotation) and the ‘lowest-conformer-first’ selection rule of each initial structure from the conformation database. We call this searching scheme in CONFLEX the ‘reservoir filling algorithm’, because the whole scheme resembles the process of pouring water into an empty reservoir or dam.

12.2.3 Parallelized and grid-based CONFLEX

CONFLEX can efficiently handle molecules of up to 100 or so atoms if it works on a computer having a single CPU. In order to apply CONFLEX algorithms to molecules larger

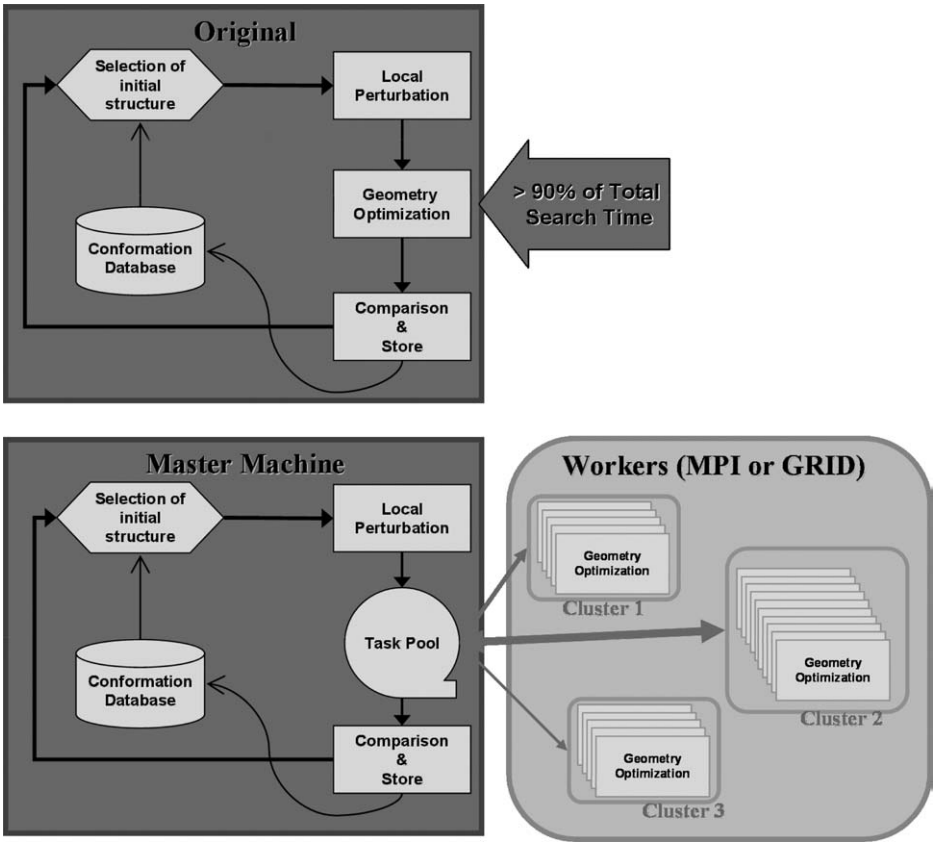


Fig. 12.3. Flow diagram of the original CONFLEX search algorithm (upper) and its parallelized procedures of MPI and grid technology (lower).

than a few hundred atoms of a molecule, CONFLEX has been improved by using high performance and parallel computing techniques. Because the geometry-optimization part of this algorithm requires over 90 percent of the computing time of the search, it was obvious that this procedure should be parallelized to increase efficiency. After applying some parallelization models, we found that the optimized 'master/workers' (M/W) model is the best match of parallelization techniques for the CONFLEX search algorithm [5].

Modifications of the search procedures are as follows: after trial structures are generated, they are temporarily stored in a task pool on the master node, then each worker node is dynamically supplied with one trial structure from the master node. After an optimization on a worker node is finished, the worker is immediately supplied with another trial structure. When all of the trial structures related to a given initial structure are optimized, only the master procedure is used for comparison.

In order to examine the efficiency of the improved algorithm, both CONFLEX-MPI and CONFLEX-G were developed by using the MPI and OmniRPC libraries [6], respectively. Many conformation searches of small peptides were performed on our local cluster

(16 worker processors) and the grid testbed (64–112 workers). The speed-up of conformation search of N-acetyl octa-alanine methylester $(Ala)_8$ has been compared using three kinds of M/W parallelized CONFLEX-MPI on a local cluster. The best performance reached an almost linear speed-up by the optimized M/W codes of the parallel CONFLEX [7].

Using RPCs, CONFLEX-G allocates the processes of trial structure optimization, which are performed by the computation nodes of a PC cluster in CONFLEX-MPI, to the computational nodes of each cluster in the grid environment. There are two computations by the worker programs in CONFLEX-G. One is the initialization of a worker program, and the other is the calculation of trial structure optimization. First, the OmniRPC facility of the AIM is adapted for initialization of a worker program. This facility automatically calls the initialization function in the worker program, once the client program invokes the worker program in a remote node. After the Initialize function has been called to set up common initialization data, a worker program can reuse these data efficiently. Second, in order to calculate trial structure optimization in a worker program, the worker program must receive the data, such as the coordinates of the trial structure and the internal energy state. After finishing each trial structure optimization of a worker program, the result is returned to the client program. Since the portion of the trial structure optimization can be designed independently, we have parallelized this portion using asynchronous RPCs on the client side. Since the OmniRPC system has an easy round-robin scheduler, we do not have to write the code explicitly for load balance, while RPCs are allocated automatically to idle workers.

The grid testbed was constructed by computing resources at the University of Tsukuba, the Toyohashi University of Technology (TUT) and the National Institute of Advanced Industrial Science and Technology (AIST). Table 12.1 shows the computing resources used for this study. In all of the CONFLEX-G experiments, the client program was executed on the master node of the Dennis cluster at the University of Tsukuba. Figure 12.4 shows the speed-up for the sample molecules N-acetyl methylester derivatives of tetra-alanine $(Ala)_4$ and hexdeca-alanine $(Ala)_{16}$ when using different combinations of clusters in our grid testbed. In the case of $(Ala)_{16}$, CONFLEX-G achieved a speed-up by using two or more PC clusters in our grid testbed [8]. This was because (1) the calculation time on the worker program was long, and (2) the overhead of network latency and the invoking of worker programs became relatively small and could be hidden. Figure 12.5 shows the performance improvement by the OmniRPC AIM facility in the case of a large molecule, $(Ala)_{16}$.

Table 12.1 Machine configurations in the grid testbed

Site	Cluster name	Machine	Authentication	Number of nodes	Number of CPUs
University of Tsukuba	Dennis	Dual Xeon 2.4 GHz	Globus, SSH	14	28
	Alice	Dual Athlon 1800+	Globus, SSH	18	36
TUT	Toyo	Dual Athlon 2600+	SSH	8	16
AIST	Ume	Dual Pentium3 1.4 GHz	Globus, SSH3	2	64

Fig. 12.4. Speed-up ratio, which is based on the elapsed time of CONFLEX-G using one worker in the Dennis cluster. 'AlaX04' (light blue) and 'AlaX16' (purple) correspond to $(Ala)_4$ and $(Ala)_{16}$, respectively.

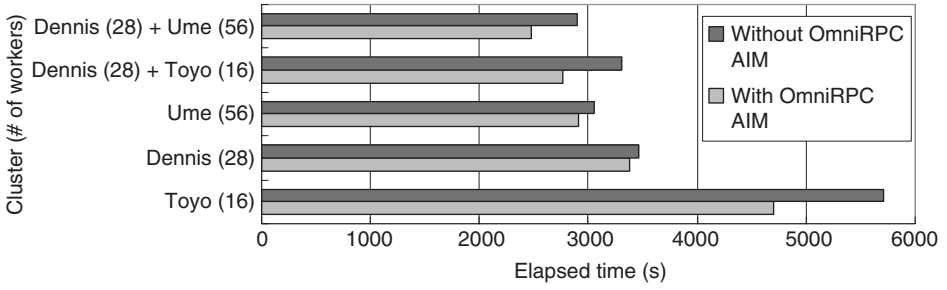


Fig. 12.5. Performance of CONFLEX-G with and without the OmniRPC facility of automatic initializable module for $(Ala)_{16}$.

12.2.4 Application of CONFLEX to peptide folding

We have applied M/W parallelization and grid technologies into CONFLEX, and shown that the performance of the parallelized CONFLEX enables exploration of the lower-energy region of the conformational space of drug molecules and small peptides within an available elapsed time using a PC cluster or grid environment. It is possible to make it that a series of molecules selected from a two-dimensional structure database is subject to a conformation search by using parallel CONFLEX, and the important conformers found are stored in a three-dimensional conformation database described in our Drug-ML. This was one of the first purposes of this project 'grid drug discovery'. As another application of parallel CONFLEX, we propose here a computational approach to prediction for a folding process of a small peptide by using

a combination of an exhaustive conformational space search and a conformation clustering technique, which are accelerated with both general parallel processing and grid technology.

If we know the energy-potential surface, at least, all the important transition states between conformers of a molecule, we can understand their dynamical conformational interconversion processes based on transition state theory. Although transition states of such a large molecule are usually unknown, some conformational similarity indices can approximately evaluate the barrier of the inter-conversion. As one of the candidate equations of similarity indices, a conformational distance in an angular space based on all peptide's backbone torsion angles (ϕ and ψ) is useful for such evaluation. Based on this assumption, we developed a simple program for finding a series of conformers between widely separated conformers in its conformational space by using a modified single-linkage clustering technique based on the similarity indices of a conformational distance matrix.

As an example, a folding process of a 12-residue peptide (1LE0) known as a stable β-hairpin tryptophan zipper [9], is predicted here. In order to complete requirements for searching a folding pathway, an exhaustive conformation search of low-energy regions was also executed as follows: in each CONFLEX search cycle, all backbone and sidechains' torsion angles of each initial structure were stepwise rotated to generate trial structures. Refinement of all trial structures generated was efficiently performed by using two optimization methods; steepest decent as a pre-optimization of crude structures, and a full-matrix Newton–Raphson method to reach to each energy minimum on the potential energy hypersurface. We applied the MMFF94s, known as one of the most reliable force fields for small organic molecules, and a dielectric continuum model for evaluation of the relative conformation energies in water (where the dielectric constant is equal to 78.4). Duplications among their optimized structures were reduced by checking the similarity index between two conformers. Finally, 22 897 unique conformers of 1LE0 were found by using a parallel CONFLEX search with 16 workers (7.2 days). Figure12.6 shows the energy changes of a predicted folding process from strand form to β-hairpin conformation of 1LE0 based on our approximation of the transition state theory as mentioned above. Theoretical CD (circular Dichroism) and UV spectra assigned to the representative structures were also calculated, as shown in Fig. 12.6 [10]. It is interesting to note that the sum of conformational distances between the neighboring conformers along the reaction coordinates corresponds to the reaction time of this folding process, if the conformational distance between two minima is correlated with the corresponding transition state. Systematic studies on the possible relationship between barrier heights and the conformational distances would be interesting [11]. It is also surprisingly noted that theoretical CD and UV spectra of the most stable conformer are almost in good agreement with the observed one.

12.3 REPLICA EXCHANGE MOLECULAR DYNAMICS TOOLKIT FOR DRUG-RECEPTOR DOCKING

We have developed a toolkit for replica exchange molecular dynamics (REMD) which realizes efficient conformational sampling. Object-oriented design is employed to adapt

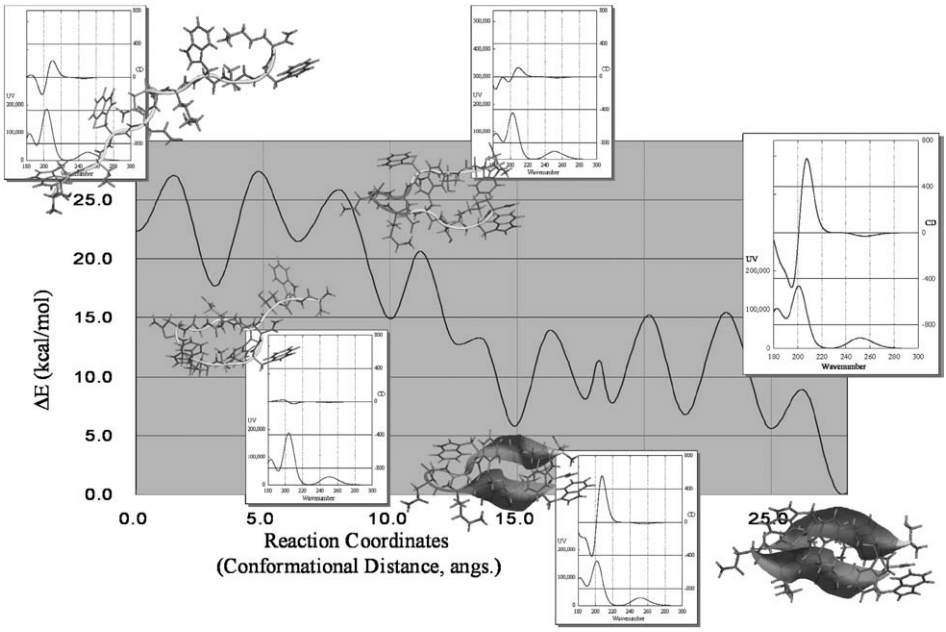


Fig. 12.6. Energy changes of a predicted folding process between a strand and β-sheet form of Typ-Zip peptide. Three dimensional-images and theoretical UV (lower)/CD (upper) specra ofrepresentative structures at 0, 3, 10, 15 and 28Å in conformation distance (similarity index) from the strand form are also shown.

the toolkit to molecular dynamics program Mindy (serial processing version of NAMD) with CHARMM force fields without significant modifications. The effectiveness of REMD was examined with the polyalanine, $(Ala)_{10}$, and compared with REM1. The second structures were sampled efficiently in the trajectory. Linear speed-up and 94 percent parallel efficiency will eight CPUs were observed on Linux clusters with MPI.

Molecular simulation of the drug-receptor docking process is essential in drug discovery to screen out less promising candidate molecules. This process does not exactly proceed in a lock-and-key fashion whereas lock-and-key plays an important role in drug design. Protein flexibility makes it more dynamic so that the receptor can dynamically fit a ligand molecule. In order to simulate such a process, full consideration of molecular structural fluctuations is usually required. This atomistic level simulation can improve the accuracy of drug-receptor affinity and is expected to enhance the productivity of new drugs.

But there are two major difficulties. One is prohibitively large computational cost to sample numerous conformations by using an all atom potential energy function. It costs much more than the scoring function currently used in a docking program and the multiple-minima problem prevents sampling. A possible resolution of this difficulty is the replica exchange method which is an extended ensemble method. This method is usually called parallel tempering, replica Monte-Carlo method, multiple Markov chain method, and so on.

The replica exchange method is an integrated method consisting of molecular structure exchange, structure sampling, and force field calculation. Structure exchange between statistical ensembles accelerates the conformational sampling. Besides, the replica exchange method is very suitable for parallelization so the integration can be achieved on a network: grid computing environment. Since integration on a grid usually might cause large and complicated software systems, this is the other difficulty for the simulation. A conventional way for development often ends up with the notorious anti-pattern 'reinvent the wheel'. Why do they not reuse their software components? That is because the software subsystems depend on each other via data structures for biomolecules, force fields, sampling methods and parallelization. The slight modification for data structure causes a complete modification of the whole software, that is, the anti-pattern 'reinvent the wheel' and its obvious solution is an object-oriented approach.

Hence we have developed an object-oriented framework for statistical mechanics simulations to integrate several physical concepts, and its corresponding software components into a replica exchange method.

12.3.1 Object-oriented framework for statistical mechanics simulations

We have used C++ abstraction mechanisms to concentrate on the meaning and relationships of physical concepts, to encapsulate irrelevant details, such as data structure for force fields, and algorithms for parallelization. It was also intended to reduce the dependencies between components for reusability.

Physical concepts are reflected in these rectangles, C++ classes. The relationships between the concepts are represented by arrows and lines in Fig. 12.7. Step ensemble is an ordinary statistical ensemble, which is generated by Monte-Carlo sampling. Dynamical ensemble is another type of ensemble, which is generated by molecular dynamics calculation. According to the foundation of statistical mechanics, ergodic hypothesis, these two ensembles can give the same statistical distribution in finite sampling. Besides, the message protocols to these classes are exactly the same, so that we generalize these classes to the abstract ensemble class. Although the relationship between ensemble and extended ensemble is not shown here, extended ensemble does not refer to these two concrete classes, but to this abstract class.

Statistical ensemble is a set of microscopic states of a molecule. Ensemble object should have an object representing a microscopic state, so walker classes were made to encapsulate these microscopic data. We use this encapsulation of data structure to prevent the adaptation to other software packages from causing a total modification of the system. The toolkit was adapted to molecular dynamics program Mindy (NAMD) [12] with CHARMm force fields without significant modifications.

12.3.2 Performance evaluation

We examined the effectiveness of the toolkit with polyalanine $(Ala)_{10}$. The force field was CHARMM19 taken from X-PLOR. The molecular conformations were generated through constant temperature molecular dynamics (NVT MD) with the Gaussian thermostating technique, where the trajectory was taken for 200 ps (200 000 steps) with time

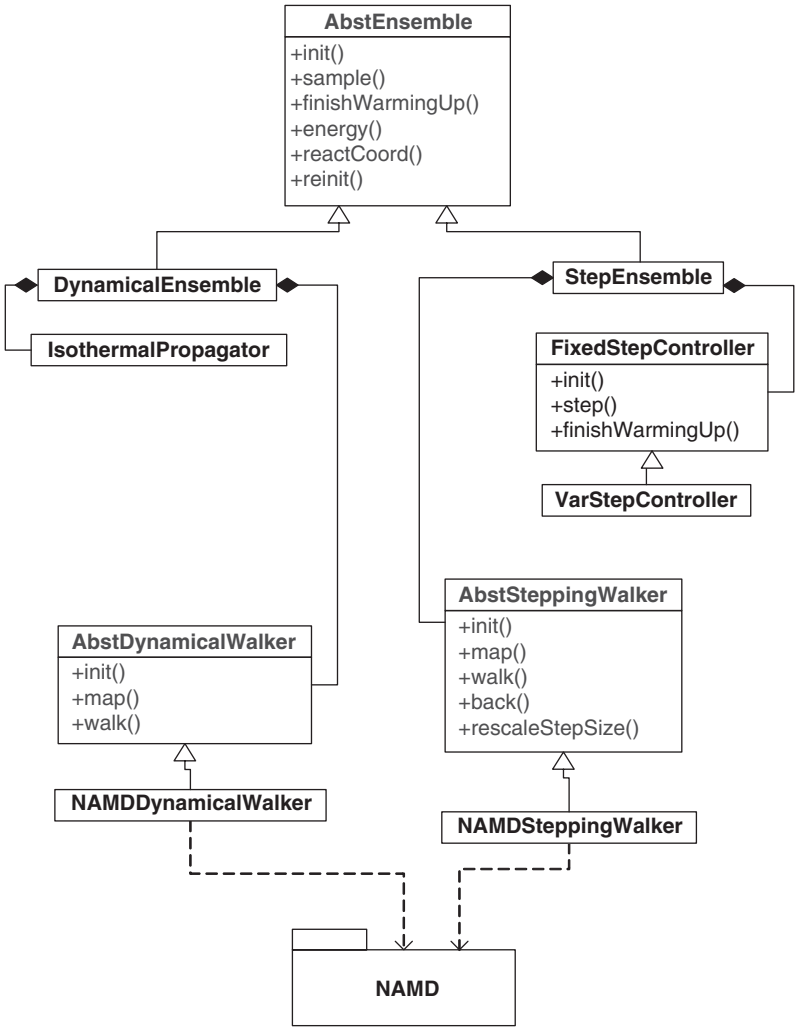


Fig. 12.7. Class diagram for the ensemble and walker layers.

step 1.0fs. The temperatures of eight replicas were exponentially distributed temperatures from 168.2 K to 565.7 K. The resultant time series of the potential energies and structures are shown in Fig. 12.8.

Random walks of energies support the effectiveness of conformational sampling. The structures at the highest temperature are shown as ribbons, describing the peptide backbone. The second structure, a helix, is partially broken and changing its shape drastically. At the lowest temperature, the helix structure is preserved.

In these snapshots, atoms are also shown as transparent spheres. The fluctuation of atoms does not change the second structure at all at the lowest temperature. This behavior is explained by energy landscape theory. Namely, energy landscape theory assumes

the effective energy surface has a funnel shape, so that chain entropy falls down and the number of conformations decreases as the energy goes down to the bottom. Therefore, when the ensemble temperature is high, a variety of second structures are sampled, but when the temperature is low, just one second structure can be sampled.

These structure changes in REMD simulation were further compared with those in the ordinary replica exchange method. REMD is based on MD sampling, whereas REM is based on MC sampling. It is known that MD is more advantageous for sampling conformations of larger systems.

In the case of being combined with the replica exchange method, our result suggests that REMD based on MD would be more advantageous than conventional MD and MC. Especially, the structures of the ensembles change more frequently in REMD than in REM during the same number of simulation steps.

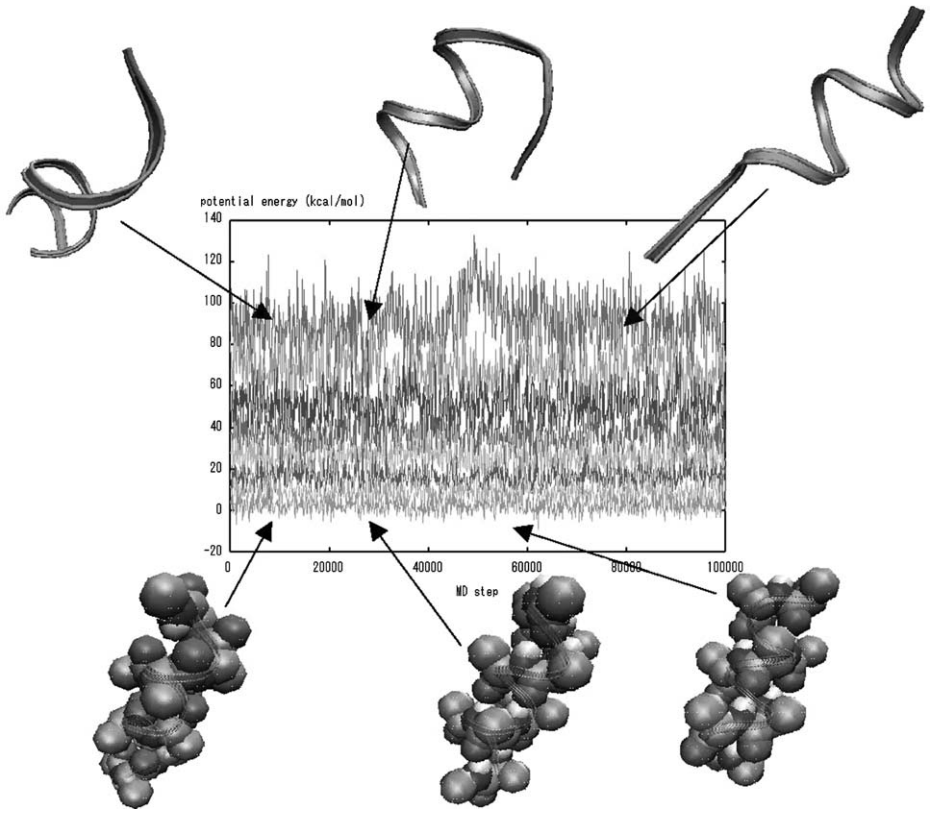


Fig. 12.8. Time series of the potential energies and structures of $(Ala)_{10}$. NVT MD Gaussian thermostat, time step 1.0fs, trajectory 200ps (200 000 steps). Eight replicas: exponentially distributed temperatures from 168.2 K to 565.7 K (see Color Plate 20).

12.4 DNA AND ESTROGEN RECEPTOR INTERACTION REVEALED BY FRAGMENT MOLECULAR ORBITAL CALCULATION

The interaction between ligand and DNA is a very important gene-based drug design. In order to investigate the interaction controlling coordination of ligands to DNA in detail, the molecular orbital calculation was carried out for a complex of the estrogen response element of DNA (DNA-ERE) and the DNA binding domain of the estrogen receptor (ER-DBD) using the fragment molecular orbital (FMO) calculation technique. The estrogen receptor is well known as one of the nuclear receptors concerning endocrine disruptors. The DNA-ERE + ER-DBD geometry used for FMO calculation was determined by molecular dynamics simulation. The binding energy of ER-DBD to DNA-ERE was analyzed by inter-fragment interaction energies (IFIEs) in the FMO calculation. This is considered to be an instructive case study of drug–receptor interaction by FMO.

12.4.1 DNA and estrogen receptor interaction

Estrogen receptor (ER) families belong to a superfamily of ligand-induced nuclear steroid hormone receptors. These families repress or enhance transcription by identifying and binding as a dimer to specific DNA target sequences, such as the estrogen response element. In the DNA binding domain of ER, zinc finger (Zn-finger) motifs formed by eight cysteine residues and two zinc ions play a critical role in the specific binding between DNA-ERE and ER-DBD.

There are many experimental and theoretical studies about the specific binding of Zn-finger in [13–16]. The structural aspect of the binding between DNA-ERE and ER-DBD is revealed in that one of the α-helixes of Zn-finger interacts with the primary groove of DNA-ERE and another α-helix positioned in the secondary groove specifies the sequences of DNA-ERE [13]. However, the detailed structure and mechanism of the specific binding are not known. There have been some molecular dynamics (MD) studies in which the binding structure of DNA-ERE and ER-DBD were reproduced [14].

On the other hand, the mechanism of the transcriptional factor is not understood at all. Change of the electronic structure of DNA was found to be the key process, but the system was too large for the molecular orbital calculation. The MD calculation can treat a large system, but the change of the electronic structure cannot be considered.

Non-empirical molecular orbital calculation is essential to investigate the binding between DNA and ER-DBD because of strong Coulomb interaction caused by ionic phosphate of DNA and Zn-finger motifs of ER-DBD, and weak interaction induced between ER-DBD and bases at the consensus sequences of DNA. ER-DBD has Zn atoms, which are very difficult to handle by the force field with fixed charge. Furthermore charge reorganization caused by solvent effect is also important for the binding between DNA and ER-DBD.

Challenging studies applying the non-empirical molecular orbital method to proteins and enzymes have been carried out [17–19], but it is a very time-consuming calculation even using the most up to date computer. The fragment molecular orbital (FMO) method [20–27] is the approximation for the MO calculation of a huge molecule, for example, protein and DNA, and reproduces the total energy of the MO calculation surprisingly well with moderate computational effort.

Inclusion of the solvent molecules in the FMO calculation is very important to understand the role of the solvent molecules in the binding between DNA and ER-DBD because DNA has the anionic phosphate groups and then DNA does not exist as a stable structure in the gas phase. Furthermore, ER-DBD has Zn-finger motifs, largely charged motifs, and solvation must affect the electronic state of ER-DBD. The FMO calculation makes it possible to include a large number of solvent molecules, thus the solvent effect in the binding of ER-DBD to DNA-ERE can be estimated.

In this study, the FMO calculation of the DNA-ERE and ER-DBD system was carried out in order to reveal the interaction between DNA-ERE and ER-DBD. The effect from the solvent molecules was also examined.

The ABINIT-MP program package [28] was used in the all calculations in this study.

12.4.2 Analysis of DNA-ERE and ER-DBD systems by FMO

The FMO calculation of DNA-ERE and ER-DBD was carried out with and without the water molecules and counter-ions. The complex structure as shown in Fig. 12.9 was constructed with the molecular dynamics simulation [14]. The consensus sequence consists of the palindrome sequence of six base pairs (TGACCT). DNA-ERE was divided into the

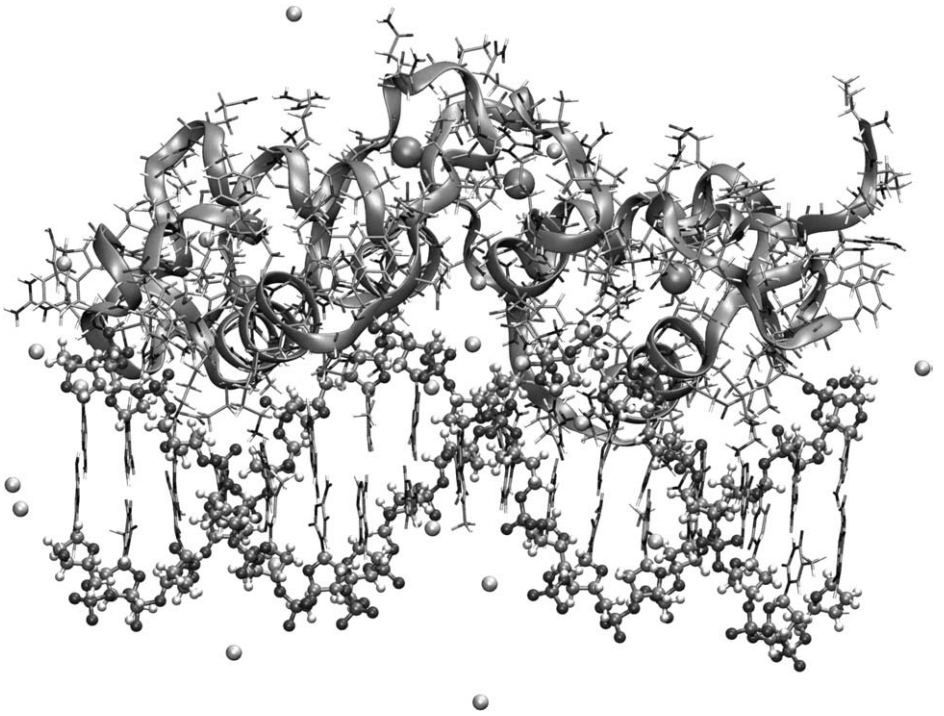


Fig. 12.9. The structure of complex between DNA ERE and ER DBD. The complex is surrounded by counter ions and water molecules, although water molecules are not shown for clarity.

base fragments and the phosphate and sugar fragments. ER-DBD was divided into each amino acid except for the cysteine residues around Zn.

The water molecules and counter-ions were selected by the nearest neighbor of DNA-ERE and ER-DBD. The numbers of water molecules within 2, 3 and 4 Å were 394, 964 and 1 361 molecules, respectively. Two types of fragmentation were adopted for the FMO calculation. Type 1 was a water cluster model, in which water molecules in the 2, 3 and 4 Å solvent shells were divided to 115, 271 and 364 clusters, respectively. Type 2 was a solvated model, in which water molecules were added to the nearest fragment of DNA-ERE and ER-DBD. The solvated model was adopted for only the 2 Å solvent shell.

The FMO calculation was carried out with the HF/STO-3G level.

The FMO calculations of DNA-ERE and ER-DBD with and without the solvent molecules were carried out. The results were examined by focusing on the IFIEs of the DNA-ERE.

The interaction energies between each DNA base pair and the ER-DBD monomer without solvent molecules are shown in Fig. 12.10. Each base pair in the DNA-ERE is divided into two fragments: one is phosphate and sugar, and the other fragment is base.

There are two peaks of interactions between ER-DBD and phosphate/sugar at AT (or TA) in DNA-ERE. These interactions are mainly due to electrostatic interaction between the minus charge of phosphate and the plus charge of ER-DBD, and roughly fix the position and direction of binding of ER-DBD to DNA-ERE. This interaction works as a fork sticking ER-DBD to DNA backwards. There are other interactions between ER-DBD and bases at the consensus sequences of DNA-ERE. These interactions are contact interactions such as hydrogen bonding, and have an important role for recognition of the consensus sequence. Namely, this effect works as a kind of glue for fine alignment of ER-DBD and DNA-ERE.

From Fig. 12.10, the role of ER-DBD dimer is clear. The ER-DBD dimer looks to have a fork with two spikes: a strong one and a little weak one. The fork fixes the position of

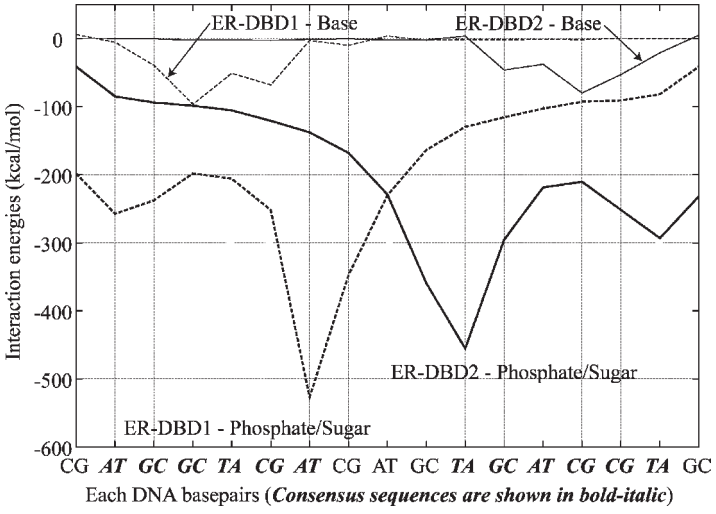


Fig. 12.10. Interaction energies between ER-DBD and base pair in DNA-ERE.

coordination at the DNA and the direction for rough alignment because the monomer which has one spike fixes only the position not direction.

The interaction energies between DNA-ERE and each ER-DBD are listed in Table 12.2. DNA-ERE and ER-DBD have −32 and +5 kcal/mol charges, respectively. Thus interactions between DNA-ERE and ER-DBD and between ER-DBD monomers without solvent molecules are basically attractive and repulsive, respectively.

In cases with solvent molecules by type 1 (water cluster model), the interaction between DNA-ERE and ER-DBD becomes weaker, as the solvent shell becomes thicker. These changes of the interaction between DNA-ERE and ER-DBD are due to the change of the electronic structure of DNA-ERE and ER-DBD and cannot be estimated by molecular mechanics with fixed charges. Maximum change of the interaction between DNA-ERE and $(ER\text{-}DBD)_2$ is 211 kcal/mol. On the other hand, the attractive interaction between ER-DBD monomers becomes stronger as the solvent shell becomes thicker. Maximum change of ER-DBD monomers is −21 kcal/mol. These results are due to the dielectric shielding effect of the solvent molecules.

In cases with solvent molecules by type 2 (solvated model), all interactions are stronger than the interaction without solvent molecules. The interactions for type 2 include extra interaction with water molecules solvated. Changes of interactions are −224 kcal/mol for DNA-ERE – $(ER\text{-}DBD)_2$ and −74 kcal/mol for ER-DBD monomers. These increased attractive interactions are mainly due to hydrogen bonds via water molecules. In the interaction between DNA-ERE and ER-DBD, enlargement of the interaction by hydrogen bonds via water molecules acts in the opposite direction to the dielectric shielding effect shown in type 1, and will be cancelled out. In the interaction between ER-DBD monomers, however, both effects act in the same direction and are enhanced by each other.

The dielectric shielding effect and the hydrogen bonds via water molecules largely affect interactions of ER-DBD with DNA-ERE. Namely, the water molecules are gluing the ER-DBD and DNA by a hydrogen bonding network. Explicit treatment of the electronic state by the FMO calculation reveals such roles of water molecules clearly.

In summary, the ER-DBD has a spike caused by strong Coulomb interaction between ionic phosphate of DNA and Zn-finger motifs of ER-DBD. The weak interaction induced between ER-DBD and bases at the consensus sequences of DNA acts as a kind of glue to bind them. Therefore, the dimer of ER-DBD became a fork with two spikes, a long one and a little short one, which determines the position and direction of the coordination to DNA. Then the dimer makes the fine alignment very efficiently.

Table 12.2 Role of solvent molecules in the interaction energies in kcal/mol

	DNA – ER1	DNA – ER2	DNA – $(ER)_2$	ER1 – ER2
No water	−3 543	−3 637	−7 180	16
+water 2Å (type 1)	−3 517	−3 618	−7 134	16
+water 3Å (type 1)	−3 473	−3 591	−7 064	4
+water 4Å (type 1)	−3 418	−3 551	−6 969	−5
+water 2Å (type 2)	−3 680	−3 724	−7 404	−58

In the solvation effect, the binding energies are enlarged by indirect interaction via water molecules, although the electrostatic interactions are weakened by the dielectric shielding effect. The water molecules shield long-range Coulomb interaction for rough alignments between the zinc finger of ER-DBD and charge on the DNA main chain. On the other hand, the water molecules glue ER-DBD and DNA by a hydrogen bonding network.

We now plan to analyze the importance of dispersion interaction caused by electron correlation in the DNA chain.

For fine drug design, the knowledge of interactions between protein and organic molecule is very important. Therefore conventional molecular mechanics is applied for rough but wide and rapid search. For fine search the combination of FMO and MM, or FMO should be applied to have sophisticated interaction analysis. Such a phase-down technique for searching area is especially important in high throughput drug screening.

12.5 DRUG MARKUP LANGUAGE AND VIRTUAL SCREENING ON GRID

12.5.1 Drug ML database

Improving technologies such as high throughput screening and combinatorial chemistry have enabled pharmaceutical companies to screen a thousand drug candidates in a day. Consequently, a demand for faster and more efficient *in silico* screening has been increasing. In reply to this increasing demand we have developed an integrated database system written in 'Drug ML'. This database system, implemented on the grid environment, can manage various calculated and experimental information uniformly. Recently, by the prosperity of XML (extensible markup language), the markup language has been used as a common data structure in various fields of science and business. Drug ML is a markup language designed for drug discovery. In the field of chemistry chemical markup language (CML [3]) was proposed and has been used widely. We extended CML in order to treat drug–receptor complexes and conformational isomers and to include various molecular properties as well as molecular structures in the database. The Drug ML database system makes various applications on the grid environment connecting each other. Figure 12.11 shows the data structure of Drug ML, in which the following three tags are introduced:

(a) 'Universe tag' describes a complex structure where multiple molecules are involved. For example, in the database we can express a complex between HIV-1 protease and its inhibitor.
(b) 'Conformation tag' describes conformational isomers of a molecule. We can handle a set of conformers obtained from the exhaustive conformational search.
(c) 'Descriptor tag' describes various structural descriptors of a molecule. These descriptors are used for QSAR analyses of a series of compounds.

The database system based on Drug ML locates at the center of the platform for drug discovery, and it enables us to connect various data seamlessly on the platform. Figure 12.12 shows the architecture of the database system. It consists of the following four

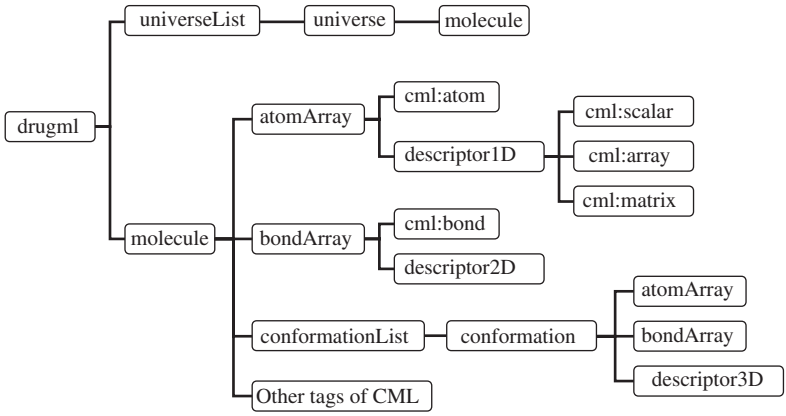


Fig. 12.11. Structure of Drug ML

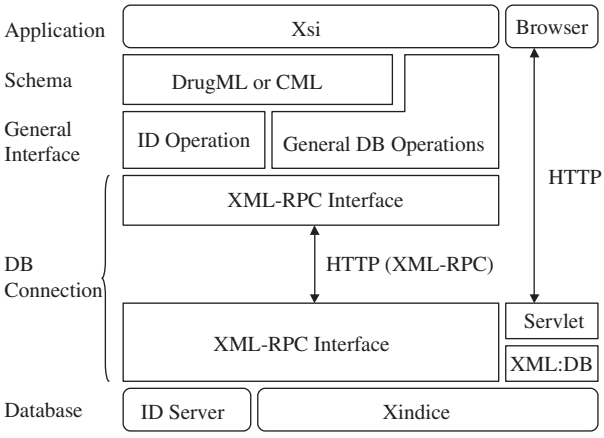


Fig. 12.12. Architecture of database system.

layers: (1) application, (2) schema (data structure), (3) API for the database operation, and (4) connection module to the database. Through this architecture the system can absorb differences in languages, formats and data structures.

12.5.2 Virtual screening on the grid environment

In order to confirm the effectiveness of the system mentioned in Section 12.5, we did two computational experiments: a docking calculation and virtual screening. First we performed a docking calculation between the viracept (the HIV-1 protease inhibitor) and the HIV-1 protease. We assumed only the structure of HIV-1 protease but not the three-dimensional structure of viracept. The flow of the calculation is as follows: (1) generate conformations of viracept by the Monte-Carlo method, (2) define the binding site of

HIV protease by Proshape [29], (3) calculate WHIM (weighted holistic invariant molecular [30,31]) descriptor of HIV-1 protease and viracept, (4) get alignments between HIV-1 protease and viracept using the WHIM descriptor, and (5) perform the docking calculation in parallel using OmniRPC as explained in Section 12.2.1. In these calculations, we used Xsi (pronounced ku-su-shi [32]) for the Monte Carlo method, calculation of WHIM descriptor and docking simulation. Xsi is a simulation environment, on which various types of calculations for drug design can be combined and executed. As a result, we found that the speed of docking calculations can be linearly increased up to at least about 80 CPUs and the rmsd (root mean square distance) of the calculated structure and the X-ray structure of viracept was 1.77 Å, which is shown in Fig. 12.13.

Secondly we performed a virtual screening, in which 888 randomly selected ligands with five known HIV-1 protease inhibitors were tested via docking simulation. The flowchart of virtual screening is shown in Fig. 12.14. First of all we generated a map of the binding site of a target receptor. Then we generated stable conformations of ligands in the database, aligned them with the map of the binding site, and randomly rotated each conformation. We used the 'ligand alignment' module implemented in Xsi in order to optimize alignment ligand conformations with the binding site. Consequently five HIV-1 protease inhibitors were ranked as low ranks and the calculation time was about 320 minutes in 40CPUs on the grid environment.

We confirmed potential effectiveness of our system based on grid and XML technologies by two computational experiments. In the future we will brush up Drug ML and the database system and will perform large-scale computational experiments using more CPUs.

12.6 CONCLUDING REMARKS

A little over four decades have passed since the discovery of the so-called quantitative structure–activity relationship. As well as recent progress in experimental technologies in drug discovery and rapid incensement of genomic information, recent advances in

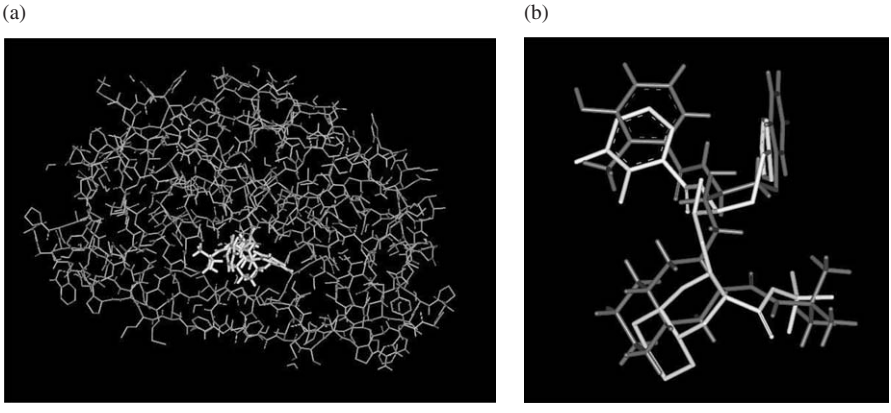


Fig. 12.13. (a) Complex between HIV 1-protease (gray line) and viracept (white line). (b) Comparison between computed (white line) and X-ray (gray line) structures of viracept.

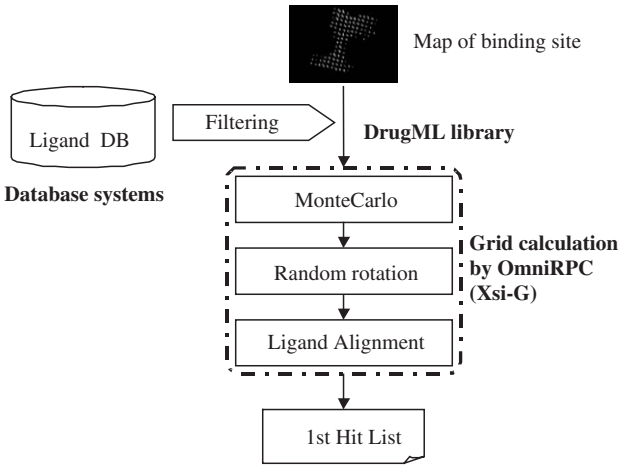


Fig. 12.14. Flow chart of virtual screening.

modern molecular calculations and simulations have been opening up a new era in the field of drug discovery. Computational techniques such as parallel computing and grid will enable us to probe the drug-receptor mediated phenomenon directly. As we have discussed, these methods and techniques, in conjunction with experimental studies, should bring a new approach to drug discovery. In conclusion, much work and development remains, but we now have powerful computational tools for drug discovery at our disposal.

12.7 ACKNOWLEDGEMENTS

This work has been supported by Research and Development for Applying Advanced Computational Science and Technology of the Japan Science and Technology Agency (ACT-JST) in Japan.

12.8 REFERENCES

1 http://www.ud.com/home.htm
2 http://www.combechem.org/
3 http://www.XML-cml.org/
4 H. Goto and E. Osawa, J. Am. Chem. Soc., 111 (1989) 8950; H. Goto and E. Osawa, J. Mol. Struct., 285 (1993) 157; H. Goto and E. Osawa, J. Chem. Soc., Perkin Trans., 2 (1993) 187; H. Goto, S. Obata, T. Kamakura, N. Nakayama and K. Ohta, CONFLEX5, Conflex Corp., Tokyo, 2005.
5 The latest version of CONFLEX including the parallel version is distributed from Conflex Corp., http://www.conflex.us/.
6 M. Sato, T. Boku and D. Takahashi, CCGrid2003 (2003) 219.
7 H.Goto, T. Takahashi, Y. Takata, K. Ohta and U Nagashima, Nanotech., 1 (2003) 32.
8 Y. Nakajima, M. Sato, T. Boku, D. Takahashi and H. Goto, Proceedings of 2004 International Conference on Supercomputing, 2004, p 154.
9 A.G. Cochran, N.J. Skelton and M.A. Starovasnik, PNAS, 98 (2001) 5578.

10 N. Harada and K. Nakanishi, Circular dichroic spectroscopy — exciton coupling in organic stereochemistry, University Science Books, Mill Valley, CA, US, (1983); R.W.J. Zijlstra, W.F. Jager, B. Lange, P. T. Duijnen, B.L. Feringa, H. Goto, A. Saito, N. Koumura and N. Harada, J. Org. Chem., 64 (1999) 1667.
11 H. Goto, Chem. Phys. Lett., 292 (1998) 254; H.Goto, Tetrahedron, 48 (1992) 7131.
12 Mindy is obtainable from NAMD source code by J. Gullingsrud, http://www.ks.uiuc.edu/Development/MDTools/mindy/ ; L. Kale, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan and K. Schulten, J. Comp. Phys., 151 (1999) 283.
13 J.W.R. Schwabe, L. Chapman J.T. Finch and D. Rhodes, Cell, 75 (1993) 567.
14 M.A.L. Eriksson and L. Nilsson, Eur. Biophys J., 28 (1999) 102.
15 L.Y. Low, H. Hernandez, C.V. Robinson, R. O'Brien, J.G. Grossmann, J.E. Ladbury and B. Luisi, J. Mol. Bio., 319 (2002) 87.
16 T. Dudev and C. Lim, J. Am. Chem. Soc., 124 (2002) 6759.
17 M. Challacombe and E. Schwegler, J. Chem. Phys., 106 (1997) 5526.
18 C. Van Alsenoy, C.-H. Yu, A. Peeters, J.M.L. Martin and L. Schafer, J. Phys. Chem. A, 102 (1998) 2246.
19 F. Sato, T. Yoshihiro, M. Era and H. Kashiwagi, Chem. Phys. Lett., 341 (2001) 645.
20 K. Kitaura, T. Sawai, T. Asada, T. Nakano and M. Uebayashi, Chem. Phys., Lett., 312 (1999) 319.
21 K. Kitaura, E. Ikeo, T. Asada, T. Nakano and M. Uebayashi, Chem. Phys. Lett., 313 (1999) 701.
22 T. Nakano, T. Kaminuma, T. Sato, Y. Akiyama, M. Uebayashi and K. Kitaura, Phys. Lett., 318 (2000) 614.
23 K. Kitaura, S.Sugiki, T. Nakano, Y. Komeji and M. Uebayashi, Chem. Phys. Lett., 336 (2001) 163.
24 T. Nakano, T. Kaminuma, T. Sato, K. Fukuzawa, Y. Akiyama, M. Uebayashi and K. Kitaura, Chem. Phys. Lett., 351 (2002) 475.
25 H. Sekino, Y. Sengoku, S. Sugiki and N. Kurita, Chem. Phys. Lett., 378 (2003) 589.
26 Y. Inadomi, T. Nakano, K. Kitaura and U. Nagashima, Chem. Phys. Lett., 364 (2002) 139.
27 S. Sugiki, N. Kurita, Y. Sengoku, and H. Sekino, Chem. Phys. Lett., 382 (2003) 611.
28 ABINIT-MP program package Ver.20021029.
29 http://csb.stanford.edu/koehl/ProShape/
30 R. Todeschin and P. Gramatica, New 3D molecular descriptors: the WHIM theory and QSAR applications, in H. Kubinyi, G. Folkers and Y.C. Martin (Eds.), 3D QSAR in drug design, Vol. 2, Kluwer/Escom, Dordrecht, 1998, pp. 355–380.
31 R. Todeschini, M. Lasagni and E. Marengo, Theory, J. Chemometrics, 8 (1994) 263.
32 http://www.mizuho-ir.co.jp/science/xsi/

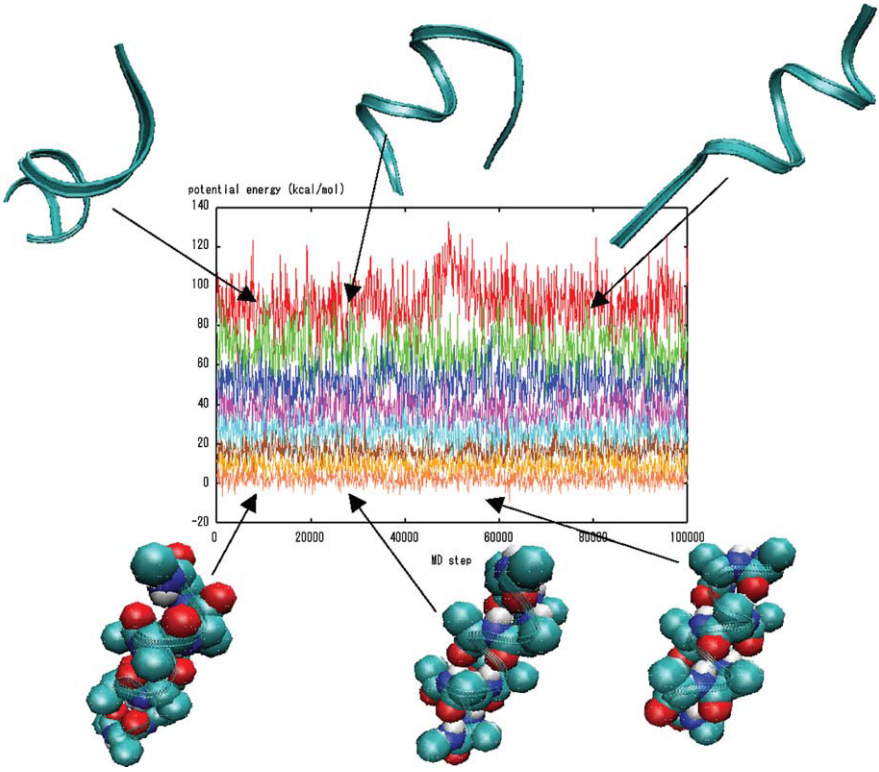


Plate 20. Time series of the potential energies and structures of $(Ala)_{10}$. NVT MD Gaussian thermostat, time step 1.0fs, trajectory 200ps (200 000 steps). Eight replicas: exponentially distributed temperatures from 168.2 K to 565.7 K (illustration appears on page 239 of this volume).

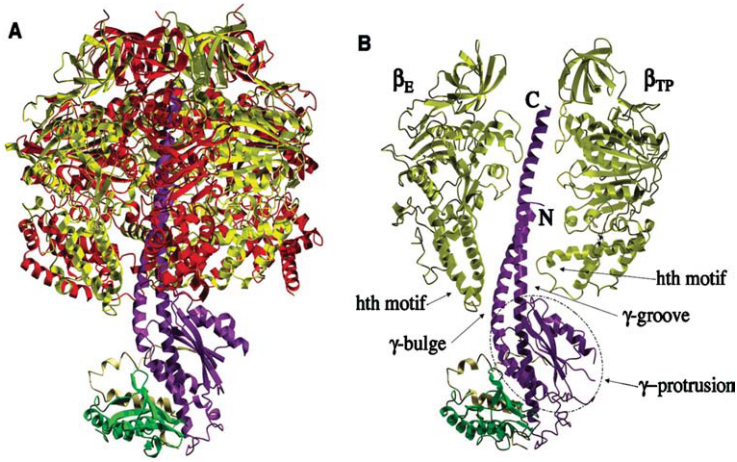


Plate 21. Crystal structure (A) of F_1-ATPase. A comparison between the open catalytic site β_E and one of the closed sites β_{TP} is shown in (B) (illustration appears on page 250 of this volume).

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 13

Thermodynamics and kinetic analysis of F_oF_1-ATPase

Yi Qin Gao[1], Wei Yang[2] and Martin Karplus[2,3]

[1]*Department of Chemistry, Texas A&M University, College Station, TX 77845, U.S.A*
[2]*Department of Chemistry and Chemical Biology, Harvard University, MA 02138, U.S.A*
[3]*Laboratoire de Chimie Biophysique, ISIS, Université Louis Pasteur, 67000 Strasbourg France*

Abstract

In this review, we discuss the thermodynamics and its implications on kinetics for ATP synthesis and hydrolysis by F_oF_1-ATP synthase, a rotary molecular machine. We summarize mainly some of the recent efforts in building a simple theoretical model for this enzyme and focus on showing how simulations and experimental results are combined together to yield new insight into the chemomechanical coupling mechanism. It is shown that the negative cooperativity in nucleotide binding and the positive cooperativity in catalysis together render F_1-ATPase an efficient molecular machine. It is also shown that, through proton–motive–force–driven conformational changes, efficient ATP synthesis and hydrolysis are both achieved under cellular conditions. We also discuss the possible biological significance of these thermodynamic and kinetic properties of this 'splendid enzyme'.

13.1 INTRODUCTION

One of the central questions of life science lies in how ATP (adenosine tri-phosphate) molecules are produced and used as the 'energy currency' in living cells [1–3]. In eukaryotic cells as well as in some bacteria, ATP is created using a splendid molecular machine–F_oF_1-ATP synthase [4]. F_oF_1-ATP synthase exists in the mitochondria inner membrane, the chloraplast membrane and bacteria membrane. It synthesizes ATP from ADP (adenosine di-phosphate) and Pi (a phosphate group) by consuming the proton motive force generated by respiration in mitochondria, photolysis in chloroplast and photosynthetic bacteria, or even glycolysis in several bacteria [5]. In some cases it also hydrolyzes ATP to maintain the proton gradient across the membrane [4].

F_oF_1-ATP synthase contains two parts, F_1 and F_o. The F_1 portion of the mitochondria ATP synthase protrudes into the matrix of mitochondrion and has access to dissolved ADP and Pi. It has a stoichiometry of $\alpha_3\beta_3\gamma\varepsilon\delta$ [4,6–8]. The α and β-subunits are arranged alternately around, and thus interact in a different manner with the asymmetric coiled-coil part of the

γ-subunit [9,10]. There are three catalytic and three non-catalytic binding sites for nucleotides in F_1-ATPase at the interfaces of β and α-subunits. The catalytic sites are mainly formed by the residues from the β-subunit and the noncatalytic ones are composed mainly of the residues from the α-subunits. The structure of F_1-ATPase [9] is shown in Fig. 13.1. The F_o part of the ATP synthase is largely embedded into the mitochondrial inner membrane. It contains multiple (9 to 14, varies with species) copies of the subunit *c*, each of which, in turn, harbors proton binding sites [4,6–8] .

A binding change mechanism [4] was proposed by Boyer for the catalytic function of F_oF_1-ATP synthase, which explains a large body of experimental observations, in particular, the primary use of energy input for the release of the synthesized product, ATP, and the different binding affinities at different sites, as well as the positive cooperativity in the catalytic process. Two key elements in the binding change mechanism are the asymmetry of the F_1-ATPase and the rotation of the γ-subunit. Experimental evidence was obtained for both of them, from X-ray crystallography [9,10] and single-molecule studies [11–15], respectively, about 20 years after the original proposal of the binding change mechanism [16].

The otherwise identical β-subunits were shown in X-ray structures to adopt different conformations and to be occupied by different substrates [9,10], one by an ATP analogue (β_{TP} site), another one by an ADP (β_{DP} site), and still another one either by an ADP+sulfate (β_{HC} site) or being empty (β_E site). Both β_{TP} and β_{DP} sites adopt a closed conformation with a well-formed nucleotide binding site. The β_E site takes an open conformation, and the β_{HC} site takes a 'half-closed' conformation, which is underway between that of the β_{TP} and β_E sites. According to the binding change mechanism in ATP synthesis, the proton gradient across the inner membrane of mitochondria is believed to drive the rotation of the *c* subunits of F_o (see below) which in turn triggers the rotation of the γ-subunit

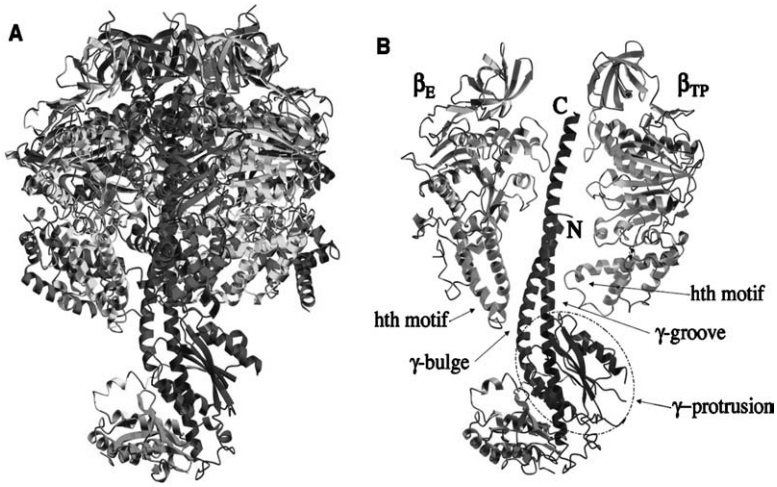


Fig. 13.1. Crystal structure (A) of F_1-ATPase. A comparison between the open catalytic site β_E and one of the closed sites β_{TP} is shown in (B) (see Color Plate 21).

in the F_1-ATPase with respect to the β-subunits [4,7,9,11]. The rotation of the γ-subunit drives the conformational changes of the β-subunits in a sequential way and at last leads to the synthesis of ATP from ADP and Pi, as well as to the subsequent ATP release. The isolated F_1 part of the ATP synthase maintains the capability of ATP hydrolysis. In the case of ATP hydrolysis, the γ-subunit is believed to rotate in the direction opposite to that in the case of ATP synthesis [9,11].

On the other hand, it is known that the mitochondria F_o portion is composed of a ring of 10 identical *c* subunits, whose structure, when isolated in solution, has been determined by NMR [17], plus several other subunits, at least some of which (e.g., the *a* subunit) appear to be involved in the coupling between F_o and F_1. A recent crystal structure [18] obtained for the rotor ring of F-type Na^+-ATPase from *Ilyobacter tartaricus* showed that the 11 *c*-subunits form a hourglass-shaped cylinder with 11-fold symmetry. Sodium ions were found to be bound in a locked conformation close to the outer surface of the cylinder near the middle of the mebrane. This 2.4 Å resolution structure suggests that the binding site opens up toward subunit *a*, believed to be part of the stator, and forms the ion channel. From single molecule experiments on F_1-ATPase [11], it has been known for some time that, in the process of hydrolysis, the rotation of the γ-subunit is counter-clockwise as viewed from the membrane and, as we have mentioned above, it has been assumed in simulations [19,20] that a clockwise rotation would be required for synthesis.

Using labeled F_oF_1-ATPsynthase incorporated into liposomes, Diez *et al*. [21] performed single-molecule FRET measurements which demonstrated that, indeed, the γ-subunit rotates in opposite directions during synthesis and hydrolysis. Itoh *et al.* [22] showed for an isolated F_1-ATPase molecule that rotating the γ-subunit with an attached magnetic bead in the clockwise direction by applying an external magnetic field resulted in ATP synthesis, confirming the predicted rotation direction. More recent experiments have revealed that the rotation of the γ-subunit induced by an external mechanical force is tightly correlated with the synthesis of ATP in the presence of an intact ε-subunit [23]. It has been demonstrated for the F_oF_1-ATP synthase complex [24] that the average output torque is large (50–60 pN nm), which is essentially the same as for the isolated F_1-ATPase [15]. Besides, it has been shown that variations of the torque are small, and that the output free energy of the loaded enzyme is a linear function of the rotational angle. Although models of how the two rotary motors are coupled have been published [25–27], the lack of high resolution structural data on the stator portion of the F_oF_1 complex have so far prevented a definitive analysis.

F_1-ATPase hydrolyzes ATP to ADP/Pi under optimum conditions with a k_{cat} value of 40 and 600 s^{-1}, depending on the organism and mutant type [4,6–8]. Unisite ATP hydrolysis by the *E. coli* F_1-ATPase (EcF1) has a rate constant of only 1.2×10^{-3} s^{-1} [28], with the product release being rate limiting. The rate enhancement for multisite ATP hydrolysis arises from two complementary contributions. The first is the separation of the chemical step from the product release step, with a speed-up of the latter by multisite catalysis; i.e., the catalysis takes place in one conformation of the γ-subunit, whereas the product releases in another one. In addition, the chemical step is accelerated by at least a factor of 600, relative to its rate in the unisite catalysis, to achieve the overall observed rate of 40 to 600 s^{-1}; the exact value is not known because only overall rates have been measured. Interestingly, the rotary catalysis accelerates both the chemical step and the nonchemical step (i.e., the product release).

Based on the binding change mechanism of Boyer [4,7], several insightful pictorial models for the function of F_1-ATPase have been proposed [4,7,8,10]. However, the binding change mechanism by itself does not provide the underlying mechanism by which the different subunits move cooperatively and how these motions lead to efficient ATP synthesis. In fact, there has been disagreement on whether all three or only two catalytic sites need to be occupied for the optimal function of the enzyme. Experimental data have been interpreted to support either the tri-site [7] or the bi-site mechanism for ATP hydrolysis by the F_1-ATPase [29]. In a series of papers, a more detailed mechanistic model using a master equation approach, based on empirical energy functions extrapolated from experimental measurements, was successfully built by Oster and coworkers for F_1-ATPase, as well as for F_oF_1-ATP synthase [30,31]. In the present review article, we focus on the thermodynamics and kinetics of ATP hydrolysis and synthesis F_oF_1-ATP synthase, as well as on the recent efforts in identifying the nucleotide binding site by linking the available experimental thermodynamic and structural information with the help of molecular dynamic simulations, and on showing how these results lead to a trisite mechanism for the ATP hydrolysis and synthesis by F_oF_1-ATP synthase.

13.2 UNDERSTANDING THE STRUCTURES USING MOLECULAR DYNAMICS SIMULATIONS

Two molecular dynamics simulations [19,20] have provided important insights concerning the F_1-ATPase mechanism. Both calculations were carried out with the rotation of the γ-subunit in the predicted synthesis direction (see Refs [21] and [22] for the experimental support of such an assumption), because it is easier to simulate the effect of the γ-subunit rotation on the β-subunit conformational change than to proceed in the reverse direction. The simulation results have demonstrated how the rotation of the γ-subunit induces the structural changes in catalytic β-subunits and explained why there is much less movement in the loaded with the ligand, but catalytically inactive, β-subunits. Both van der Waals (steric) and electrostatic interactions contribute to the coupling between the γ- and the β-subunits; the dominant electrostatic interactions occur between positive residues of both the coiled-coil portion and the globular region of the γ-subunit (the 'ionic track'), as well as negatively charged residues of the β-subunits (see Fig. 13.4 of Ref. [20]).

The importance of some of the residues involved has been explored by mutation experiments; for a recent study, which reviews some of the earlier data, see Ref. [32]. The simulations show how the rotation of the γ-subunit induces the opening motion of the β-subunits; the closing motion, once there are no steric interactions due to the γ-subunit preventing it, appears to be spontaneous; see, also Ref. [33], which describes the simulation of an isolated subunit. An ionic track of Arg and Lys residues on the protruding portion of the γ-subunit was shown to play a role in guiding the motion of the β-subunits.

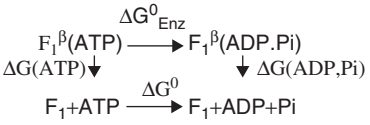
Furthermore, a recent simulation [34] was done for a β-sheet in isolation and molecular dynamics simulation of an isolated β-subunit [33] confirms the results of simulations of the entire F_1-ATPase [19,20], which indicate that the closed state is the stable one, relative to the open state, even in the absence of bound substrate. All these results have important indications on how energy is stored in the conformations of the β- and γ-subunits, as well as in their interactions, and how the energy is used to drive the motion of the enzyme.

13.3 THERMODYNAMICS OF ATP HYDROLYSIS

An essential quantity used in the binding change mechanism of Boyer [4,7] is the free energy difference, ΔG^0_{Enz}, between ATP(+H_2O) and ADP+Pi, at each catalytic site of the β-subunits. Three different binding constants for ATP, here referred to as 'tight', 'medium', and 'weak', have been measured for F1-ATPase in solution, however, it has not been possible to experimentally determine the relation of these measured solution values to the three different conformations of the β-subunits observed in the crystal structures. This situation is an unusual aspect of F_1-ATPase; in most other motor proteins (e.g., kinesin and myosin), separate structures with different binding constants exist in solution and in crystals, and so the identification is more straightforward. It is generally agreed that the 'open' subunit (β_E) contains the weak-binding site, but there has been no consensus on which of the two other subunits (β_{TP} and β_{DP}) harbors the tight-binding site, and which one the medium affinity site.

Clearly, to determine the role played by each of the β–subunits in the rotary catalysis mechanism of F_1-ATPase, it is essential to resolve the uncertainty concerning their binding affinities; that is, to make a connection (the 'missing link') between the microscopic (structural) and macroscopic (solution) data. This has been achieved recently [35] by combining 'alchemical' free energy difference simulations with experimental data. The free energy simulations were used to calculate the standard free energy change, ΔG^0_{Enz}, of the hydrolysis reaction in the various β-subunit sites of the crystal structure. The bound ATP/H_2O was calculated to have a free energy similar to that of ADP/Pi in the β_{TP} site, whereas the free energy in the β_{DP} site strongly favors ADP/Pi relative to ATP. Experiments have shown that, under unisite hydrolysis conditions (i.e., at low ATP concentrations where only one ATP is bound to the enzyme), the free energies of ATP/H_2O and ADP/Pi in the occupied site are nearly the same; the measured ΔG^0_{Enz} value is -0.6 kcal/mol in the *E. coli* enzyme [36]. Because unisite hydrolysis is expected to take place in the tight site, the simulation results can be used to identify the β_{TP} and β_{DP} subunits as the ones with the tight and medium ATP affinity, respectively. Given these identifications, experimental measurements for EcF_1 of the binding constants for ATP and ADP bar Pi are used to obtain quantitative values of ΔG^0_{Enz} for each of the sites in the latter enzyme.

This was done by employing the thermodynamic cycle shown in Scheme 13.1 where $F_1{}^\beta$ refers to each of the catalytic subunits β_{TP}, β_{DP}, β_E, or β_{HC} identified in the crystal structures. The quantity ΔG^0 is the standard hydrolysis free energy of ATP in solution under the presence of Mg^{2+} in its physiological concentrations ($\Delta G^0_{Enz} = -7.3$ kcal/mol for $[Mg^{2+}]=3$ mM [37]). The quantities, ΔG(ATP) and ΔG(ADP,Pi) are the standard binding free energies of ATP and ADP+Pi at that given β site under consideration. Based on the nucleotide concentration dependence of the number of occupied sites, Senior and coworkers [38] obtained three ATP binding constants for EcF_1; they are <0.1 μM, around 0.5 μM and around 25 μM. Using the same technique, the binding constants for ADP, in



Scheme 13.1

the presence of 5 mM Pi, were determined to be 0.05 μM, 29 μ M and 29 μM [39]. Gruber and Capaldi studied ATP binding to EcF_1 with the β and γ-subunits cross linked, so that no multi-site hydrolysis could occur [40]; they obtained the values of <0.1 μM, 2 μM, and 30 μM, respectively. Also, the binding constant of ATP at the tight-binding site in EcF_1 has been determined directly to be about 0.2 nM by using radioactive ATP [4,36]. All these experimental values were obtained at 2.5 mM Mg^{2+}, the physiological concentration. For the dissociation constant of Pi from a β-site, a lower limit of 10 mM for all three sites was estimated for EcF_1 by Senior and co-workers [7].

The measured ATP binding affinity is 5.5 kcal/mol greater in the tight site (K_D = 0.2 nM [7]) than in the loose site (K_D = 2 μM [40]). Combining this value with the fact that the calculated ΔG^0_{Enz} decreases by 10.6 kcal/mol in going from the tight to the loose ATP site (i.e., the calculated values are 1.4 and −9.2 kcal/mol, respectively; see Table 13.1), we find that the binding affinity for ADP+Pi is about 5.1 kcal/mol greater in the β_{DP} site than in the β_{TP} site. These results suggest that the β_{DP} is the tight site for ADP and the loose site for ATP, and we thus obtain for this site $\Delta G^0_{Enz} = -9.8$ kcal/mol, again consistent with the simulation result (−9.2 kcal/mol [35]). This agreement between the calculated and estimated results confirm that the tight-binding site for ATP is the loose-binding site for ADP and *vice versa*. This result plays an essential role in the proposed mechanism for F_1-ATPase hydrolysis, as described below. The ΔG^0_{Enz} values of the β_E and β_{HC} sites were estimated in a similar way, which are about −7.8 kcal/mol (no simulation was made for this site because of the uncertainty of placing the ligands in an empty site) and −14.4 kcal/mol (simulation result is −12.7 kcal/mol [35]), respectively.

Having the binding sites identified and binding affinities of substrates assigned, the free energy change during ATP hydrolysis can be described as follows:

- An ATP binds to the empty (β_E) site with a free energy change of −2.8 kcal/mol, as determined from K_D = 25 μM and the cellular concentration of ATP, which is equal to 3 mM [41].
- Then, the conformation of the β_E site changes to that of the β_{TP} site, due to the free energy stabilization of about −7.0 kcal/mol, corresponding to the decrease of K_D from 25 μM to 0.2 nM.
- The inward movement of the β-subunit involved in this conformational change drives the γ-subunitrotation, as suggested by simulations [19,20]; i.e., it is the first source of energy coupling that is anessential part of the rotary binding change mechanism.

The free energy difference between ATP and ADP+Pi is nearly zero at the β_{TP} site, so that ATP hydrolysis could begin once the β_{TP} site is formed. Meanwhile, a conformational change towards that of β_{DP} is likely to be involved, since the hydrolysis rate is faster than in the unisite catalytical process. With the γ-subunit rotation, which is induced by the change of β_E to β_{TP}, the former β_{TP} site is transformed to a β_{DP} site.

During this process, the ATP hydrolysis reaction is expected to be essentially complete, since ΔG° in the β_{DP} site is −9.8 kcal/mol. Because the β_{DP} site has a higher binding affinity for ADP+Pi than β_{TP}, the transformation of β_{TP} to β_{DP} provides a second driving force for γ-subunit rotation; the free energy stabilization is about −4.4 kcal/mol. The rotation of the γ-subunit results in the opening of the β_{DP} site and to release ADP, Pi. This process requires about 2.6 kcal/mol, which is the binding free energy of ADP, Pi at the β_{DP} site.

Table 13.1 Binding constants K_D, binding affinities (kcal/mol), and standard free energy differences between ATP and ADP+Pi (kcal/mol) at various binding sites

	β_{TP}	β_{DP}	β_E	β_{HC}
K_D for ATP	0.2 nM	2 μM	25 μM	5 mM
(affinity)	(−13.24)	(−7.78)	(−6.28)	(−3.14)
K_D for ADP	29 μM	0.05 μM	29 μM	25 μM
(affinity)	(−6.19)	(−9.96)	(−6.19)	(−6.27)
K_D for Pi	0.5 M	0.5 M	0.5 M	1 mM
(affinity)	(−0.37)	(−0.37)	(−0.37)	(−4.09)
ΔG°_{Enz}(estimated)	−0.63	−9.8	−7.8	−14.3
ΔG°_{Enz}(calculated)	1.4	−9.2	—	−12.7

13.4 IMPLICATION ON THE KINETICS OF ATP HYDROLYSIS BY F_1-ATPASE

One of the most important findings of the thermodynamic analysis mentioned above is that the strong-binding site for ADP, Pi is the β_{DP} site, in contrast to the strong-binding site for ATP, H_2O, which is the β_{TP} site. The thermodynamic analysis was used to construct a free energy diagram for each subunit as a function of the γ-rotation angle. This showed how the cooperativity between the subunits during the rotation cycle leads to the increased rate of rotation observed at high concentrations of ATP. Further, it has made possible the development of a chemical kinetic model for ATP hydrolysis, in which the rate constants are associated with specific configurations of the β-subunits.

The model, which indicates that all three catalytic sites are usually occupied under optimum hydrolysis conditions (a so-called 'tri-site' mechanism), has explained a series of experimental data, including the ATP concentration dependence of the rate of hydrolysis and catalytic site occupation for both the *Escherichia coli* F_1-ATPase (EcF_1) and thermophilic *Bacillus PS3* F_1-ATPase (TF_1), which exhibit different behavior. The kinetic model was able to explain the simple Michaelis–Menten behavior of thermophilic F_1-ATPase (TF_1), which had been used as the evidence supporting the bi-site mechanism [13,31]; during steady-state ATP hydrolysis, the β_{DP} site is occupied by ADP/Pi instead of ATP and therefore the TF_1 functions under a tri-site mechanism even at ATP concentrations as low as several nanomolar. It was concluded that the ATP concentration dependence of the enzyme activity itself does not provide a way of distinguishing a bi- from tri-site mechanism.

Another set of single molecule experiments [12] provides data that can be used to test certain aspects of the model described above. A fluorescent analogue (Cy3-ATP) of ATP was used to follow the binding of the ligands to single molecules during the hydrolysis cycle. This has been facilitated by the fact that the reaction is significantly slower with Cy3-ATP than ATP. It was possible to ascertain that a given substrate molecule (Cy3-ATP), or its product form (Cy3-ADP), remains attached until the γ-subunit has rotated by at least 240°; regretably, the fluorescent tag does not distinguish between Cy3-ATP and Cy3-ADP.

Further, in the so-called waiting state for ATP binding at low ATP concentration, only one of the three sites was empty, in accord with the tri-site mechanism. Both of these

conclusions are in agreement with the model in Ref. [42]; see, in particular, Figure 13.2 of Ref. [42] and its discussion. As in earlier studies by the Kinosita group [13,14], two pauses occur during the 120° rotation cycle of the γ-subunit, at an angle of 90° in the experiment with a wild-type system [13], and at 80° in modified systems (either mutant F_1-ATPase or modified substrates or both, all of which slow down the reaction by more than an order of magnitude) [12,14]. In none of the studies was it possible to identify the processes that occur during the two pauses. The total product (ADP and Pi) release or hydrolysis plus release of one part of the product (e.g., Pi) were suggested as plausible possibilities in several papers [12–14]. The thermodynamic/kinetic model [42] assumes the total product release, albeit this is not an essential aspect of the model. Although the exact nature of the pauses was not identified, and when and where the ADP and Pi are released during the cycle was not determined by direct experiments, as will be seen later, the analysis of the inhibition results suggests that the release of ADP and Pi are the rate-limiting steps.

13.5 THERMODYNAMICS IN ATP SYNTHESIS

In turn, the identification of the binding sites for nucleotides and the thermodynamic information obtained for ATP hydrolysis helps us in understanding the mechanism of ATP synthesis by F_oF_1-ATP synthase. F_oF_1-ATP synthase is capable of catalyzing either ATP hydrolysis or ATP synthesis in the presence of similar concentrations of ATP and ADP, Pi. What the enzyme does is determined by whether or not there is a large enough external free energy input. In the presence of a large proton motive force the enzyme catalyzes ATP synthesis, otherwise it catalyzes ATP hydrolysis [4,7]. Based on analysis of binding affinities of nucleotides and Pi it was speculated [42] that there exist different binding sites for the substrates in ATP hydrolysis (ATP) and synthesis (ADP, Pi): the β_E site is more likely to be the site for ATP binding in ATP hydrolysis and the β_{HC} site be the site for ADP and Pi binding in ATP synthesis.

The challenge that F_oF_1-ATP synthase faces in catalyzing the synthesis of ATP is not only to accelerate the reaction rates, which most enzymes do, but also to couple the chemical reaction to mechanical processes in order to synthesize chemical compounds that are thermodynamically unstable from reactants that are more stable in solutions. This remarkable enzyme fulfills its function by making use of a series of conformation changes induced by externally applied forces (proton motive force or forces applied *in vitro*), each serves its unique purpose.

The synthesis of ATP starts with the binding of reactants, ADP, Pi, to an open nucleotide binding site in the F_1-ATPase. In the absence of proton motive force, the open β-subunit takes a β_E conformation, which binds dominantly ATP and would not serve for the purpose of ATP synthesis. The proton motive force drives the rotation of the γ-subunit in the synthesis direction and, as a result, changes the conformation of the open nucleotide binding site to a half-closed conformation, β_{HC}. As we learned from the thermodynamic analysis for the binding of nucleotide to the different conformations of the β-subunits, the β_{HC} site binds ADP, Pi and, to be more efficient for the reactant binding, it binds ATP only weakly. The binding of ADP, Pi is both thermodynamically and kinetically favorable at this site. With ADP, Pi bound to the β_{HC} site, further rotation of the γ-subunit changes the conformation of

this site into a β_{DP}, which is the tight site for ADP, Pi. This site also isolates the bound ADP, Pi from the solution, so that the release of ADP, Pi also becomes kinetically slow.

In the next step, the β-subunit changes to a β_{TP} conformation, again driven by the rotation of the γ-subunit. Although the β_{TP} is a loose site for ADP, Pi binding, ADP, Pi release from this site is slow due to the isolation of its nucleotide binding site from solution. In the mean time, during the conformational changes of $\beta_{HC} \rightarrow \beta_{DP} \rightarrow \beta_{TP}$, the binding affinity of the β site for ATP increases. Following this sequence of conformational change of the β-subunit, the free energy of the ATP bound to this site is lowered with respect to that of the bound ADP, Pi. In the β_{TP} site, ATP and ADP, Pi have similar free energies, so that ATP synthesis occurs.

Free energy simulations [35] suggest that ATP synthesis occurs in a conformation similar to β_{TP}, but with the local structural changes induced by rotation of the γ-subunit to about 220°, referred as β_{TP*} in the following. The sensitivity of the reaction free energy to small displacements of the large number of charged residues in the active site is one of the key features of the catalytic β-subunits of F_1-ATPase; e.g., the small difference between the very similar structures of the β_{TP} and β_{DP} sites has been shown by free energy simulations to result in a difference in the hydrolysis reaction free energy of as much as 9 kcal/mol, with the residues αArg373 and βArg189 making the most important contribution [35].

The role of the residue αArg373 in F_1-ATPase hydrolysis, which is the only α-subunit residue contributing to the β-subunit catalytic site, has been suggested [43,44] to be analogous to that of the so-called 'Arginine Finger' in GTPase [19]. Structural studies have shown that αArg373 can be at a distance of 10 Å from the phosphate in the β_{TP} site [35,44], in contrast to its position in other structures [35,45,46]. Also, if αArg373 is mutated, the hydrolysis is slowed down significantly and the synthesis is eliminated [43]. Calculations of ATP unbinding from the catalytic site confirm that αArg373 moves away when ATP is no longer present, and that αArg373 is essential for biasing the reaction free energy in the direction toward hydrolysis, while at the same time, lowering the activation barrier [35]. Conversely, a displacement of αArg373 from the active site by a conformational change, due to the rotation of the γ-subunit, shifts the thermodynamic potential so as to favor ATP synthesis.

Meanwhile, the residue βArg189, whose equivalent is not present in GTPase, has also been found to be important for ATP synthesis [4,43]. Overall, the Arg pair (αArg373 and βArg189) and their relative positions pay the major contribution to switching between hydrolysis and synthesis. Structural studies [44] and simulations suggest that αArg373 acts to 'transmit conformational signals across the α/β-subunit catalytic interface' [47], as a part of the chemomechanical coupling to the γ-subunit. In β_{TP*}, the residue βGlu188 is in approximately the same position, relative to the ATP γ-phosphate group, as in the β_{TP} structures [9,45]; it polarizes the water molecule that is involved into the synthesis reaction in β_{TP*} and into the hydrolysis reaction in β_{DP}.

Along with the βGlu188, some other residues like βLys162 are essential for the catalysis of synthesis and hydrolysis; small displacements of these residues can significantly change the reaction rate so these play an important role in the kinetic control. For instance, during the ATP synthesis, after the formation of ATP at the β_{TP*} site, the displacement of catalytic residues prevents ATP from being hydrolyzed to ADP, Pi with high chemical potential. Even the mutation of charged residues not in direct contact with the ligand, such as βArg246, has been shown recently to promote loss of activity [48].

This illustrates the intricacy of the residue contributions to the catalytic region of the β-subunits required to fulfill the unique functions of regulating absolute binding, relative binding, chemical catalysis and conformational signal change.

13.6 ATP HYDROLYSIS VERSUS ATP SYNTHESIS BY F_OF_1-ATP SYNTHASE

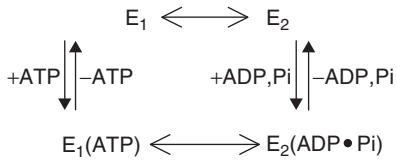
13.6.1 An $E_1 E_2$ mechanism

In this section we describe a unified model for ATP hydrolysis and synthesis by F_oF_1-ATP synthase where two different conformations of the enzyme participate in the catalysis [50]. This model is a minimal one, and it is possible that additional intermediate conformations are involved, but they seem to be not important in the explanation of the available data. It is likely that, as discussed above, the two conformations in question correspond to the two X-ray structures of bovine heart mitochondrial F_1-ATPase [9,10]. In one structure, the β_{TP} site is occupied by an ATP analog (AMPPNP), the β_{DP} site is occupied by an ADP, and the β_E site is empty. This structure is the one to which ATP would bind as a reactant for ATP hydrolysis. The other structure has ADP.AlF_4^{-1} bound at both β_{TP} and β_{DP} sites, and ADP+sultate at the β_{HC} site.

The β_{DP} and β_{TP} sites are very similar in the two structures and have similar properties as shown by free energy simulations [35]. However, the β_{HC} ('half closed') site is very different from the β_E site. In particular, it has a binding site for Pi and is assumed here to be the structure involved in binding the reactants, ADP and Pi, from the solution in the synthesis reaction. This assumption is in accord with the conclusion [42], based on the free energy simulations and a comparison with experimental measurements of substrate binding in ATP hydrolysis and synthesis, that there exist different binding sites for ATP and ADP, Pi. It was shown experimentally [7] that the dissociation constant of ATP increases from tens of micromolar in the hydrolysis condition to several millimolar in the synthesis condition and that for Pi decreases from >10 mM to about 1 mM. These experimental results indicate that there exist different binding sites of reactants according to whether ATP hydrolysis or synthesis is the dominant process, consistent with the thermodynamic analysis given above.

Therefore, as discussed earlier, the β_E site is likely to be the ATP binding site during the tri-site ATP hydrolysis, whereas in the presence of a large enough proton motive force, where ATP synthesis dominates, this site is transformed into a β_{HC} site and binds ADP/Pi. The β_{HC}-subunit binds ADP and Pi, because in its bonded state Pi has a lower free energy than in solution; also the rate of binding is expected to be relatively fast. (The value for the ADP constant for *E. coli* F1-ATPase is about $10^6\ M^{-1}s^{-1}$ [7]; but any estimate of that for Pi is not known). Since the potential of ATP in the β_{HC} site is lower than that in solution, there is little interference (inhibition) from ATP binding, even when it is present at a concentration similar to that of ADP, Pi.

The rotation of the γ-subunit by 90°, after ADP and Pi are bound, transforms β_{HC} into β_{DP}; this conformational change does not induce synthesis, since the reaction free energy still strongly favors ADP, Pi. A further rotation by 120° changes β_{DP} to β_{TP}, the high affinity site for ATP. The free energies of bound ATP and ADP, Pi are approximately equal at β_{TP}, so that the synthesis can begin, but its rate would be much slower ($0.04\ s^{-1}$ in *E. Coli*)



Scheme 13.2

than the observed maximal rate (10–100 s^{-1} in *E. Coli*) [7]. A conformational change is required to shift the equilibrium toward ATP, and increase the rate of the synthesis.

The rotation of the γ-subunit to complete the 360° cycle creates the β_E site, which binds ATP less strongly, as required for product release. As the β_E site is approached, the free energy of ATP becomes higher than that of ADP, Pi. A lower value of the free energy of ATP in the β_E site, relative to that in solution, is required for optimization of hydrolysis, as well as synthesis; i.e., β_E is the binding site for the ATP substrate in hydrolysis, as well as the release site for ATP after synthesis. The openness of the β_E site makes the release and binding rate rapid enough, and thus it should not be rate-limiting under normal conditions.

The catalytic cycle for ATP synthesis and hydrolysis by F1-ATPase is illustrated in Scheme 13.2.

Two different conformations of F_1-ATPase participate in the reaction cycle, in formal analogy to the E_1E_2 model. The structure E_1 has the subunit conformations β_{TP}, β_{DP} and β_E [9], while E_2 consists of β_{TP}, β_{DP} and β_{HC} [10]. In ATP synthesis, the reaction begins with a β-subunit rotation of approximately 30°, which changes E_1 to E_2 and most importantly β_E, β_{HC}, which has a binding site for Pi as well as ADP. As mentioned earlier, binding of the substrates Pi and ADP to the β_{HC} subunit takes place without rotation.

Once the substrates are bound, a 90° rotation occurs, which changes E_2 to E_1. ATP synthesis occurs during this step and is followed by product release from β_E. (In ATP hydrolysis, the ATP substrate binds to the β_E subunit of E_1 and the hydrolysis products are released from the β_{HC}-subunit of E_2). In synthesis, the rotation induced by the proton motive force (or equivalently an external force [22,23]) creates the ADP, Pi binding pocket in β_{HC}, lowering the reaction barrier (see below) and accelerating the release of ATP by forming β_E. In hydrolysis, the rotation of the β-subunit is produced by the differential binding of ATP and ADP, Pi to the various subunit conformations.

The direction and rate of the reaction is primarily determined by the magnitude of the proton motive force. This property of the E_1E_2 mechanism is essential for an enzyme like F_oF_1-ATP synthase, which works in both directions with high efficiency, in that the maximum rate constant can be obtained for synthesis and hydrolysis at essentially the same cellular concentrations of ATP, ADP, and Pi (i.e. [ATP] ~ 3 mM, [ADP] ~ 0.4 mM, [Pi] ~ 6mM, and [Mg^{2+}] ~ 3 mM [24]).

13.6.2 Kinetics

The above-described model was successfully used to study various kinetic properties of the enzyme under study. Importantly, the K_M (Michaelis–Menten constant) values of the reactants also decrease with the increase of the solution concentration of the products. This

Table 13.2 Inhibition parameters

	K_i (Pi)	K_i (ADP)	K_i (ATP)
Experimental	>10 mM	<100 μM	5 mM
Established	900 mM	30 μM	7.5 mM
	[Pi]	[ADP]	[ATP]
Cellular Concentration	~6 mM	~0.4 mM	~3 mM

result is an essential property of the enzyme, which makes the K_M value of each ligand (ATP, ADP, and Pi) low enough, so that the maximum rates of both ATP hydrolysis and synthesis can be achieved. The K_M values of all three ligands are all smaller than their cellular concentrations. In contrast to the E_1E_2 mechanism proposed here, competitive binding between the reactant and product at the same binding site would allow at most only one of the two reactions to achieve a maximum rate constant at any given concentrations of ligands. Such a mechanism would not describe the biological function of F_oF_1-ATP synthase. Nor does it agree with the experimental observations, which show that the K_M values for ADP and Pi (29 μM and 1 mM) in ATP synthesis and that for ATP (29 μM) in hydrolysis are all lower than their cellular concentrations.

At negligible ADP concentrations, the inhibition concentration of Pi, K_i(Pi), at which the ATP hydrolysis rate is one half of its maximum value, is about 900 mM and agrees with the experimental lower limit of 10 mM. Similarly, at low Pi concentrations, the K_i(ADP) is hown to be 30 mM, again in agreement with the experiments (<100 mM). The estimated inhibition concentration of ATP, K_i(ATP), at which the maximum rate of ATP synthesis is one half of its value. K_i(ATP) is estimated to be 7.5 mM, while the experimental value is 5 mM. All products (ATP in synthesis and ADP, Pi in hydrolysis) inhibit the reaction; i.e., the higher the concentration(s) of the product(s) the smaller the maximum rate constant. In addition, the present model gives inhibition concentrations for both ATP hydrolysis and synthesis, which are consistent with the experimental results (Table 13.2). From the analysis, it follows that at cellular conditions, only ADP plays a significant role in regulating the function of the enzyme through product inhibition, which is consistent with experimental observations.

The ATP and Pi cellular concentrations are lower than their inhibition concentrations and thus are much less important. The different role of ADP from ATP and Pi is consistent with the fact that ADP has a concentration more than 10 times lower than ATP and Pi and thus, its change leads to a sensitive regulation of the enzymatic activity and has only minor influence on the concentrations of the other two. As demonstrated above, we note that the structure-based kinetic model agrees with the available data and provides predictions that can be tested by experiments. A prediction of the model is that not only the maximum ATP synthesis rate, but also the Michaelis–Menten constant for ADP and Pi, decrease with increasing ATP concentration.

13.7 SUMMARY

Our understanding of how biomolecular motors function and of their role in the machinery of the cell is advancing rapidly. As we learn more, the wonderful complexity and

effectiveness of these motors as parts of cellular networks becomes all the more impressive. The description given here of the current understanding being achieved for F_1-ATPase by both theoretical and experimental advances is expected to be of use in the analysis of other motor proteins. For example, the kinesin activity at saturating concentrations of ATP, like F_1-ATPase, has been found to be linearly rather than exponentially dependent on the external load, which suggests that the kinesin also uses a nearly constant free energy gradient when it translocates along its track [49]. Another example is sarcroplasmic reticulum calcium ATPase (Ca^{2+}-ATPase), which uses the free energy derived from the hydrolysis of ATP to pump Ca^{2+} from the cytoplasm to the inside of the sarcoplasmic reticulum and causes relaxation of the heart beat, as well as of other muscles. Ca^{2+}-ATPase belongs to the class of P-type ATPases [51], which differ from the F,V-type ATPases by the fact that ATP is not hydrolyzed directly to ADP, Pi, but that instead two chemical reactions (phosphorylation and dephosphorylation) are involved. There exists experimental evidence [52,53] that both the phosphorylation and the dephosphorylation of the Ca^{2+}-ATPase are reversible; i.e., they are associated with very small free energy changes. This suggests that Ca^{2+}-ATPase, like F_1-ATPase, optimizes its use of free energy by coupling the available free energy to protein conformational changes. The similarities between these different types of motor enzymes suggest that there exists a common strategy for the efficient use of chemical energy.

We have focused here on protein motors that have been shown to be possessed of a largely constant stoichiometry, where leakage is negligible; for example, for every ATP hydrolyzed, F_1-ATPase rotates 120° and kinesin translocates about 8 nm. Other protein motors such as myosin II [53] and dynein [54,55] appear to have variable stoichiometry and they may have different strategies for the use of chemical energy, possibly adjusted to the changing external load. These two types of motors thus appear to be different F_1-ATPase in terms of the coupling between the chemical reaction and the conformational change of the protein involved in the function. The use of different chemomechanical coupling mechanisms by different motor proteins may reflect their unique biological functions and their environment.

The analysis of several different motor proteins seems to give a general mechanism by which the chemical energy is used to do mechanical work. The chemical energy is not used directly by the enzyme, instead it is first converted to binding affinity at the transition state of the chemical reaction, and the binding affinity is further used to drive the conformational changes of the motor proteins. It is also shown in the present analysis how the free energy change is tightly coupled to the conformational changes and thus can be used for the motor protein to function against external load. A nearly linear potential energy profile is produced as a result of this tight coupling and has great advantages over potential energy surfaces resulting from weak coupling. The tight coupling between the conformational change and the free energy utilization is more important for motor proteins working at high loads (near equilibrium). Since F-ATPase is an enzyme that functions efficiently in both directions (ATP hydrolysis and synthesis), the use of chemical energy by this enzyme is optimized by having a linear relation between the potential energy and its rotation angle and by linking its ATP binding affinity to the stable conformation of the protein (the binding to the empty site), thus to the external force.

With the rapid advance of experimental techniques, our knowledge on molecular motors expands at an unprecedented speed. More details on motor proteins are becoming

available from structure and single molecular spectroscopic studies. Computational and theoretical studies at different levels of detail are expected to make important contributions to the understanding of these systems. Rationalizing of the general chemomechanical coupling mechanism of motor proteins, as well as their specific functions in cells, will inevitably help us in answering many of the fundamental questions in biological systems: how its viability is maintained, how energy flow is controlled, how the cells reproduce themselves, how biological systems fight with various diseases related to the malfunction of these proteins etc. On the other hand, in the shadow of an approaching world-wide energy crisis, constructing more efficient machines and/or machines using alternative fuels becomes more and more urgent. We hope that a better understanding of how nature designs efficient and beautifully regulated machines will help us in all these efforts.

13.8 ACKNOWLEDGEMENTS

Much of the material in this review is conceptually adapted from the articles M. Karplus and Y.Q. Gao, Current opinion in structural biology, 14, 250–259 and Y.Q. Gao, W. Yang and M. Karplus, Cell, 123, 195–205. Y Q G acknowledges the support from the Camile and Henry Dreyfus Foundation. M K is supported by an N I H grant.

13.9 REFERENCES

1 B. Alberts, Cell, 92 (1998) 291–294.
2 M. Schliwa, (Ed.) Molecular Motors**,** Wiley-VCH, Weinheim, 2003.
3 R.D. Vale, Cell, 112 (2003) 467–480.
4 P.D. Boyer, Annu. Rev. Biochem., 66 (1997) 717–749.
5 D. Nicholls, and S.J. Ferguson, Bioenergetics 3, Academic Press, Amsterdam, 2002.
6 P.D. Boyer, Biochim. Biophys. Acta, 1140 (1993) 215–250.
7 J., Weber and A.E. Senior, Biochim. Biophys. Acta, 1319 (1998) 19–58.
8 W.S. Allison, Acc. Chem. Res., 31 (1998) 819–826.
9 J.P. Abrahams, A.G.W. Leslie, R. Lutter and J.E. Walker, Nature, 370 (1994) 621–628.
10 R.I. Menz, J.E. Walker and A.G.W. Leslie, Cell, 106 (2001) 331–341.
11 H. Noji, R. Yasuda, M. Yoshida, K. Kinosita and H. Itoh, Nature, 386 (1997) 299–302.
12 T. Nishizaka, K. Oiwa, H. Noji, S. Kimura, E. Muneyuki, M. Yoshida and K. Kinosita, Nat. Struct. Biol., 11 (2004) 142–148.
13 R. Yasuda, H. Noji, M. Yoshida, K. Kinosita and H. Itoh, Nature, 410 (2001) 898–904.
14 K. Shimabukuro, R. Yasuda, E. Muneyuki, K.Y. Hara, K. Kinosita and M. Yoshida, Proc. Natl. Acad. Sci. USA, 100 (2003) 14731–14736.
15 R. Yasuda, H. Noji, K. Kinosita and M. Yoshida, Cell, 93 (1998) 1117–1124.
16 P.D. Boyer, in Membrane Bioenergetics, C.P. Lee, C. Schatz and L. Ernster, (Eds.), Addison-Wesley, Waltham, 1979, pp. 461–477.
17 M.E. Girvin, V.K. Rastogi, F. Abildgaard, J.L. Markley and R.H. Fillingame, Biochemistry, 37 (1998) 8817–8824.
18 T. Meier, P. Polzer, K. Diederichs, W. Welte and P. Dimroth, Science, 308 (2005) 659–662.
19 R.A. Böckmann and H. Grubmüller, Nature Struct. Biol., 9 (2002) 198–202.
20 J. Ma, T.C. Flynn, Q. Cui, A. Leslie, J.E. Walker and M. Karplus, Structure, 10 (2002) 921–931.
21 M. Diez, B. Zimmermann, M. Börsch, M. König, E. Schweinberger, S. Steigmiller, R. Reuter, S. Felekyan, V. Kudryavtsev, C.A.M. Seidel and P. Gräber, Nat. Struct. Biol., (2004) 135–141.

22 H. Itoh, A. Takahashi, K. Adachi, H. Noji, R. Yasuda, M. Yoshida and J.R. Kinoshita, Nature, 427 (2004) 465–468.
23 Y. Rondelez, G. Tresset, T. Nakashima, Y. Kato-Yamada, H. Fujita, S. Takeuchi and H. Noji, Nature, 433 (2005) 773–777.
24 O. Panke, D.A. Cherepanov, K. Gumbiowski, S. Engelbrecht and W. Junge, Biophys. J., 81 (2001) 1220–1233.
25 R.H. Fillingame, C.M. Angevine and O.Y. Dmitriev, FEBS Lett., 555 (2003) 29–34.
26 G. Oster and H.Y. Wang, Trends Cell Biol., 13 (2003) 114–121.
27 R.A. Capaldi, R. Aggeler, Trends Biochem. Sci., 27 (2002) 154–160.
28 A.E. Senior, J. Bioenerg. Biomemb., 24 (1992) 479–484.
29 Y.M. Milgrom, M.B. Murataliev and P.D. Boyer, Biochem. J., 330 (1998) 1037–1043.
30 H.Y. Wang and G. Oster, Nature, 396 (1998) 279–282.
31 J.H. Xing, H.Y. Wang, von C. Ballmoos, P. Dimroth and G. Oster, Biophys. J., 87 (2004) 2148-2163; J. Xing, J. Liao and G. Oster, Proc. Natl. Acad. Sci. 102 (2005) 16539–16546.
32 M.D. Greene and W.D. Frasch, J. Biol. Chem., 278 (2003) 51594–51598.
33 R.A. Böckmann and H. Grubmüller, Biophys. J., 85 (2003) 1482–1491.
34 S. Sun, D. Chandler, A.R. Dinner and G. Oster, Eur. J. Biophys., 32 (2003) 676–683.
35 W. Yang, Y.Q. Gao, Q. Cui, J. Ma and M. Karplus, Proc. Natl. Acad. Sci. USA, 100 (2003) 874–879.
36 A. E. Senior and J. Bioenerg. Biomembr., 24 (1992) 479–484.
37 L. Stryer, Biochemistry, Freeman, New York, 1995, p. 446.
38 J. Weber, S. Wilke-Mounts, R.S.F. Lee, E. Grell and A.E. Senior, J. Biol. Chem., 268 (1993) 20126–20133.
39 J. Weber, S. Wilke-Mounts, E. Grell and A.E. Senior, J. Biol. Chem., 269 (1994) 11261–11268.
40 G. Gruber and R.A. Capaldi, Biochemistry, 35 (1996) 3875–3879.
41 E. R. Kashket, Biochemistry, 21 (1982) 5534-5538.
42 Y. Gao, W. Yang, R.A. Marcus and M. Karplus, Proc. Natl. Acad. Sci. USA, 100 (2003) 11339–11344.
43 A.E. Senior, S. Nadanaciva and J. Weber, Biochim. Biophys. Acta - Bioenergetics, 1153 (2002), 188–211.
44. R. Kagawa, M.G. Montgomery, K. Braig, A.G. W. Leslie and J.E. Walker, EMBO J., 23 (2004) 2734–2744.
45 K. Braig, R.I. Menz, M.G. Montgomery, A.G. W. Leslie and J.E. Walker, Structure (London), 8 (2000) 567–573.
46 M. Diez, B. Zimmermann, M. Borsch, M. Konig, E. Schweinberger, S. Steigmiller, R. Reuter, S. Felekyan, V. Kudryavtsev, C.A.M. Seidel and P. Graber, Nature Struct. Biol., 11 (2004) 135–141.
47 K. Kinosita, Jr., K. Adachi and H. Itoh, Annu. Rev. Biophys. Biomol. Struct., 33 (2004) 245–268.
48 M.D. Greene and W.D. Frasch, J. Biol. Chem., 278 (2003) 51594–51598.
49 Y.Q. Gao, W. Yang and M. Karplus, Cell, 123 (2005) 195–205.
50 K. Visscher, M.J. Schnitzer and S.M. Block, Nature, 400 (1999) 184–189.
51 G.A. Scarborough, J. Bioenerg. Biomem., 34 (2002) 235–250.
52 W.P. Jencks, J. Biol. Chem., 264 (1989) 18855–18858.
53 K. Kitamura, M. Tokunaga, A.H. Iwanes and T. Yanagida, Nature, 397 (1999) 129–134.
54 H. Kojima, M. Kikumoto, H. Sakakibara and K. Oiwa, J. Biol. Phys., 28 (2002) 335–345.
55 R. Mallik, B.C. Carter, S.A. Lex, S.J. King and S.P. Gross, Nature, 427 (2004) 649–652.

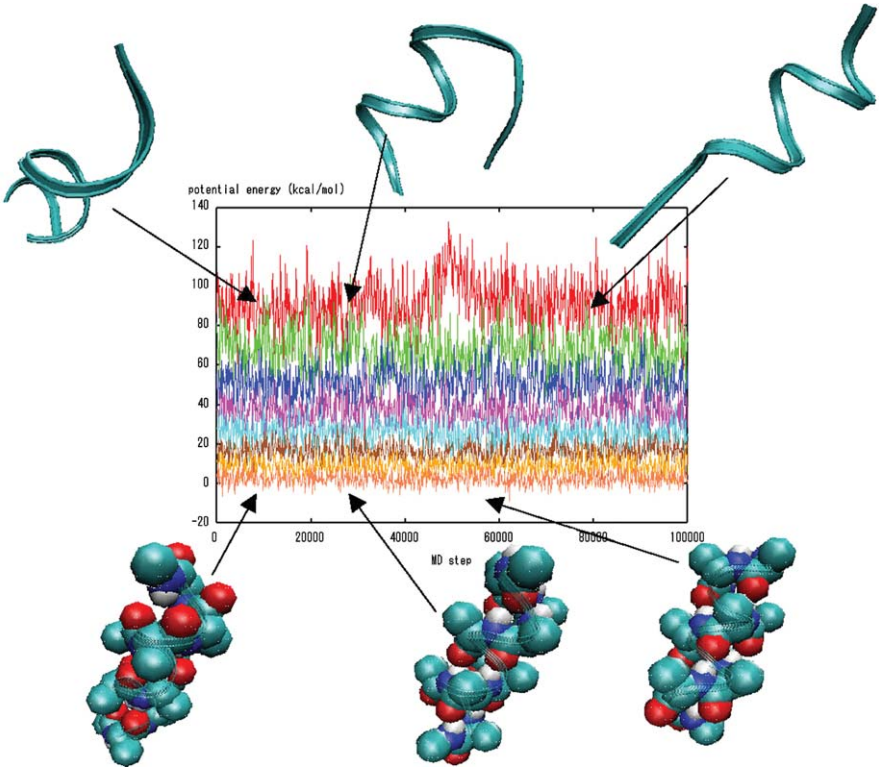


Plate 20. Time series of the potential energies and structures of $(Ala)_{10}$. NVT MD Gaussian thermostat, time step 1.0fs, trajectory 200ps (200 000 steps). Eight replicas: exponentially distributed temperatures from 168.2 K to 565.7 K (illustration appears on page 239 of this volume).

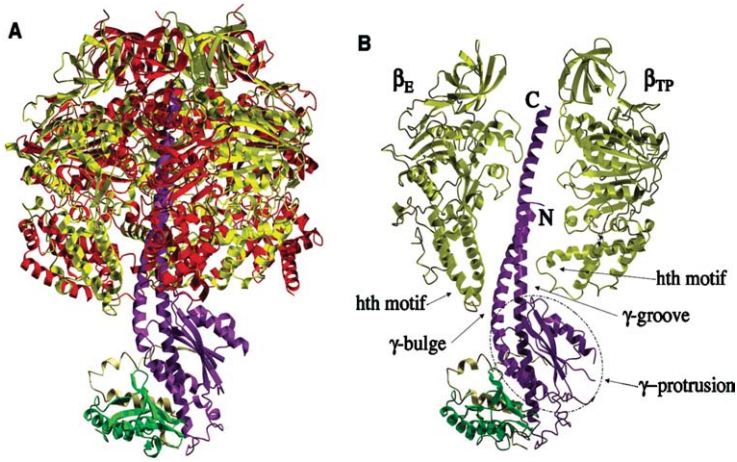


Plate 21. Crystal structure (A) of F_1-ATPase. A comparison between the open catalytic site β_E and one of the closed sites β_{TP} is shown in (B) (illustration appears on page 250 of this volume).

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 14

Monte Carlo method: some applications to problems in protein science

Petras J. Kundrotas*

Department of Biosciences at Novum Research Park, Karolinska Institutet, SE-141 57 Huddinge, Sweden

Abstract

In this chapter, Monte Carlo methods applicable to studies of protein-like chains will be reviewed and recent Monte Carlo results for flexible polypeptide chains will be surveyed.

14.1 INTRODUCTION

The term 'Monte Carlo method' is a very general concept. Monte Carlo (MC) methods are employed in a wide variety of fields including economics, finance, physics, chemistry, engineering, and even the study of traffic flows. Each field that employs MC methods has its own way of applying them, but there is always one thing in common. All MC algorithms use random numbers and probabilities in order to examine some problem and approximate outputs of simulations. The MC methods give us a simple and reliable way to investigate complex systems, which are often extremely hard (or even impossible in principle) to examine by means of other types of techniques. These methods provide approximate solutions to a variety of mathematical problems by performing statistical sampling experiments on a computer.

Monte Carlo methods were originally used under the more generic name 'statistical sampling'. The contemporary term 'Monte Carlo method' was introduced during World War II, around 1944, thus dubbing the secret work at Los Alamos which involved a direct simulation of the probabilistic problems concerned with random neutron diffusion in fissile materials. The expression 'Monte Carlo' was suggested by John von Neumann and Stanislaw Ulam [1,2], and was inspired by the gambling casinos in the city of Monte Carlo in Monaco, where the Manhattan Project team met to create their methods. Ulam was a

*Present address: Computational Biophysics and Bioinformatics, Department of Physics and Astronomy, Clemson University, Clemson SC 29634, South Carolina, USA

gambler and frequent visitor of Monte Carlo casinos, so that, by choosing this term, he would probably like to stress similarity between the statistical simulation technique and the game of chance, since any casino roulette can be considered as a simple random number generator. The analogy of Monte Carlo methods to games of chance is a good one, but one should always keep in mind that the 'gambling game' here is a physical system, and the outcome of the game is not a pot of money or stack of chips (unless simulated), but rather a solution to some interesting and important physical–chemical problem.

A number of isolated and undeveloped instances of 'Monte Carlo'-like methods emerged even much earlier than that which appeared under the Manhattan project in 1944. The earliest attempt to apply a stochastic method to some real problem dates probably back to the second half of the nineteenth century, when a number of people performed experiments, in which they threw a needle in a haphazard manner onto a board ruled with parallel straight lines and inferred the value of $\pi = 3.14\ldots$ from observations of the number of intersections between needle and lines. The theoretical background for this experiment (which is known as 'Buffon's needle') was described even a century earlier, in the eighteenth century, by Buffon [3,4]. There are many sources on the Internet about the earlier history of the Monte Carlo method and curious readers are reffered to there, for example, to Ref. [5].

This chapter gives an overview of one of the numerous branches of the Monte Carlo methods which is applicable to biological macromolecules, in particular for calculations of ionization equilibrium in proteins. The branch involved is based on the Metropolis algorithm [6] for generation of a sequence of system states (Markov chain) and has its roots in the classical statistical mechanics with a Boltzmann distribution of states (with the energy as the cornerstone parameter). After a brief introduction to the general principles of the Monte Carlo method and to some necessary (although probably boring, as the author realizes) mathematical background, the beauty and power of the Monte Carlo method will be demonstrated on only two examples (out of innumerable possible applications) of the problems emerging in protein science. It should be kept in mind that the Monte Carlo method, described in this chapter, is only a small subset of the much broader pool of the Monte Carlo methods, for which excellent overviews can be found, e.g., in the books by Landau and Binder [7,8].

14.2 BACKGROUND

14.2.1 The Monte Carlo method

14.2.1.1 Main idea of the Monte Carlo method

Very often, a physical system of interest is possessed of plenty of possible and admissible states, and it is *a priori* unknown, in which state the system can be found at a certain moment in time. It is convenient then to use the language of probabilities for description of the system's properties and to operate with average values of any quantity of interest (or observable, A)

$$\bar{A} = \frac{\sum_{k=1}^{N} A_k \times p(k)}{Z} \tag{1}$$

where N is the number of system states, $p(k)$ is a probability of the system being in the state k and A_k is the value of the observable A in the state k. The notation Z here is the normalization factor that equals, in general, to the total number of observations of the quantity A in the system

$$Z = \sum_{k=1}^{N} p(k) \tag{2}$$

When N is a huge number (very often unfeasible to calculate, as will be demonstrated below for proteins), it is highly desirable to find an approximate way to calculate $\bar{A}$. As has already been mentioned, the Monte Carlo method provides an approximate solution to this mathematical problem by performing statistical sampling experiments on a computer.

The main idea of the Monte Carlo method is to replace the full statistical average (1) by simple arithmetic average (henceforth called the Monte Carlo average of the observable A,$\langle A \rangle$) over a subset of the system states

$$\langle A \rangle = \frac{1}{N_{\mathrm{MC}}} \sum_{k=1}^{N_{\mathrm{MC}}} A_k \tag{3}$$

where N_{MC} is the number of Monte Carlo states over which the average (3) is performed. Construction of this subset is based on a Markov process, which is a random walk with a selected probability for generating a new state under the condition that this move is independent of the previous history of the system (i.e., the probability of any state of the system depends *only* on the probability of its previous state). The Markov process is used in Monte Carlo calculations, because its long-lasting repetition eventually brings the system into its most likely state. For a thermodynamic process, this means reaching an equilibrium state of the system, and, hence, the Markov process somehow mimics the way a real system reaches its most likely state at a given temperature of the surroundings.

Since $N_{\mathrm{MC}} << N$, a reasonable question arises whether the Monte Carlo average (3) is a good approximation of the statistical average (1). The probability theory (essentials of which for the Monte Carlo method can be found, e.g., in the book [7]) tells us that $\langle A \rangle \approx \bar{A}$, if the detailed balance condition holds

$$p(k) \times p(k \to k') = p(k') \times p(k' \to k) \tag{4}$$

where k and k' are any two states of the system and $p(k \to k')$ is a probability for the system to undergo transition from the state k to the state k' (transition probability). The detailed balance condition (4) implies that at equilibrium the average number of transitions from the state k to the state k' is the same as the average number of reverse transitions $k' \to k$). In other words, if transitions for a system in equilibrium obey the detailed balance condition, the system will remain in equilibrium. When the system departs from its equilibrium, then this ratio between the probabilities favours reinstatement of the equilibrium.

Another important condition ensuring reliability of Monte Carlo results is ergodicity of the Markov process. Ergodicity means that for any system state there should be non-zero probability (it could be negligibly small, though) for reaching it by the Markov process from any other system state.

In classical statistical thermodynamics, the states of a system in equilibrium are assumed to be distributed according to the Boltzmann distribution

$$p(k) = \exp\left(-\frac{W(k)}{RT}\right) \tag{5}$$

where T is the temperature measured in Kelvin (K) and R is a constant that transforms temperature into energy. Particular value, dimensions and name of this constant depend on the energy units used. For instance, if the energy is measured in Joules (J) then this constant is called the Boltzmann constant (k_B) and equals 1.3807×10^{-23} J K^{-1}. In protein science the energy is often measured in kcal/mol and R is then called the universal gas constant ($R = k_B\times N_A$ with N_A being the Avogadro number) and equals 1.9872×10^{-3} kcal mol^{-1} K^{-1} (this notation and value will be used throughout this chapter). The notation $W(k)$ stands for the energy of a system in a state numbered by the index k. The exact expression of $W(k)$ depends on the context of the considered problem and on the chosen model. An appropriate selection of $W(k)$ is state of the art in science and it is always half way to success in solving a particular task. Later in the chapter two choices of $W(k)$ expressions for two different problems in protein science will be presented and discussed in detail.

Then, the normalization factor (2) takes the well-known form of the partition function

$$Z = \sum_k \exp\left(-\frac{W(k)}{RT}\right) \tag{6}$$

the logarithm of which is the Gibbs free energy

$$G = -RT\ln Z \tag{7}$$

which describes thermodynamic properties of a classical system with plenty of states. The statistical average (1) and the detailed balance condition (4) are then rewritten in the form

$$\langle A\rangle = \frac{1}{Z}\sum_k A_k \exp\left(-\frac{W(k)}{RT}\right) \tag{8}$$

$$\frac{p(k\to k')}{p(k'\to k)} = \exp\left(-\frac{W(k') - W(k)}{RT}\right) \tag{9}$$

Assuming that the system always goes to a state with a lower energy, one arrives at the formula for the transitional probability

$$p(k\to k') = \begin{cases} 1, & \text{if } W(k') < W(k) \\ \exp\left(-\dfrac{W(k') - W(k)}{RT}\right), & \text{if } W(k') > W(k) \end{cases} \tag{10}$$

The formula (10) is the cornerstone of the Metropolis Monte Carlo algorithm [6]. This algorithm satisfies both detailed balance and ergodicity conditions, and its simplicity makes it suitable for many diverse applications, especially for those that can be described by numerous modifications of the two-state classical Ising model [9]. As will be shown below,

the range of applications for this class of statistical models (and hence the Metropolis algorithm) also includes complicated biomacromolecules such as proteins, for example.

14.2.1.2 Essentials of the Metropolis Monte Carlo algorithm

The generic step of the Metropolis Monte Carlo algorithm consists of the following steps:

- Calculate the energy $W(k)$ of the considered system in the current state k according to the chosen model of system interactions.
- Change the state of the system. For example, for proteins such a change can constitute a change in protonation (deprotonation) degree of a chosen titratable site and/or consist of a spatial rearrangement of protein atoms (conformational change).
- Calculate the energy $W(k')$ for the changed system and the transition probability (10).
- Generate a random number $0 < r < 1$.
- Accept the change of the system state, if $r \leqslant p$, otherwise reject the change.

In order to get reliable Monte Carlo results this generic step must be repeated sufficiently many times, and the quantities of interest should be calculated from time to time and stored for further averaging procedures according to the formula (3).

'Sufficiently many times' means here such an amount of observations that adding new values of the measurable quantity(-ies) to the MC average (3) will not change the average value within the limits of given accuracy. This can be illustrated on the very well-known example of evaluating the number $\pi = 3.14157\ldots$ using darts. A simple dart board consists of one square and one circular region. Furthermore, the circle touches each side of the square at a single point as shown in Fig. 14.1. By throwing darts in some random sequence, the number π can be evaluated from the formula

$$\pi = 4 \times \frac{N_1}{N} \tag{11}$$

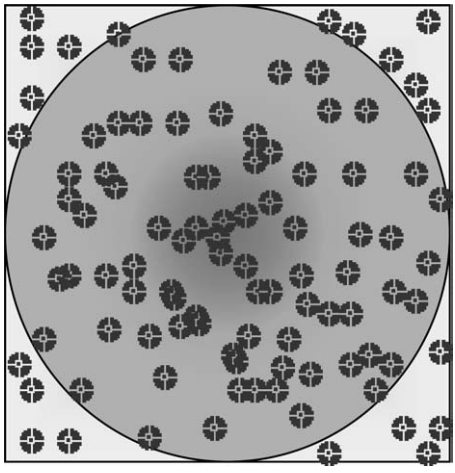


Fig. 14.1 The simple dart board after throwing 100 darts (see Color Plate 22).

where N_1 is the number of darts inside the circle and N is the total number of darts hitting the board. As seen from Fig. 14.1, throwing 100 darts is clearly not enough for obtaining a feasible π evaluation (for the situation depicted in Fig. 14.1, 79 out of 100 darts are inside the circle and hence $\pi = 3.160$). Only after the exhausting job of throwing 5000 darts can one obtain a value of π to two decimal places ($\pi \approx 3.14$) which will be not changed further by throwing more darts. This is illustrated in Fig. 14.2, where three hypothetical ways of achieving reliable evaluation of π are shown for three different persons.

Figure 14.2 also illustrates another important aspect of how to check the feasibility of acquired Monte Carlo results: the value of the final average should not depend on the pathway of achieving this value. Computer programs for Monte Carlo calculations use the so-called generators of pseudo-random numbers, i.e., a piece of computer code that generates a sequence of pseudo-random numbers controlled by a special parameter (the so-called 'seed') of such a generator. Then, the simplest way of assuring the reliability of Monte Carlo results is to repeat the calculations several times with different seeds for the generator and check whether the resulting averages differ essentially from each other or not. If such a difference is large, this is very often a sign that the considered model system has not achieved its true equilibrium, but instead got trapped into some intermediate metastable state.

To acquire reliable Monte Carlo results, one should assure statistical independence of states included in the MC average (3). The theory which gives a precise answer for the independence of states in Markov chains (the sequence of consecutively generated system states) is cumbersome, but fortunately there is always some simple advice to be followed. In this case, before calculating a quantity of interest for the sake of incorporating it into the Monte Carlo average, it is recommended to perform at least as many times of idle attempts to change system state as the considered system has changeable elements (e.g., atoms, spatial positions or titration states which can be altered during the simulations).

And, finally, several words must be said about difficulties in dealing with the algorithm described above (the world is not perfect!). One of the main obstacles in applying the Metropolis Monte Carlo (MMC) algorithm to complex systems is the problem of metastable states. When the system under study has plenty of states with energy values close to the minimum value of $W(k)$ (local energy minima), then straightforward application of the MMC

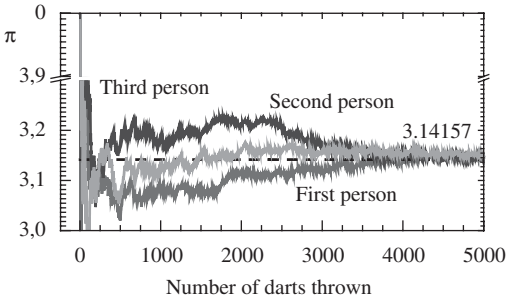


Fig. 14.2. Three possible pathways of achieving the feasible π estimation for π.

algorithm can lead to trapping of the system in one of those states, especially at low temperatures ('low' means here that the energy of thermal fluctuations is considerably smaller then the height of an energy barrier between local minima which results in a low acceptance ratio p (10) for transitions between the local minima). In the examples described below, this problem is surmounted by the so-called simulated annealing method, when, at a certain temperature, several Monte Carlo runs with different initial configurations are performed, and the final average is estimated both within every single MC run and among them. Another alternative to this is the so-called method of expanded ensembles [10], when a change in a system state implies that even those parameters can be rendered variables, which in the conventional MC algorithm are maintained constant, such as temperature, external fields, etc. (a subset of these methods with variable temperature is known as the simulated tempering algorithm [11]). It should be noted however, that the problem of sampling efficiency for the different MC algorithms is still under debate.

14.2.2 Some other necessary background

14.2.2.1 Electrostatic energy of a microscopic state of a protein

Any large unfolded protein has many rotational degrees-of-freedom (in the simplest case, rotations around the C–C_α and N–C_α bonds of a backbone) and can adopt many different spatial configurations. From the statistical physics point of view, each of those configurations can be considered as a microscopic state of a system uniquely described by a set of state variables. Let us, for the sake of simplicity, consider here the case when only titratable residues contribute to overall electrostatic interactions. Then, a microscopic state of a protein with N_t titratable groups can be represented by two sets of state variables. The first set reflects the fact that electrostatic energy is dependent on mutual allocation of interacting charges and comprises a set of spatial coordinates for titratable groups, $\{\mathbf{r}_1...\mathbf{r}_{N_t}\}$. Each vector $\mathbf{r}_i$ (hereafter index $i = 1, 2, \ldots, N_t$ counts the titratable residues) contains all the information about the spatial position of ith residue, for example, three Cartesian coordinates. The second set stands for charge states of titratable residues. It is convenient to assign a deprotonation variable x_i to an ith group so that the variable $x_i = 0$ when a proton is bound to a titratable site of ith residue and $x_i = 1$ in the case of proton absence. Then, a microscopic state is completely described by a $(2\times N_t)$-dimensional vector $k = \{\mathbf{r}_1,..., \mathbf{r}_{N_t}, x_1,... x_{N_t}\}$ consisting of the conformational part $\mathcal{R} \equiv \{\mathbf{r}_1,..., \mathbf{r}_{N_t}\}$ and charge-state part $\mathbf{X} = \{x_1,..., x_{N_t}\}$. Bearing this in mind, the electrostatic energy of the protein molecule in a state k, assuming that all the charges are point charges, can be cast in the following general form:

$$W(\mathbf{k}) = e\sum_{i=1}^{N_t}(\mathcal{P}_i - x_i) \times \left[\frac{1}{2}\sum_{\substack{j=1\\ j\neq i}}^{N_t}\varphi_i(q_j) + \varphi_i(q_{per}) + \Delta\varphi_i(q_i)\right] + \sum_{i=1}^{N_t}\mu_i x_i, \qquad (12)$$

where summations are performed over titratable groups, e is the elementary charge and $\mathcal{P}_i = 1$ for a basic group and $\mathcal{P}_i = 0$ for an acid group. The potential $\varphi_i(q_j)$ is created by a charge of the jth group, $\varphi_i(q_{per})$ stands for a potential created by all the permanent (non-titratable) charges of a protein molecule, such as backbone dipoles, etc. The quantity $\Delta\varphi_i(q_i)$ is the difference between the charge self-energies when the ith group is alone in a solvent (model compound) and when it is a part of the protein molecule.

The chemical potential μ_i in (12) is related to the concentration of the protons $[H^+]$ in surrounding solvent [12]

$$\mu_i = \mu_i^0 - RT \ln(\gamma_{H^+}[H^+]) \tag{13}$$

where γ_{H^+} is the activity coefficient of the protons, T is the temperature (in K), R is the universal gas constant and μ_i^0 is a residue-specific constant which might be viewed as a change in the self-energy of the ith titratable group of the model compound during proton binding

$$\mu_i^0 = RT \ln \mathcal{K}_i \tag{14}$$

with $\mathcal{K}_i$ being an equilibrium constant for the chemical reaction of the proton binding/unbinding to the ith group. Usually, the relations (13) and (14) are expressed in terms of $pH = -\log_{10}(\gamma_{H^+}[H^+])$ and $pK_{i,mod} = -\log_{10}\mathcal{K}_i$. Then, introducing notations

$$\Delta pK_{i,per} = \frac{e\varphi_i(q_{per})}{RT \ln(10)}$$

$$\Delta pK_{i,sol} = \frac{e\Delta\varphi_i(q_{per})}{RT \ln(10)} \tag{15}$$

Equation (12) can be recast in the more customary, frequently used in the literature form (see, e.g., Refs [13,14])

$$W(\mathbf{k}) = W_0 + \sum_{i=1}^{N_t} \sum_{\substack{j=1 \\ j\neq 1}}^{N_t} W_{ij} (\mathcal{P}_i - x_i)(\mathcal{P}_j - x_j)$$

$$- RT \ln(10) \sum_{i=1}^{N_t} (pH - pK_{i,int}) x_i \tag{16}$$

where W_{ij} is a pairwise interaction potential and

$$pK_{i,int} = pK_{i,mod} + \Delta pK_{i,per} + \Delta pK_{i,sol} \tag{17}$$

The term

$$W_0 = RT \ln(10) \sum_{i=1}^{N_t} \left(\Delta pK_{i,per} + \Delta pK_{i,sol} \right) \mathcal{P}_i \tag{18}$$

in Eq. (16) describes contributions of the permanent charges in proteins and of desolvation penalties in the electrostatic energy for the fully protonated state (all $x_i = 0$) of a protein molecule. Equation (16) gives the electrostatic energy necessary to create a microscopic state k from a reference state with zero electrostatic energy, which corresponds then to the fully protonated model compound, due to the choice of deprotonation variables. The quantity $pK_{i,int}$ (17) is then the pK value of an ith titratable group in a protein molecule with all other titratable groups being kept in their neutral state (protonated state for acid groups and deprotonated state for basic groups), as given in Ref. [15].

14.2.2.2 Electrostatic part of the protein Gibbs free energy

In the case of an unfolded protein, the summation in the statistical sum (6) is carried out over all the possible spatial locations of titratable groups, determined by the protein sequence. Further, for each of these locations, the summation over all the possible combination of deprotonation variables, x_i, must also be performed. Note, that if the protein molecule had only one state (e.g., fixed spatial structure and 'frozen' distribution of charges), then the energy $W(k)$ (16) would coincide with the Gibbs free energy G (7).

In studying pH-dependent properties of unfolded proteins, it is useful to obtain an expression for the derivative of G with respect to pH. Generally, it is written as

$$\frac{\partial \mathrm{G}}{\partial(\mathrm{pH})} = -\frac{1}{\mathrm{Z}}\sum_{\mathrm{k}}\left[\exp\left(-\frac{W(\mathbf{k})}{\mathrm{RT}}\right)\times\frac{\partial W(\mathbf{k})}{\partial(\mathrm{pH})}\right] \tag{19}$$

For the energy $W(k)$ in the form (16), the rightmost derivative in (19) adopts the form

$$\frac{\partial W(\boldsymbol{k})}{\partial(\mathrm{pH})} = \mathrm{RT}\ 1\mathrm{n}(10)\times\sum_{i=1}^{N_\mathrm{t}} x_i \tag{20}$$

Taking into account the definition of the statistical average and introducing the notation $X = \sum_{i=1}^{N_\mathrm{t}} x_i$ the expression (19) can be rewritten as

$$\frac{\partial G}{\partial(\mathrm{pH})} = -RT\ 1\mathrm{n}(10)\times\overline{X} \tag{21}$$

where $\overline{X}$ is average number of protons covalently bonded to the protein molecule (proton uptake) at given pH. Relation (21) implies that the Gibbs free energy at given pH_0 can be calculated by integrating the $\overline{X}$ (pH) dependence

$$G(\mathrm{pH}_0) = -RT\ 1\mathrm{n}(10)\int_{-\infty}^{\mathrm{pH}_0}\overline{\boldsymbol{X}}\,\mathrm{d}(\mathrm{pH})\times G_0 \tag{22}$$

The integration constant G_0 in (22) is the free energy of a protein in the fully protonated state. If the spatial structure of a molecule is fixed (as for native proteins, for instance), G_0 takes a simple form

$$G_{0,\mathrm{fixed}} = W_0 + \sum_{i=1}^{N_\mathrm{t}}\sum_{\substack{j=1\\ j\neq i}}^{N_\mathrm{t}} W_{ij}\mathcal{P}_i\mathcal{P}_j \tag{23}$$

In unfolded proteins, however, G_0 also contains a conformational average, and its calculation becomes unfeasible. Therefore, the Eq. (22) can be used only to obtain the relative contribution of electrostatic interactions into free energy. The most important application area of Eq. (22) is estimation of the electrostatic contribution into the pH-dependent free energy of denaturation for a given molecule

$$\Delta\Delta G(\mathrm{pH}_0) \equiv (G_\mathrm{D} - G_\mathrm{N}) - \Delta G_0 = -RT\ 1\mathrm{n}(10)\int_{-\infty}^{\mathrm{pH}_0}\Delta\overline{X}\,\mathrm{d}(\mathrm{pH}) \tag{24}$$

where $\Delta\overline{X} = \overline{X}_D - \overline{X}_N$ is the difference in proton uptakes between the denatured and native states of a protein at given pH. The term ΔG_0 is the pH-independent part of the free energy of denaturation and comprises integration constants, as well as likely non-electrostatic contributions. Sometimes it is convenient to use the net charge of a molecule, q, rather than the proton uptake $\overline{X}$. This is done through the relation $q = e \times (N_B - \overline{X})$ (hereafter, N_B denotes the number of basic groups in a protein). Equation (24) can, in principle, be used for evaluation of the relative difference in electrostatic free energies between any two states of the system involved, or between states of two different systems. The latter allows for studying influences of mutations on protein properties (then, the indices D and N would stand for the wild-type and mutant protein, respectively).

14.2.2.3 Hendersson–Hasselbalch titration

The partition function Z (6) can be simplified significantly by replacing the potentials in (12) by their averaged counterparts that do not depend on the microscopic state of all sites but one (mean-field approximation). Then Z can be split into N_t independent parts

$$Z = \prod_{i=1}^{N_t} Z_i(x_i) \tag{25}$$

where $Z_i(x_i)$ is the partition function of an ith site, that can be expressed in the form

$$Z_i(x_i) = \sum_{(x_i)} 10^{x_i(\mathrm{pH}-\mathrm{p}K_i)} \tag{26}$$

The quantity pK_i in (26) includes now the electrostatic pK shift

$$\mathrm{p}K_i = \mathrm{p}K_{i,\mathrm{int}} + \Delta\mathrm{p}K_{i,\mathrm{el}} \tag{27}$$

that, by analogy with (15), can be expressed as

$$\Delta\mathrm{p}K_{i,\mathrm{el}} = \frac{e\overline{\varphi}_i}{RT\ln(10)} \tag{28}$$

The potential $\overline{\varphi}_i$ is an averaged one from all other titratable residues and is independent of both spatial positions and protonation states of these charges. The summation in (26) is performed over protonation states of an ith site and in the case of the two-state model (x_i = 0,1), the average degree of deprotonation of the ith site takes the form

$$\overline{x}_i = \frac{10^{(\mathrm{pH}-\mathrm{p}k_i)}}{1+10^{(\mathrm{pH}-\mathrm{p}k_i)}} \tag{29}$$

The potential $\overline{\varphi}_i$ in (28) and consequently the pK_i value in (29) can be dependent on some external factors, such as pH and temperature. The case of pH-dependent pK_i values with strong electrostatic interactions between the titratable sites was described in detail in Ref. [16] and here only the case of the pK_is being constant is examined. Then, Eq. (29) describes sigmoidal Hendersson-Hasselbalch (HH) titration and the pK_i value has a meaning of the

pH value at which there is 50 percent probability for a proton to be bound to this residue, and therefore it is often called the $pK_{1/2}$ value.

In the case of HH titration, the electrostatic contribution into the protein denaturation free energy (24) can be simplified using the expression for the indefinite integral

$$\int \frac{10^x}{1 + 10^x}\, dx = \frac{1}{\ln(10)} \ln(1+10^x) \tag{30}$$

Then, taking into account that $\overline{X}_{D(N)} = \sum_{i=1}^{N_t} \overline{x}_{i,D(N)}$, at given pH_0

$$\Delta\Delta G = RT\left[\sum_{i=1}^{N_t}\left\{\ln\left(1+10^{pH_0-pK_{D,i}}\right) - \ln\left(1+10^{pH_0-pK_{N,i}}\right)\right\}\right] \tag{31}$$

If the pH_0 value is in the extreme basic region ($pH_0 >> pK_i$ for all i), where a protein has all its titratable groups deprotonated, then $\Delta\Delta G$ has the meaning of a complete electrostatic contribution and can be expressed in a simple form

$$\Delta\Delta G = RT\ln(10)\sum_{i=1}^{N_t}\Delta pK_i \tag{32}$$

with $\Delta pK_i = pK_{D,i} - pK_{N,i}$ being the difference in pK_i values between native and denatured forms or, generally speaking, between any two forms of a protein.

14.3 MONTE CARLO SIMULATIONS OF IONIZATION PROPERTIES FOR UNFOLDED PROTEINS

14.3.1 Importance of accounting for electrostatic interactions in unfolded proteins

Stability of proteins at given conditions is determined by the difference in the Gibbs free energies, ΔG, between folded and unfolded (or denatured) conformations of a molecule. The theoretical prediction of ΔG and its components is a prerequisite for the correct understanding of various functional properties of proteins. The electrostatic component of ΔG can be deduced from the pH-dependence of the denaturation energy, where electrostatic interactions are a predominant factor. Electrostatic interactions cannot be measured directly, and therefore theoretical efforts are focused on prediction of measurable quantities dependent on electrostatic interactions. These are, for instance, the protonation/deprotonation equilibrium constants of the titratable groups in proteins measured by NMR. Any theoretical efforts toward prediction of these constants (or their equivalent, the pK values) are mainly concentrated on native proteins, and studies aimed at analysing electrostatic properties of unfolded proteins are scarce.

Conventionally, the denatured state is modelled by means of a 'zero' approximation where the electrostatic interactions are set to zero and the titratable groups are characterized by pK values of amino acids with the alpha amino and carboxyl groups blocked by

non-titratable residues. The p*K* values of such model compounds can be determined experimentally, which renders this approximation an appropriate reference state for p*K* calculations of native proteins (see, for review Ref. [17]). It becomes, however, non-applicable for prediction of other quantities, such as the electrostatic term of unfolding energy, because the electrostatic interactions are appreciable in unfolded proteins as well [18–22]. For instance, Fersht and co-workers in a series of studies [19–21] have demonstrated that the p*K* values of the acidic groups in barnase are on average by 0.4 pH units lower than those of the model compounds. Sheafer *et al.* [23] used the extended conformation for the calculation of electrostatic interactions in the denatured state. In this model, the titratable groups are characterized by maximum solvent accessibility, which corresponds to the commonly accepted assumption that titratable groups are fully hydrated in the denatured state. A similar model has been proposed by Warwicker [24].

Denatured protein is again represented by the extended conformation, but locations of the ionizable sites are defined by additional parameters. Yang and Honig [25] have employed a hybrid approach to analyse the influence of solution pH and ionic strength on the stability of sperm whale apomyoglobin. The unfolded state of this molecule has also been modelled by the 'zero' approximation, whereas intermediate states of the protein have been described using a combination of the p*K* values for the 'zero' approximation and for the native state. A more general approach to modelling of the unfolded state has been proposed by Elcock [26], by employing a simple molecular mechanics protocol that uses the native state as a starting point.

In spite of good agreement with experimental data, all the above models have a severe limitation, since they are considering only one possible distribution of protein charges (a single conformation), which, in general, may not be representative for an unfolded protein. A generic model for calculation of electrostatic interactions in denatured proteins has been proposed in the works [27,28]. Along with this, in his series of papers [29–31], Zhou has proposed a Gaussian-chain model for treating residual charge–charge interactions in denatured proteins. These two approaches are based on very similar concepts and objectives and have been successfully applied to calculate the free energy of denaturation for several proteins. In this section, we present a survey of the Monte Carlo studies of a spherical model of the denatured state of proteins, which is based on a few physically plausible parameters and allows the accuracy of explicit calculations to be achieved.

14.3.2 Spherical model of unfolded proteins

14.3.2.1 Basic concepts

There are three basic assumptions behind the spherical model of unfolded proteins. The first one is related to the fact that an unfolded protein has to be considered an average over all the possible conformations of a flexible polypeptide chain. Therefore, it is plausible to suggest that the unfolded protein 'adopts' (on average, of course) the most symmetric shape in nature, i.e., a sphere, inside which the largest part of the protein atoms reside. The dielectric constant ε_p inside the sphere may have any value between ε_p of the native protein (which is usually taken between two or four) and the dielectric permittivity of the solvent, ε_s. To our knowledge, there are no consistent experimental data on ε_p, therefore, ε_p is so far an adjustable parameter in the spherical model of the denatured state.

The second fundamental assumption is that the titratable sites of a denatured protein in equilibrium are located on the surface of the sphere. This assumption is justified by the fact that because of differences in the self-energies, an unconfined charge is expelled from the protein interior (a medium with a low dielectric constant) towards the solvent (a medium with a high dielectric constant) [32]. Since charges belong to protein moiety and, due to the polypeptide chain flexibility, it is plausible to assume that titratable sites of a denatured protein in equilibrium tend to adopt positions on the dielectric boundary.

The radius of the sphere is assumed to be equal to the radius of gyration, R_g, of a protein in its denatured state. According to the definition (Ref. [15], p. 151), the radius of gyration, R_g, of a molecule in a given conformation k is expressed as

$$R_g(k) = \sqrt{\frac{\sum_n m_n r_n^2}{\sum_n m_n}} \tag{33}$$

where the summation is performed over all atoms of a molecule, r_n is the distance between an nth atom and the centre of the mass for a given conformation of the flexible chain, and m_n is the mass of the nth atom. In general, due to changes in conformational population, R_g of protein molecules is not a constant. It depends on pH, ionic strength, temperature, etc. (for recent experimental data on R_g(pH,T) for a number of proteins see, e.g., Refs., [33-43]). However, due to conformational freedom of unfolded proteins, direct calculation of R_g from Eq. (33) is a computationally cumbersome task which requires explicit consideration of all the protein atoms, covalent bonds and angles for a large number of conformations. As far as a model of a denatured state would treat the protein moiety as a continuum medium, it is highly desirable to find an expression for R_g which depends on as few parameters as possible. As was shown in Ref. [28], R_g for unfolded proteins can be evaluated from the relation

$$R_g = \sqrt{27 + 5 \times N_r} \tag{34}$$

Despite its simplicity, Eq. (34) gives R_g values that show reasonable agreement both with experimentally measured R_g for a number of unfolded proteins and with the values obtained by another theoretical method (for details, see Table 1 in Ref. [28]). It should be noted, however, that Eq. (34) was obtained from pure geometrical considerations and therefore is applicable for proteins under certain conditions only.

Thus, an unfolded protein is represented in terms of the spherical model by a virtual chain linking titratable sites of the protein and located on a surface of a dielectric sphere with radius equal to the radius of gyration for the protein considered. The sphere is immersed into a solvent, and the chain is free to move across the surface under certain conditions (Fig. 14.3).

14.3.2.2 Electrostatic energy in the spherical model

For calculations of the pair-wise electrostatic interactions W_{ij} in Eq. (16) we employed a formalism originally developed by Tanford and Kirkwood [44] for a spherical representation of native protein molecules. Thereupon W_{ij} splits into three parts

$$W_{ij} = 332 \times \left(W_{ij}^{(1)} + W_{ij}^{(2)} + W_{ij}^{(3)} \right) \tag{35}$$

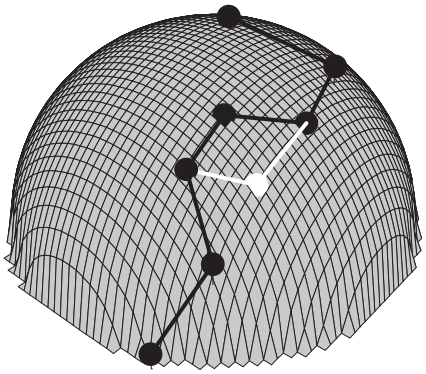


Fig. 14.3. Schematic representation of an unfolded protein as a virtual chain (thick solid line) connecting titratable sites of a protein (black filled circles) and located on a surface of a dielectric sphere (grey body) in the spherical model. Thin solid lines represent the spherical grid used in the calculations and the white fragment shows a possible move of the virtual chain. For clarity, only one half of the sphere is displayed in the figure.

The coefficient 332 is used in order to obtain energy in kcal/mol units. The first term on the right hand side of Eq. (35) is a simple Coulomb interaction

$$W_{ij}^{(1)} = \frac{q_i q_j}{\varepsilon_p r_{ij}} \tag{36}$$

where r_{ij} is the distance between the charges q_i and q_j of the ith and jth titratable groups, respectively, and ε_p is the dielectric permittivity of the protein moiety. The second term in (35) arises from the fact that protein and surrounding solvent are different dielectric media:

$$W_{ij}^{(2)} = \frac{1}{\varepsilon_p R_g} \sum_{n=0}^{\infty} \frac{(n+1)(\varepsilon_s - \varepsilon_p)}{(n+1)\varepsilon_s + n\varepsilon_p} P_n(y) \tag{37}$$

where n is an integer number and $P_n(y)$ are the nth order Legendre polynomials of the variable $y = 1 - (r_{ij}^2 / 2R_g^2)$. The third term in (35) reflects the influence of ionic strength

$$W_{ij}^{(3)} = \frac{1}{\varepsilon_s(R_g + \Delta R)} \left\{ \frac{x}{1+x} + \sum_{n=1}^{\infty} \left[\frac{2n+1}{2n-1} \times \left(\frac{\varepsilon_s}{(n+1)\varepsilon_s + n\varepsilon_p} \right) \right. \right. \times \left. \left. \frac{Z^{2n} x^2 P_n(y)}{\dfrac{K_{n+1}(x)}{K_{n-1}(x)} + Z^{2n+1} \dfrac{n(\varepsilon_s - \varepsilon_p)}{(n+1)\varepsilon_s + n\varepsilon_p} \dfrac{x^2}{4n^2 - 1}} \right] \right\} \tag{38}$$

where ΔR is a radius of a salt ion in the solvent (in our calculations the value $\Delta R = 2$ Å), $z = R_g/(R_g + \Delta R)$, $x = \chi(R_g + \Delta R)$, χ is the Debye–Hückel parameter, and the polynomials

$$K_n(x) = \sum_s \frac{2^s n!(2n-s)!}{s!(2n)!(n-s)!} x^s,\ s = 0, 1, \ldots, n \tag{39}$$

Because the polypeptide chain is considered flexible, the interactions of the titratable sites with the other charged components, such as peptide dipoles, can be neglected ($\Delta pK_{i,per} = 0$ in Eq. (17)). It has been shown [32], that $\Delta pK_{i,sol}$ is sensitive to the dielectric environment. It is then desirable to obtain values of $\Delta pK_{i,sol}$ for different types of titratable groups before applying the spherical model to denatured proteins. For this purpose, the calculations for seven types of titratable groups (glutamic and aspartic acids, lysine, arginine, cysteine, tyrosine and histidine) located on the surface of a dielectric sphere have been carried out for different radii and dielectric constants (ε_p) inside the sphere. The values of $\Delta pK_{i,sol}$ were calculated using the formula:

$$\Delta pK_{i,sol} = \frac{1}{2RT\ 1n(10)} \left[\sum_m q_m^d \left(\varphi^d_{protein}(\mathbf{r}_m) - \varphi^d_{solvent}(\mathbf{r}_m) \right) \right.$$

$$\left. - \sum_m q_m^p \left(\varphi^{\,p}_{protein}(\mathbf{r}_m) - \varphi^{\,p}_{solvent}(\mathbf{r}_m) \right) \right] \tag{40}$$

where q_m^d and q_m^p are the charge values of an atom m of the titratable group i in its deprotonated (d) and protonated (p) forms, respectively. The potential fields $\varphi^{p(d)}_{protein}(\mathbf{r}_m)$ and $\varphi^{p(d)}_{solvent}(\mathbf{r}_m)$ are created by the charges, q_m, of protonated (deprotonated) residue in protein and in solvent, respectively. The calculations were performed numerically using the finite difference method. The calculated $\Delta pK_{i,sol}$ values display a clearly pronounced dependence on ε_p (see Fig. 14.4, where typical results of calculations are shown).

14.3.2.3 Coding of protein sequence

Another important aspect in statistical modelling of unfolded proteins is how the position of the titratable groups along the amino acid sequence can be introduced into a statistical model and how this influences electrostatic interactions. Our initial proposal [10] was to describe the sequence by a quantity that depends only on the number, m, of non-titratable residues between two titratable residues adjoined along the sequence. For this purpose, a parameter

$$d(m) = \frac{1}{N_r - m} \sum_{l=1}^{N_r - m} \left\langle d_{l,l+m} \right\rangle \tag{41}$$

has been introduced. In the latter expression, index l enumerates peptide units along the sequence and $\left\langle d_{l,l+m} \right\rangle$ is the average distance between titratable sites at positions l and $l+m$. This parameter represents a statistically expected distance between two ionizable sites separated by m peptide units along the protein sequence. The values of $d(m)$ have been evaluated by means of computer simulations of polylysine chains of various lengths. Then, the pK calculations have been performed using single values of $d(m)$, matching the maximum in the corresponding distributions (except for the case $m = 1$, where two equally probable values were considered). This approach has been proven to give an overall good description of ionization properties of unfolded proteins [10]. Nevertheless, we are aware that the computational simplicity gained by the use of a single value $d(m)$, determined on the basis of simulations of polylysine polypeptide chains, may lead in some cases to a distorted picture of ionization properties. First, titratable residues differ in length and in number of internal rotational degrees-of-freedom,

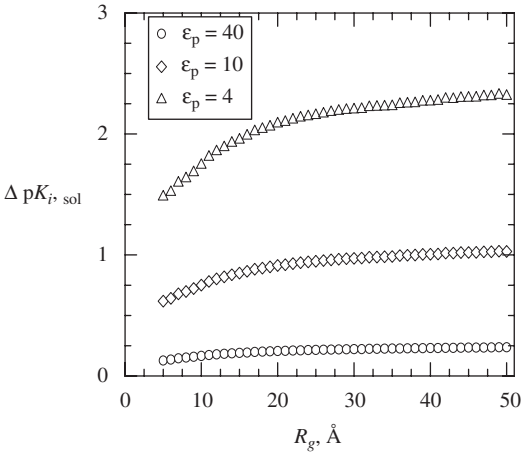


Fig. 14.4. Calculated $\Delta pK_{i,sol}$ values for the glutamic acid as a function of radius, R_g, of the dielectric sphere at different values of dielectric constant, ε_p, inside the sphere. The results were obtained for simplified charge distributions. For the deprotonated state a charge of -0.5 p.u. was assigned to carboxyl oxygens. For the protonated state all atoms were taken as neutral.

a feature that cannot be described by homogeneous chains. Second, the determination of $d(m)$ as the maximum of the statistical appearance can be a good approximation for narrow bell-shaped distributions (like the Gaussian distribution with a small dispersion, for instance), but may be inadequate in the case of small m, e.g., for two neighbouring titratable residues along the sequence.

Therefore, Monte Carlo studies of the $d(m)$ parameter as functions of m and chain composition were continued in Ref. [45], where the histograms $d(m)$ were adopted for the distances between titratable atoms of titratable residues ($O_{\delta 1}$ and $O_{\delta 2}$ atoms for Asp, $O_{\varepsilon 1}$ and $O_{\varepsilon 2}$ atoms for Glu, N_ζ atom for Lys, NH1 and NH2 atoms for Arg, $N_{\delta 1}$ and $N_{\varepsilon 2}$ for His and OH atom for Tyr). The detailed analysis of the adopted histograms is given in Ref. [45], so that here only the main results and conclusions are presented. Figure 14.5 shows $d(m)$ distributions obtained for poly-aspartic and poly-glutamic acids, poly-arginine and poly-lysine, for m ranging from one to six. The calculations show that $d(m)$ distributions become insensitive to the length of side chains when charges are separated by ~ 10 rotational degrees- of-freedom ($m \geqslant 4$). This means that, for the proximate titratable sites ($m < 4$), the statistically expected distances cannot be expressed either by a single value (average or maximum) of the distribution $d(m)$, or by the Gaussian-chain distribution. As m increases, the average of $d(m)$ obeys a power law with the exponential factor, that is, within the limits of statistical accuracy, independent of the sidechain type, with the only exception being the histidine sidechain, which is probably related to an asymmetric location of two charge sites with respect to the C_β–C_γ bond. This asymmetry is probably also the cause of the $d(m)$ shape for the poly-histidine chains, which is more complex compared with the other poly-chains considered.

It is interesting to compare the distance distributions calculated by means of the Gaussian-chain formalism (see formula (5) in Ref. [31]) to the distance distributions obtained from our calculations. Such a comparison is depicted in Fig. 14.5. As can be seen,

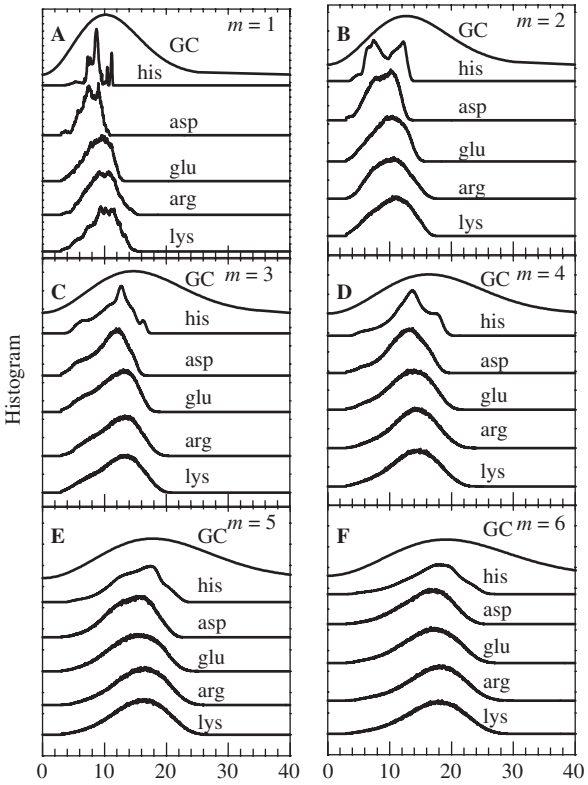


Fig. 14.5. Distributions of distances between titratable sites of sidechains separated along the sequence by $m-1$ residues obtained from the Gaussian-chain model (curves labelled GC) and from computer simulations for poly-asp, poly-glu, poly-arg, poly-lys and poly-his chains with m ranging from 1 to 6. The upper five curves in the panels are shifted upwards for clarity.

the Gaussian-chain distributions even for large m stretch towards distances that are not allowed in polypeptide chains by the sterical collisions criteria and by covalent bonds (Fig. 14.5). E.g., for the residues neighbouring in the sequence ($m = 1$), the allowed distances are from 3.0 Å to 12 Å (for the shortest asp sidechain) or to 16 Å (for the longest arg sidechain), while the Gaussian-chain distribution allows in practice all distances between 0 and 30 Å. The tails of the Gaussian-chain distributions in the sterically forbidden regions contribute ~ 20 percent of interaction energy for the neighbouring arg sidechains and ~30 percent for the neighbouring Asp sidechains. This may cause a misinterpretation of the role of the charge-charge interactions in determining ionization properties of unfolded proteins. Nevertheless, it should be noted that the Gaussian-chain model produces results that agree well with the experimental data for a number of proteins [30,31,46].

14.3.2.4 Calculation of free energy profile

The method described above produces histograms of distances between titratable sites of the charged residues in a system where only hard-sphere potentials are taken into consideration.

At this point, a natural question arises as to whether these histograms can be of practical relevance for real proteins at physiological conditions, when all those sites usually bear a charge and therefore charge–charge interactions cannot be neglected. The positive answer to this question comes from statistical physics, where it is very well known that histograms for the quantities of interest in the absence of certain interactions can be easily recalculated (within certain limits, of course) for those interactions existing in a system. For the distributions considered in this chapter this can be done by calculating a free energy [47,48]

$$\Delta G_{ij}(r) = -\beta^{-1} \ln \frac{d(m, r) \times \exp\left(-\beta W_{ij}(r)\right)}{\sum_{r} d(m, r) \times \exp\left(-\beta W_{ij}(r)\right)} \tag{42}$$

where $m = |i-j|$ is a separation parameter between the groups i and j along the sequence, $\beta = 1/RT$ with R and T being the universal gas constant, and the temperature (in K), respectively. The quantity $d(m,r)$ stands for number of states, for which titratable sites of a pair of groups with separation parameter m are r Å apart from each other. We used the $d(m,r)$ values from the corresponding distance distributions obtained for generic chains (see the previous section). The quantity $W_{ij}(r)$ stands for the energy of a system composed of two charges in the state, when charges are separated by r Å, and can be calculated by the expression

$$W_{ij}(r) = 332 \times q_i q_j \times W(r) \tag{43}$$

where q_i and q_j are the charges of groups i and j, respectively, and $W(r)$ is the interaction potential between two charges separated by distance r and located on the surface of protein dielectric moiety (for details see Ref. [28]). At its equilibrium, the system tends to adopt a distance between the charges corresponding to the minimum of the free energy. Hence, the functions $\Delta G_{ij}(r)$ for each pair of titratable groups provide information about the energetically favourable lengths of the virtual chain segments.

The application of the function $\Delta G_{ij}(r)$ is illustrated in Fig. 14.6 for four typical examples of titratable site pairs in the drkNSH3 sequence. Since the definition of $\Delta G_{ij}(r)$ comprises charge–charge interactions only, the following two cases can be clearly distinguished for each pair: (**i**) both sites in a pair are charged (black lines in Fig. 14.6) and (**ii**) at least one site is neutral (grey lines in Fig. 14.6). In the latter case, the position of the $\Delta G_{ij}(r)$ minimum coincides with the position of the maximum at corresponding distance distribution, $d(m)$. Thus, the use of a single value at maximum distance distribution is partly justified, at least for extreme pH values and for acidic-basic pairs, when one residue in such a pair is, as a rule, neutral. But, the situation is more complex at physiological pH, where all titratable groups bear a charge.

As can be seen in Fig. 14.6, the location of the minimum shifts toward shorter distances in the case of opposite charges (see panel (C), as an example) and for the like charges the move is reversed, as illustrated in panel (D). The $\Delta G_{ij}(r)$ function could have a complex shape with several minima of almost the same depth (see, e.g., panel (A) of Fig. 14.6, where $\Delta G_{ij}(r)$ is shown for the Asp15–Glu16 pair). The magnitude of energy changes with distance, however, is comparable to the energy of thermal fluctuations.

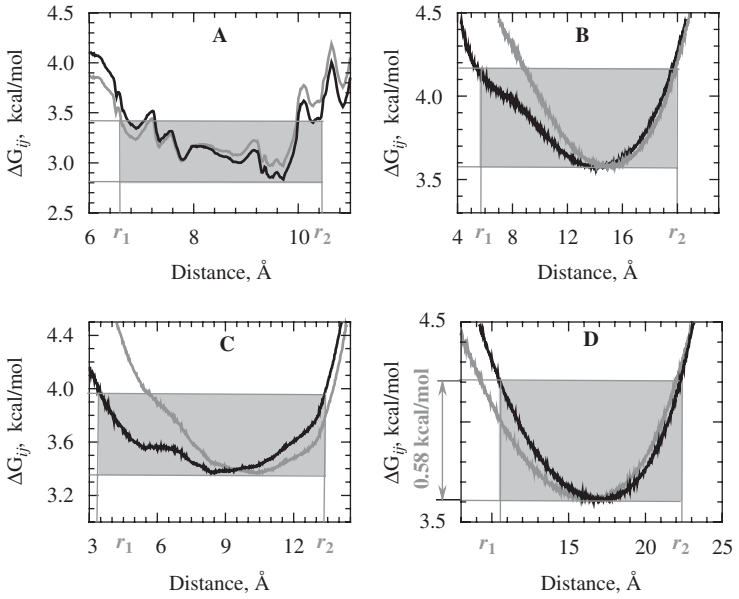


Fig. 14.6. Free energy profiles, ΔG_{ij}, (formula (42)) for Asp15–Glu16 (panel (A)), Glu16–Arg20 (panel (B)), Asp42–Lys44 (panel (C)), and Lys21–Lys26 (panel D) pairs of titratable residues in drkNSH3 protein at 20 °C. Black lines represent ΔG_{ij} when both residues in the corresponding pair are charged while grey lines stand for ΔG_{ij} in the case when at least one residue in the pair is neutral. Shadowed regions show ΔG_{ij} regions (of width ~0.6 kcal/mol) for which intervals of equally probable distances for the pairs are selected in the model.

Therefore, it is plausible to suggest that all states for which the difference between the $\Delta G_{ij}(r)$ value at given distance r and the minimal ΔG_{ij} value is less than 0.6 kcal/mol, are nearly equally populated. Grey rectangles in Fig. 14.6 illustrate how such an interval of equally probable distances is chosen for a certain pair of residues. Provided that the lower boundary of the rectangle is at the level of $\Delta G_{ij}(r)$ minimum and rectangle height is 0.6 kcal/mol, the intersection of upper rectangle boundary with the distance distribution, when both residues are charged, gives the interval, $[r_1, r_2]$, of allowed distances for the pair of residues in question.

14.3.3 Details of the Monte Carlo algorithm

The pK value of an individual titratable site i can be obtained by calculating the degree of deprotonation for this group $<x_i>$ as a function of pH. At given pH and temperatures, $<x_i>$ can be accurately determined on the basis of statistical mechanical calculations, and then the statistical average (8) transforms into the following expression

$$<x_i> = \frac{1}{Z}\sum_{\mathbf{k}}^{L} x_i^{\mathbf{k}} \exp\left(-\frac{W(\mathbf{k})}{RT}\right) \tag{44}$$

Here Z is the partition function (6) and $W(k)$ is the electrostatic energy of the molecule in the state k defined by Eqs (16) and (17). The summation in Eq. (44) runs over all L microscopic states of the protein molecule. Assuming a single structure for a native protein, such a microscopic state is uniquely defined by an N_t-dimensional vector $\mathbf{k} = \{x_1,...x_{N_t}\}$, which gives $L = 2^{N_t}$. The assumption of a single structure does not hold for unfolded proteins, since titratable sites can change their spatial allocation. The vector $\mathbf{k}$ becomes then $(2 \times N)$-dimensional $\mathbf{k} = \{\mathbf{r}_1...\mathbf{r}_{N_t}, x_1,...x_{N_t}\}$. This increases L dramatically and makes the direct summation in Eq. (44) unfeasible, even for short polypeptide chains ($N_t \sim 10$). The Monte Carlo technique provides an appropriate solution to this problem by replacing the direct summation in Eq. (44) with the Monte Carlo average (3)

$$<x_i>=\frac{1}{L_{MC}} \sum_{k=1}^{L_{MC}} x_i^k \tag{45}$$

over the number $L_{MC}<<L$ of states generated randomly according to the probability (10).

The protein molecule is represented as a virtual chain with N elements, each of which corresponds to one titratable site. According to the assumptions described above, the chain elements (i.e., the titratable sites) are located on the surface of a sphere that comprises the material of the denatured protein. The possible positions of the chain elements are predetermined by a set of uniformly distributed points on the surface of the sphere forming a spherical grid (thin solid lines in Fig. 14.3). The minimum distance between the grid points, d_{min}, is about 3 Å, which is close to the minimum distance between the charges of an ion pair in proteins.

An initial configuration of the virtual chain is generated by arbitrarily placing the first point on the spherical grid and by placing each next point so that **(i)** the chain does not become self-intersecting and **(ii)** the distance from the previous point satisfies the constraint $d(m) \pm d_{min}$. When generating the state k' from k either the protonation state ($x_i(k)$ changes to $x_i(k')$ in equation (16) or allocation (W_{ij} ($\mathbf{k}$) $\rightarrow W_{ij}$ ($\mathbf{k'}$) in Eq. 16) of a randomly chosen titratable site is altered. In the case of reallocation of site i, the distance constraints with respect to both the (i–1)th and the (i+1)th sites are taken into account (thick white line in Fig. 14.3). The choice between the two types of alterations is performed in a random way. The set of deprotonation variables was stored after repeating the above generation procedure N times, which comprises one MC step per site (MCS/S). An average of over 20 MC runs with different initial configurations of the virtual chain was also performed. Each run consists of 10000 MCS/S with the first 2000 MCS/S being discarded from consideration. The average over MC runs essentially reduces computational time, since it extinguishes problems related to a trapping of the system in a metastable state. For instance, a significant scattering of points is observed in the single-residue titration curve obtained from one very long (107 MCS/S) MC run while the average over 20 relatively short MC runs gives a smooth titration curve (see Fig. 14.7).

The results of calculations for the spherical model were obtained for T = 20°C, ionic strength 0.1 M, dielectric permittivity of the solvent ε_s = 80 and for various values of the dielectric permittivity, ε_p, of the protein moiety. The following set of standard pK values [49, 50]: p$K_{i,mod}$(Asp) = 4.0; p$K_{i,mod}$(Glu) = 4.4; p$K_{i,mod}$(His) = 6.3; p$K_{i,mod}$(Cys) = 9.1; p$K_{i,mod}$(Tyr) = 9.4; p$K_{i,mod}$(Lys) =10.4 and p$K_{i,mod}$(Asp) = 12.0 was used.

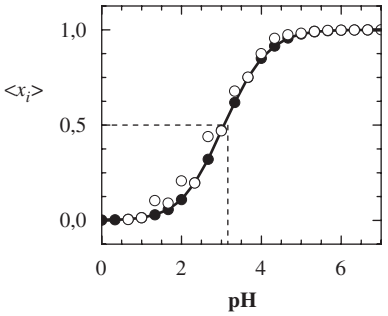


Fig. 14.7. Titration curve for the Asp12 residue of barnase obtained from one long (10^7 MCS/S, open circles) and 20 short (10^4 MCS/S, filled circles) MC runs. The dashed lines show the determination of the pK value.

14.3.4 Applications of the spherical model of unfolded proteins

14.3.4.1 Barnase

Barnase from *Bacillus amyloliquefaciens* is a protein with $N_r = 110$ residues and Eq. (34) gives a radius of gyration $R_g = 24$ Å. This value is between the experimental observations, which vary from 15.9 Å to 34 Å [51]. The protein contains 37 titratable groups, which are distributed over 997 points of the spherical grid with the distance between the neighbouring points $d_{\min} = 2.7$ Å. Since experimental results for the proteins considered in the present review are available for the acidic pH region only, hereafter we present the pK values only for acidic groups, although they were calculated for every type of the titratable groups. The only exception here is the His18 group of barnase, for which the pK value was measured directly [52].

The pK values for barnase are presented in Table 14.1 along with the experimental pK values for the acidic groups for native barnase as reported by Oliveberg *et al.* [20]. In the same work the following experimental pK values were reported for the denatured state of barnase: pK = 3.50 for aspartic acids and pK = 3.70 for glutamic acids. These experimental pK values are, in fact, averaged values obtained from a theoretical analysis of pH-dependent difference ΔQ between the net charge of the barnase molecule in the native and denatured states. ΔQ, in turn, was calculated as the derivative of unfolding free energy ΔG obtained from thermal unfolding experiments [19]. Such methodology does not provide Q for both native and denatured states separately while the MC simulations give Q only for the denatured state. Thus, the pK values rather than net charge are to be compared here.

The pK values calculated for the acidic groups are sufficiently close to, but systematically below the experimental pK values. This discrepancy is most likely due to an overestimation of electrostatic energy, which can have several sources. One of them is the dielectric constant of the protein moiety, ε_p, which is an adjustable parameter in the present calculations. Results of calculation for various ε_p values (see Table 3 in Ref. [28]) indicate that the pK values move towards the model pK values with ε_p increasing and for $\varepsilon_p = 40$ the electrostatic pK shifts practically coincide with the experimentally obtained pK shifts.

For the His18 residue, the agreement between the calculated and measured pK values is good for all ε_p, because in this pH region the net charge of the molecule is close to zero,

Table 14.1 pK values for native and denatured barnase

Residue	pK_i^{N}[a]	pK_i^{0}[b]	pK_i[c]	$pK_{el,i}$[d]
Asp8	3.10	4.00	3.70	−0.75
Asp12	3.50	4.00	3.58	−1.02
Asp22	3.30	4.00	3.56	−0.89
Asp44	3.60	4.00	3.74	−0.71
Asp54	2.20	4.00	3.69	−0.76
Asp75	3.10	4.00	3.48	−0.97
Asp86	4.20	4.00	3.68	−0.77
Asp93	1.50	4.00	3.53	−0.92
Asp101	2.00	4.00	3.69	−0.76
Glu29	3.75	4.40	4.17	−0.68
Glu60	3.40	4.40	4.03	−0.82
Glu73	2.10	4.40	4.01	−0.84
His18	6.59[e]	6.30	6.70	−0.20

[a]Experimental data from Ref [20] for native barnase.
[b]Standard pK values (null approximation).
[c]Values of pK calculated by the spherical model with $\varepsilon_p = 29$.
[d]Shift of protonation/deprotonation equilibrium due to the charge-charge interaction.
[e]Experimental data from Ref. [52] for denatured barnase.

i.e., electrostatic influence is minimal. Neglecting the desolvation penalty, $\Delta pK_{i,sol}$, also leads to an overestimation of electrostatic interactions. This effect becomes clearly manifested when the electrostatic contribution into the protein stability, $\Delta\Delta G$, is calculated from Eqs (24) and (29) (see lines 3 and 4 in Fig. 14.8). Also, the increase of ε_p leads to increasing overestimation of electrostatic interactions in the denatured state of proteins (see lines 2 and 3 in Fig. 14.8). It is also apparent from the latter figure that using the 'zero' approximation as a model of denatured state produces a significant underestimation of the electrostatic interactions in the denatured state. For instance, for barnase the change of ΔG from acidic and neutral pH is $\Delta\Delta G_{exp}^{barnase} \approx 13$ kcal/mol, while the null approximation gives $\Delta\Delta G_{NA}^{barnase} \approx 23$ kcal/mol.

14.3.4.2 N-terminal domain of the ribosomal protein L9

The N-terminal domain of the ribosomal protein L9 from *Bacillus stearothermophilus* (NTL9) has 56 residues with the C-terminus amidated and the N-terminus being free. The protein sequence includes two Asp, four Glu, 11 Lys and one Arg groups. Thus, the model virtual chain for NTL9 consists of 19 segments and R_g = 18 Å. The results of ΔG calculations with the pK values for the native state of NTL9 taken from the Ref. [22] and the constant ΔG_0 in Eq. (24) chosen so that ΔG(pH 0) = 2 kcal/mol, are shown in Fig. 14.9. Again, the 'zero' approximation gives significant underestimation of electrostatic interactions in the unfolded state. For instance, the measured change of ΔG from acidic and neutral pH is $\Delta\Delta G_{exp}^{NTL9} \approx 2$ kcal/mol, while the null approximation produces $\Delta\Delta G_{NA}^{NTL9} \approx -3.5$ kcal/mol. It is also worth noting that the introduction of the desolvation effects into the model significantly improves the evaluation of the electrostatic interactions. This is illustrated in Fig. 14.9, where $\Delta\Delta G_{EL}^{NTL9}$, calculated on the basis of pair-wise electrostatic

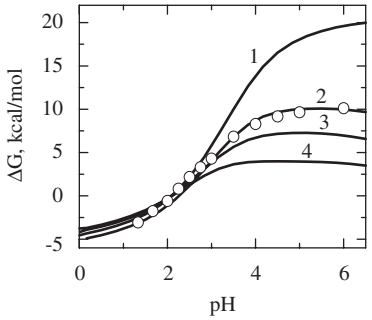


Fig. 14.8. The calculated (lines) and experimental (circles) free energy of denaturation, ΔG, as a function of pH for barnase. Line numbers refer to various representation, of unfolded state: 1. The 'zero' approximation; 2. Spherical model with $\Delta pK_{i,sol} \neq 0$ and ε_p=29; 3. Spherical model with $\Delta pK_{i,sol} \neq 0$ and ε_p=40; 4. Spherical model with $\Delta pK_{i,sol} = 0$ and ε_p=40. The constant ΔG_0 in the Eq. (24) was chosen so that ΔG(pH 2.1) = 0.

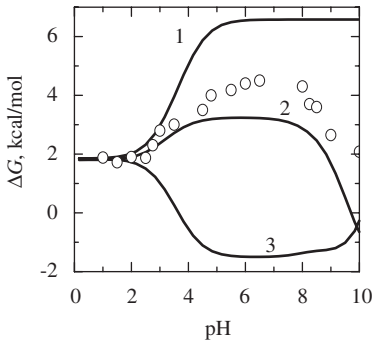


Fig. 14.9. The calculated (lines) and experimental (circles) free energy of unfolding, ΔG, as a function of pH for NTL9. Line numbers refer to various representation of unfolded state: 1. The 'zero'approximation; 2. Spherical model with $\Delta pK_{i,sol} \neq 0$ and $\varepsilon_p = 20$; 3. Spherical model with $\Delta pK_{i,sol} = 0$ and ε_p=40.

interactions only, indicates electrostatic destabilization of the native structure ($\Delta\Delta G_{EL}^{NTL9} \approx -3.5$ kcal/mol, curve 3 in the Fig. 14.9) while the experimental results (points) show the opposite.

14.3.4.3 Staphylococcal nuclease

Staphylococcal nuclease (SNase) is a highly charged protein containing 31 basic and 28 acidic groups including the amino- and carboxyl terminal groups (N_r = 149). The denatured state of SNase was modelled by a flexible chain with 61 segments and a dielectric sphere with R_g = 27.8 Å. The distance between the neighbouring points in this case was d_{min} = 3.1 Å. To our knowledge there are no experimental pK values for denatured SNase available in the literature. Therefore, we limited our consideration to comparison of protein titration curves only. In Fig. 14.10 the titration curve calculated by the model is juxtaposed with the experimental data on H^+ binding obtained by Whitten and Garcia-Moreno [53]. As can be seen, the 'zero' approximation describes SNase well within the wide pH range if the protein is dissolved in 6M GdnHCl, which is a very strong denaturant. In this case, SNase is chemically denatured

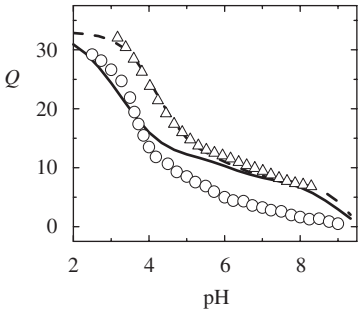


Fig. 14.10. The net charge, Q, of the SNase as function of pH. Points denote the experimental data obtained for SNase dissolved in 100mM KCl (circles) and 6 M GdnHCl (triangles). The lines show results of the calculations using the spherical model of denatured state (solid line) and the 'zero'approximation (dashed line).

within the whole pH interval investigated in Ref. [53]. It is known that GdnHCl is a highly ionized agent and therefore it screens any electrostatic interactions.

However, for the SNase dissolved in 100 mM Kcl, the calculated net charge agrees well with the experimental points for pH $\leqslant$ 4, where the protein is known to be thermally unfolded [53]. The observed, though small, discrepancy between the calculated and experimental net charges for pH $<$ 4 can be due to the factors discussed above, such as the assumption that R_g does not depend on pH. However, the agreement with experimental data in this case is significantly better than that of the null approximation (dashed line in Fig. 14.10). The net charge predicted by the null approximation is between five and ten charge units larger compared with the experimental data [53]. The seeming agreement of the results obtained by the model proposed and by the 'zero' approximations at pH between six and nine is due to the fact that both models are expected to predict all carboxyl, lysines and arginines being charged. The residual difference in the net charge in this pH region arises from the pK shifts of histidines predicted by the spherical model.

14.3.4.4 N-terminal SH3 domain of the Drosophila protein drk

Recently, Tollinger *et al*. [54,55] have developed an NMR technique which has allowed them to perform direct measurements of the pK values for acidic groups in unfolded drkNSH3 protein. This provides excellent and, to our knowledge, the only opportunity so far for a straightforward testing of various statistical models of the unfolded state developed for the calculation of electrostatic interactions.

The sequence of drkNSH3 is represented in terms of this model by a 26-segment chain (seven asp, six glu, two arg, five lys, two his, two tyr, N- and C-terminal groups) on the surface of dielectric sphere with radius R_g = 18 Å and dielectric permittivity of the moiety inside the sphere, ε_p = 20. The desolvation effect on the ionization equilibrium, ΔpK_B, was taken into account as described above (see Section 14.2.2.2).

The spherical model described in Ref. [28] uses a virtual chain with the segment lengths assumed to take single values. Thus, we refer to this approach as the rigid-chain-segment (RCS) approximation. Table 14.2 shows comparison of the pK values

Table 14.2 Comparison of experimental pK values for acidic groups of drkNSH3 protein with the pK values obtained by different model calculations

Group	Experimental[a]	GC[b]	RCS[c]	VCS[d]	Calculated[e]
Glu02	4.08	4.23	4.51	4.13	4.09
His07	7.07	6.71	6.94	7.12	7.11
Asp08	3.75	3.71	3.89	3.73	3.84
Asp14	3.97	3.94	4.00	3.89	4.01
Asp15	3.99	3.95	3.94	3.83	3.96
Glu16	4.44	4.43	4.47	4.38	4.34
Glu31	4.37	4.48	4.39	4.47	4.43
Asp32	4.05	3.99	4.01	3.94	4.06
Asp33	4.13	3.97	3.99	3.97	4.09
Glu40	4.42	4.37	4.39	4.28	4.27
Asp42	4.05	3.87	3.99	3.86	3.98
Glu45	4.45	4.29	4.49	4.39	4.34
Glu54	4.28	4.32	4.43	4.37	4.34
His58	7.83	6.85	7.06	7.21	7.23

[a]Experimental data taken from Ref. [54].
[b]Gaussian-chain model [46].
[c]Calculation based on the spherical model with fixed segment lengths proposed in Ref. [28].
[d]Calculations based on the extended spherical model with variable segment length (this work).
[e]Calculations based on the extended spherical mode with the additional assumption that centres of carboxyl oxygen atoms of the aspartic acids are 0.2 Å below the dielectric boundary.

calculated on the basis of the RCS approximation and by the Gaussian-chain model, to the experimental data for drkNSH3 from Refs [54,55]. Both models give overall good agreement with the experimental data and show, in general, small deviations from the standard pK values (pK_0 4.0 for the aspartic acids, pK_0 4.4 for the glutamic acids and pK_0 6.6 for the histidines). The pK values of the two histidines are shifted to high values, which can be expected, if one takes into account that the protein is rich of acidic groups. The small pK shifts of the glutamic and the aspartic groups are a noteworthy observation, because these shifts have different origins in the two models. In the Gaussian-chain model, the charge–charge interactions occur in a medium that has dielectric permittivity, ε_s equal to that of water, which leads to a significant diminishing of their influence.

An additional screening of EI comes from the mobile solvent ions. This would suggest the conclusion that the charge–charge interactions are not relevant for determining ionization properties of unfolded proteins. On the other hand, in the spherical model (where dielectric permittivity of a protein moiety, ε_p, and ε_s may differ considerably) the seeming small effects of electrostatic interactions stem from compensation of the charge–charge interactions and the desolvation energy [27]. E.g., for the residue Glu16 charge–charge interactions give $\Delta pK_e = -0.51$, while due to the compensation effect the resulting pK shift is $\Delta pK = \Delta pK_e + \Delta pK_B = -0.04$. In the work [56] it was shown that at physiological pH the polypeptide chain collapses to a compact structure. This leads to an effective reduction of ε_p in comparison with ε_s, and hence to an increase in

the magnitude of charge–charge interactions. Therefore, we think that the difference between ε_p and ε_s should be taken into account when electrostatic interactions in unfolded proteins are of interest. The concept of the Gaussian-chain model, although elegant in its mathematical formulation and computational simplicity, does not account for this effect.

Another noteworthy result is that both the model of Zhou [46] and the calculations based on the RCS approximation, fail to predict the p*K* value of Glu2. This residue is neighbouring along the sequence to the N-terminus, so that the number of rotational degrees-of-freedom in between for such a pair is less than between any two sidechain sites with the same separation parameter. Also, while for a cluster neighbouring along the sequence residues Asp14–Asp15–Glu16 both models give good agreement with the experiment, for a similar cluster Glu31–Asp32–Asp33 such an agreement is worse. The main distinction between the two clusters lies in the titratable groups preceding (Asp08) and following (Tyr37) these two clusters, respectively. This distinction is not reflected in the formalisms of either the RCS approximation or the Gaussian-chain model. Thus, the above observations suggest that description of distances between charged groups without taking into consideration their internal structure could be an oversimplification. Therefore, the spherical model has been reconsidered in Ref. [45] in the part concerning the choice of distance constraints for segments of the virtual chain that models an unfolded molecule.

The incorporation of the sequence of the titratable sites in the spherical model using average values calculated from pair-specific (or at least, residue-specific) distributions is attractive from the computational point of view, since average charge–charge distances depend on separation distance in a single power-law manner. Such an approach can be fruitful for studying, for example, flexible linkers (or spacers) that allow connected domains to move relative to each other with separation varying with space and time [57–59]. However, for the calculation of ionization equilibrium in unfolded proteins this is not an adequate simplification, because an essential quantity here is an average energy of the charge–charge interactions, which, in general (especially in the cases of complex asymmetrical distributions, such as those shown in Fig. 14.5(A), is not equal to the energy value at the average distance. On the other hand, inclusion of such complicated distributions explicitly into the calculation algorithm is an awkward procedure, so we have chosen a compromise.

Thus, instead of having rigid segments for a model chain, we propose to use a chain with lengths of its segments chosen randomly from an interval of equally probable distances (Section 14.2.2.4). Thus, the modified spherical model uses the variable-chain-segment, VCS, approximation. The p*K* values of drkNSH3 calculated by means of VCS approximation are shown in column (d) of Table 14.2. As seen, the agreement with experiment is better for the two acidic residues closest to the N-term. It should be noted that this improvement stems not from the VCS approximation itself but rather from using the specific terminal distance distributions (see Fig. 14.4 in Ref. [45]) that allow shorter distances for the Glu02–Nterm pair compared with the bulk distributions used in the RCS calculations. For all other residues there is no essential improvement. One can notice a tendency of p*K* values being slightly below the corresponding experimental values.

This effect can be explained by the interplay of the two main factors regulating ionization equilibrium of the titratable groups, namely, the charge–charge interactions and the non-equivalent desolvation of the aspartic and glutamic acids. The calculations of solvent accessible surface (SAS) for short flexible fragments of the drknSH3 sequence have shown that average SAS of the carboxyl oxygen atoms of the aspartic acids is ~ 3 Å^2 less compared with that of the corresponding atoms of the glutamic acids. The pK values listed in column (d) are calculated under the assumption that asp and glu are equally accessible to the solvent. In terms of the model used for evaluation of the desolvation energy, this corresponds to hydroxyl oxygen atoms situated on the surface of the dielectric boundary (for details see Ref. [27]). If one takes into account the observed difference in the average SAS of glu and asp, in terms of the desolvation energy model [27], the centres of the carboxyl oxygen atoms of the aspartic acid should be 0.2 Å under the dielectric boundary. The calculations based on this assumption are given in the rightmost column in Table 14.1. As seen, the agreement with experimental data is better in comparison with the other model.

14.3.5 Concluding remarks

In conclusion, one can say that the spherical model extended by the VCS approximation provides a physically plausible model for calculation of EI in unfolded proteins. It also illustrates the significance of the desolvation effects in the ionization equilibrium, a factor which is not considered in the current Gaussian chain model. It is obvious that the description of the desolvation effect is somewhat simplified. Nevertheless, the good agreement with the experimental data is encouraging and gives rise to the belief that the presented model can be employed for approaching other problems, such as pH-dependence of protein stability.

One of the critical parameters of the spherical model is the dielectric constant of the protein material. As mentioned above, the protonation/deprotonation equilibrium is sensitive to the value of this parameter. For instance, the calculations carried out with $\varepsilon_p = 40$ showed an apparent overestimation of the electrostatic interactions. Essentially better agreement is obtained with a lower dielectric constant (see, e.g., Fig. 14.8). This result originates from the fact that the reduction of ε_p enhances charge–charge interactions to a lesser extent than desolvation penalty. Our calculations suggest that an appropriate range for the dielectric constant of unfolded proteins is between 20 and 30.

The spherical model produces systematic overestimation of the electrostatic interactions, and one of the sources for this may be the fact that the radius of the dielectric (protein) sphere is kept constant for the whole pH range considered. Indeed, one expects that electrostatic repulsion between titratable sites may induce a swelling of the protein moiety, especially at extreme pH. Experimental evidence for such an effect has been provided by Stigter *et al.* [60], who have shown that the average coil radius of the myoglobin molecule increases dramatically with reduction of the ionic strength. In general, proteins at different conditions (such as pH and temperature) may adopt various denatured states characterized by different compactness and often by distinctive residual secondary structures. These factors are not reflected in Eq. (34), which has been obtained from pure geometrical considerations. A reconsideration of this equation can be carried out, for instance,

on the basis of Monte Carlo calculations for the generic polypeptide chain described in the next Section.

14.4 MONTE CARLO STUDIES OF DIMENSIONS OF UNFOLDED PROTEINS

14.4.1 Generic-chain representation of unfolded proteins

14.4.1.1 Basic concepts

Another approach to modelling of the unfolded state of proteins is the use of a generic polypeptide chain with charge distributions typical for proteins. The molecules of proteins are possessed of a very complicated structure, and therefore many researchers use a virtual chain as their representation. While this approach provides good description of protein properties on a macroscopic level, a microscopic insight is very often lost. On the other hand, explicit consideration of all atoms in a protein is very cumbersome and computationally demanding task. In the study [56], for the sake of simplicity, the unfolded state of protein molecules is approximated by explicit protein backbone and a single sidechain atom at the β-position (polyalanine). Calculations with sidechain atoms at the γ, δ, ε and ζ position were also carried out. The use of polyalanine as a general representation of denatured proteins in most cases is an oversimplification. However, such an approach turns out to work rather well for studying dimensions of unfolded polypeptide chains [56].

In order to extract the influence of pure electrostatic interactions on dimensions of unfolded polypeptides, it is convenient to ignore all contributions to the electrostatic energy of the protein molecule (16), $W(k)$, other than pair-wise interactions W_{ij}. Then, the pair-wise interaction energy of the chain in a conformation k consists of two parts:

$$W(\mathbf{k}) = \sum_{ij} W_{ij}^{\mathrm{hs}}(\mathbf{k}) + \sum_{ij} W_{ij}^{\mathrm{el}}(\mathbf{k}) \tag{46}$$

where the first and second sum is taken over all atoms and over all charges of the chain, respectively. The hard-sphere potential $W_{ij}^{\mathrm{hs}}(\mathbf{k})$ is approximated as

$$W_{ij}^{\mathrm{hs}}(\mathbf{k}) = \begin{cases} 0, & \text{if } r_{ij}(\mathbf{k}) > d_{ij}^{\min} \\ \infty, & \text{if } r_{ij}(\mathbf{k}) < d_{ij}^{\min} \end{cases} \tag{47}$$

Here, $r_{ij}(\mathbf{k})$ is the distance (in Å) between the atoms i and j in a conformation k and $d_{ij}^{\min}$ is a minimal allowed distance between atoms i and j. A conformational space of this generic chain should resemble a conformational space of proteins, and therefore a set of $d_{ij}^{\min}$ values, corresponding to the extreme limits of the Ramachandran criteria [61], was used. The electrostatic interactions are the strong and long-range ones and, for instance, the distance between the opposite charges could be less than the Ramachandran criteria for the C_β atoms. Hence, calculations with different values of $d_{ij}^{\min}$ for the contact $C_\beta - C_\beta$ were also performed. The electrostatic part of $W(k)$ in the Eq. (46) was calculated using the Coulomb law (36). For the sake of simplicity, all other contributions to $W(k)$ (energy

of hydrogen bonds, etc.) were ignored in the study [56], although their introduction into the simulation procedure is straightforward and does not complicate the computational task significantly.

14.4.1.2 Charge distribution and dielectric properties of the medium

The charge distribution along the protein sequence plays a significant role in determining properties of its native conformation and even its unfolded state [28]. However, the effects of interest in this study come from the number of charges rather than from a specific charge allocation along the chain. Taking into account that the frequency of titratable residues occurrence in proteins is ~25 per cent (without tyrosines and cysteines) [62], we approximate denatured proteins by flexible polyalanine chains with ionizable sites uniformly distributed along the chain with densities $\rho = (N_A + N_B)/N_r \leqslant 0.25$, where N_r is the total number of sidechains (C_β atoms), N_A is the number of the acidic sites, and N_B is the number of basic sites. Electrostatic interactions (36) are symmetric with respect to the sign of interacting charges. This allows us to simplify the task by reducing the amount of variables by altering the number, $0 \leqslant N_A^- \leqslant N_A$, of negative charges while the number of positive charges is always $N_B^+ = N_B$.

An important aspect in the calculation of electrostatic interactions is the dielectric response of the medium. For charges immersed in water, the Coulomb law (36) with uniform ε value does not hold at separation distances comparable with a few water molecules [63], where the dielectric constant is reduced. This effect can be taken into account by introducing a distance-dependent dielectric permittivity $\varepsilon(r_{ij})$ in Eq. (36). There are two groups functions for $\varepsilon(r_{ij})$ widely used in the literature. The first group consists of monotonic functions increasing with r_{ij}, like linear [64–66] or quadratic [67] ones. The second group comprises sigmoidal functions with various parameters deduced from experiments or/and from theoretical considerations [68–72]. We have chosen the function proposed in the work [72] and modified in the work [70]

$$\varepsilon(r_{ij}) = \varepsilon_s \left(1 - \frac{z^2 \exp(z)}{(\exp(z) - 1)^2} \right) \tag{48}$$

with $z = 0.4 \times r_{ij}$. The function (48) reproduces the earlier numerical calculations [68] well and has its saturation value at $r_{ij} \sim 20$ Å.

14.4.2 Details of the Monte Carlo algorithm

The simulation algorithm used for the generic polypeptide chains is similar to the Monte Carlo algorithm used for the spherical model (see Section 14.3.3). Conformations of a flexible chain were generated by randomly sampling dihedral angles φ and ψ for a backbone with standard geometry, as shown in Fig. 14.11. The first conformation is generated by consecutively building residues with random choice of φ and ψ. The position of each atom thus generated was checked according to Eq. (47) against the sterical collisions with all the previously generated atoms and is accepted only if all distances satisfy the Ramachandran criteria (see Table 14.3).

After the chain of desired length is built, each second conformation $k \geqslant 2$ is generated by changing angles φ and ψ of a single residue selected. To ensure that the way of choos-

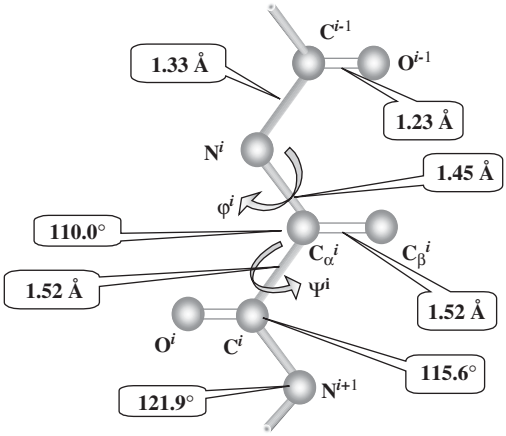


Fig. 14.11. Generation of a conformation for the flexible homopolymer chain. For details see the text.

Table 14.3 The Ramachandran criteria for the sterical collisions used in the calculations

Contact	$d_{ij}^{\min}$, Å
O — O	2.6
O — N	2.6
O — C	2.7
N — N	2.6
N — C	2.8
C — C	2.9
C — C_β	3.0
C_β — C_β	3.0

ing the residue does not influence the outcome of calculations, test simulations were performed with random and consecutive selection of the residue. The results obtained in both these cases do not differ within the limits of statistical accuracy, but the consecutive selection saves computational time. The new conformation is accepted or rejected with the probability (10) [7] with $W(k')$ and $W(k)$ being energies (46) of the conformations k' and k, respectively. If conformation k' is rejected, the conformation k becomes k'. After repeating the above procedure $2\times N_r$ times, the conformation achieved was assigned the index $\tilde{\mathbf{k}}$ (each time increased by one), and the radius of gyration, $R_g(\tilde{\mathbf{k}})$, was calculated according to the definition (33) and stored. The procedure was repeated until $\tilde{\mathbf{k}}$ reaches a certain value LMC and this hereafter is referred to as a single MC run. In order to avoid problems related to the trapping of the system in a metastable state, the calculations were repeated several times, each with another initial conformation of the chain. A final quantity, R_g, comprises then the average over $R_g(\tilde{\mathbf{k}})$ values obtained both within a single MC run (with the first L' conformations discarded) and for different MC runs

$$R_g = \frac{1}{L_R \times (L_{MC} - L')} \sum_{\tilde{\mathbf{k}}=1}^{L_R \times (L_{MC} - L')} R_g(\tilde{\mathbf{k}}) \tag{49}$$

Here L_R is the number of single MC runs. Typical parameters used in the calculations were: L_{MC} = 2 000, L' = 1 000 and L_R = 100.

14.4.3 Dimensions of non-charged chains

The radius of gyration for a flexible chain with excluded-volume interactions obeys the relation $R_g \sim N_r^{\nu}$, where ν is about 3/5 [73]. To introduce the influence of electrostatic interactions on R_g we assume that the average radius of gyration of the charged flexible chain can be approximated as

$$R_g(N_r,\text{pH}, T) \approx R_g^0 (N_r) \times f(W^{\text{el}}) \tag{50}$$

with $R_g^0(N_r)$ being R_g with the absence of electrostatic interactions (or at infinite temperature), and with an N_r-independent function $f(W^{\text{el}})$ describing the correction to R_g which comes from the electrostatic interactions. The function $f(W^{\text{el}})$ is, in general, dependent on external factors such as temperature and pH. As will be shown below, the temperature dependence of $f(W^{\text{el}})$ is weak in the temperature interval of interest. A change in pH causes variations of the net charge, so that pH-dependence of $f(W^{\text{el}})$ can be replaced by dependence on the charge composition. It is interesting to see how different sidechains presented in proteins can influence the value of R_g. For this purpose, MC simulations for homopolymer chains with N_r varying from 10 to 80 and with different sidechains represented by lysine in extended conformation truncated at β, γ, δ, ε and ζ positions have been performed

Variations of R_g^0 with N_r plotted in double logarithmic scale (Fig. 14.12) reveal the power-law behaviour

$$R_g^0 (N_r) = \tilde{R}_g^0 \times N_r^{\nu} \tag{51}$$

with exponents ν and pre-exponential factors $\tilde{R}_g^0$ listed in Table 14.4. As seen, the exponent ν does not depend, within the limits of statistical accuracy, on the type of sidechain used. The value of $\nu \sim 0.66$ is close to the value estimated for a generic flexible chain with the excluded-volume interactions [73]. The multiplier $\tilde{R}_g^0$ can be considered as the radius of gyration of a single amino acid, and hence differs for various sidechains. This deviation, however, is about 10 per cent, which, for example, results in a 3.5 Å difference between the R_g values for polyalanine and polylysine with N_r = 80 (compare with the difference of 7.3 Å between the R_g values for polyalanine and for the flexible chain without sidechains, polyglycine). This result suggests that R_g of a heteropolymer (an unfolded protein) will vary insignificantly with respect to the amino acid composition. As far as $\tilde{R}_g^0$ appears as a multiplier in Eq. (51) and hence in Eq. (50), one can consider a homopolymer (for the sake of computational simplicity we choose polyalanine) without influencing the final conclusions.

The electrostatic interactions are the strong and long-range ones, and hence the distance between the opposite charges could be less than the Ramachandran criteria for the C_β atoms. Calculations for the polyalanine with varied minimal allowed distance between the C_β atoms, $d_{\min}$ (see insert to Fig. 14.12), have, however, revealed that this influence on R_g^0 is even weaker compared with the impact of different sidechains described above.

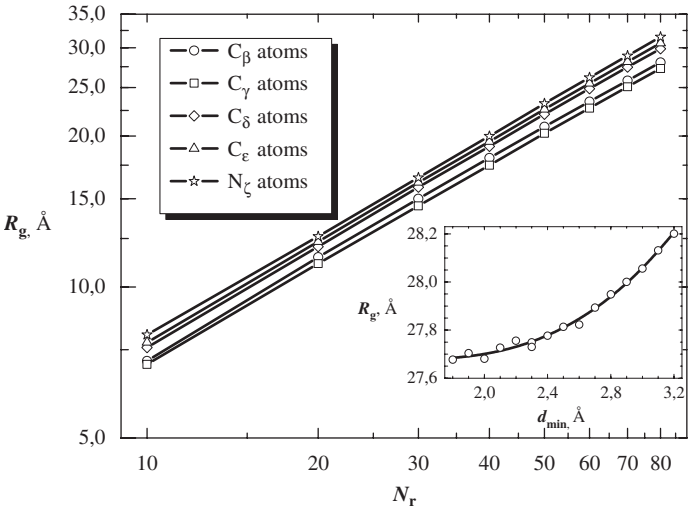


Fig. 14.12. Double logarithmic plot of the radius of gyration, R_g, calculated for the flexible homopolymer chain in the absence of electrostatic interactions, versus the number of residues, N_r, in the chain. Different symbols in the main panel correspond to different last atoms at which the lysine sidechains are truncated. Data were obtained using Ramachandran criteria (Table 14.3) for the sterical collisions between atoms. Insert shows dependence of R_g for the homopolymer with 80 residues and with the only C_β atoms in the sidechains on the sterical collision criterion, d_{min}, between the C_β atoms (all other criteria are the same as for the data in the main panel).

Table 14.4 Exponents and pre-exponential factors in the expression (51) obtained for non-charged chains with the lysine sidechains truncated at different atoms

Last atom of a sidechain	ν	$\tilde{R}_g^0$, Å
	d_{ij}^{min} (C_β — C_β) = 3.0 Å	
C_β	0.6579	1.586
C_γ	0.6532	1.568
C_δ	0.6606	1.663
C_ε	0.6614	1.698
N_ζ	0.6593	1.757
	d_{ij}^{min} (C_β — C_β) = 2.0 Å	
C_β	0.6537	1.593

14.4.4 Dimensions of flexible charged chains

14.4.4.1 Zero net charge

The case of zero net charge can be related to proteins at an isoelectric point where equal amounts of acidic and basic groups bear a charge. Chains with $N_r = 80$ residues and with density of ionizable sites $\rho = 0.25$ were used in the calculations (correspondingly, $N_A = N_B = N_A^- = N_B^+ = 10$). The charges were uniformly distributed along the chain so that a

positive charge (corresponding to a basic group) was followed by a negative (acidic group) one and so on. Calculations were carried out with the permittivity in Eq. (36) both taking a uniform value (corresponding to that of water at 25 ºC, ε_s=78.4) and being a distance-dependent function (48).

The results are summarized in Fig. 14.13. As can be seen, in the case of uniform permittivity, two clearly different states can be distinguished depending on the temperature. The first one (hereafter called the compact unfolded state or the CU-state), observed at low temperatures is characterized by radius of gyration R_g ~ 15–16 Å. This value is close to that observed for proteins of similar sizes in the 'molten globule' and native states. In the low-temperature phase an average distance between nearest opposite charges, $\langle d_{NN} \rangle$ was ~ 8 Å and, on average, seven to eight pairs of charges per conformation were close enough ($\leqslant$ 4.5 Å) to be able to form a salt bridge. The high-temperature state (hereafter denoted as expanded unfolded state or the EU-state) is characterized by an R_g value close to the R_g^0 value for a non-charged chain. Here the average distance $\langle d_{NN} \rangle$ ~ 18.5 Å, and no charges were observed that are capable of forming salt bridges.

The calculations with distance-dependent permittivity have shown a dramatic reduction of R_g (triangles in Fig. 14.13). In the whole temperature interval considered (up to 400 K), the values of R_g are close to the CU-state and weakly depend on temperature. No transition to the EU-state is observed. The calculations revealed that this transition occurs at about 600 K (data not shown). Note that the transition from the CU- to the EU-state occurs in our model at temperatures, T_0, dependent on ε. Insert B to Fig. 14.13 shows rough estimations of the $T_0(\varepsilon)$ function for the case of uniform ε value.

The two electrostatic models provide very different results within the physiologically relevant temperature interval (280–400 K). The values of R_g obtained using uniform permittivity are close to that corresponding to the radius of gyration, R_g^0, in the absence of electrostatic interactions. That would imply a domination of the thermal

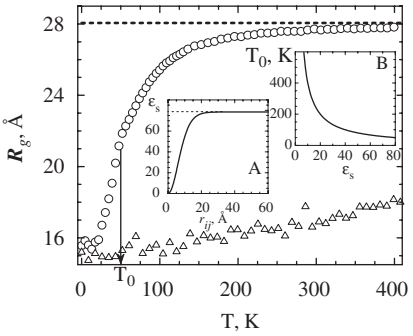


Fig. 14.13. Temperature dependence of the radius of gyration, R_g, calculated for the flexible homopolymer chain with 80 residues, with zero net charge and with the electrostatic interactions (46) between the charges calculated using the uniform value of dielectric permittivity, ε_s = 78.4 (circles), and with the distance-dependent function $\varepsilon(r_{ij})$ (triangles). The dashed line represents the R_g^0 value given by Eq. (51) for the 80-residue chain in the absence of electrostatic interactions. The function $\varepsilon(r_{ij})$ (48) with the saturation value, ε_s =78.4, used in the calculations is shown in the insert. The arrow shows temperature T_0 of the transition between two unfolded states in the case of uniform ε value. Variations of T_0 with the value of ε_s are shown in the insert B.

motion at physiological temperatures, i.e., electrostatic interactions are not relevant for the size of unfolded proteins. The calculations with distance-dependent permittivity suggest an opposite conclusion. The question arises, which of the two models adequately reflects the role of electrostatic interactions in unfolded proteins? The comparison to experimental data indicates that the model with the distance-dependent permittivity is more appropriate, so that this model will be used henceforth.

14.4.4.2 Non-zero net charge

To characterize different charge compositions ($0 \leqslant N_A^- \leqslant N_A$, $N_B^+ = N_B$) and compare the results for different N_r, the normalized net charge, $x = \Delta Q/N_B$ will be used henceforth. The quantity N_B (or N_A, which, in principle, is equivalent) is used here for normalization rather than N_r (or total number of ionizable groups, $N_c = N_A + N_B$), because it allows evaluation of R_g for the general case where $N_A \neq N_B$, using the results of model calculations with $N_A = N_B$. Then, the normalized net charge, x_{model}, of the model chain is equal to the normalized net charge, $x_{protein}$, of the protein under study at the pH value of interest. For each N_A^- value, calculations for five randomly chosen distributions of the negative charge positions over the N_A sites were performed and the results were averaged.

Figure 14.14 shows dimensions of the 80-residue chains as functions of x for the electrostatic interactions (46) calculated with the distance-dependent function (48). As can be seen, the system undergoes a smooth change from a state with prevailing attractive forces ($R_g < R_g^0$) to a state with predominant repulsive forces ($R_g > R_g^0$). Chain dimensions depend weakly on x when $x \leqslant 0.5$, and they are close to those observed at the isoelectric point. This implies that the electrostatic interactions would favour the compact unfolded (CU) state for charge states significantly far away from the isoelectric point, i.e., within a wide pH interval. For $x > 0.6$ the reduction in number of negatively charged groups leads to a statistical decrease of electrostatic attraction reflected by an increase of R_g with x. Electrostatic attraction is statistically compensated for by the electrostatic repulsion at $x = 0.8$ (20 per cent of acidic groups

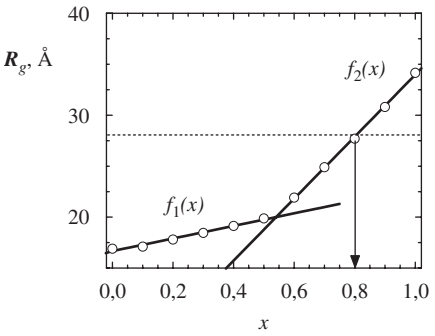


Fig. 14.14. Radius of gyration, R_g, calculated for the flexible homopolymer chain with 80 residues for various values of the normalized net charge, $x = (N_B^+ - N_A^-)/N_B$ net charge and with the electrostatic interactions calculated using the distance-dependent function $\varepsilon(r_{ij})$ displayed in the insert A to Fig. 14.14. Lines represent liner fits, $f_1(x) = 16.7 + 0.6 \times x$ (correlation coefficient is 0.9915) and $f_2(x) = 3.6 + 3.0 \times x$ (correlation coefficient is 0.9995) for $x \leqslant 0.5$ and $x \geqslant 0.6$, respectively. The arrow displays the x value at which electrostatic repulsion statistically is compensated for by the electrostatic attraction.

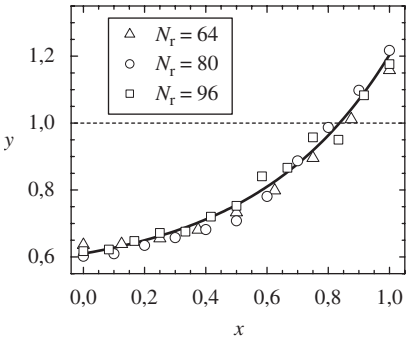


Fig. 14.15. Normalized radius of gyration, $y = R_g/R_g^0$, for the chains of different lengths (see the legend) as function of x. All data points stand for the results of calculations at 25°C and with fixed density of titratable groups $\rho = 0.25$, dashed lines represent the R_g value for the N_r-residue chain in the absence of electrostatic interactions, and the solid line is a guide for the eye.

charged), where the chain dimensions coincide with that of non-charged chains. Two regions of $R_g(x)$ can be distinguished with a crossover for $0.5 \leqslant x \leqslant 0.6$ (Fig. 14.14). The first one, $f_1(x)$ for $x \leqslant 0.5$, corresponds to the CU state, whereas the second one, $f_2(x)$ for $x \geqslant 0.6$, is characterized by a pronounced dependence of $R_g(x)$ on x and a tendency of the system to adopt the EU-state. For practical purposes it is convenient to approximate $R_g(x)$ by two linear functions $f_1(x)$ and $f_2(x)$ as illustrated in Fig. 14.14.

It is noteworthy that the electrostatic correction $f(x)$ in expression (50) is independent of chain size. Fig. 14.15 displays the results for chains of three different lengths plotted as function $y = f(x)$ where $y = R_g/R_g^0$. Indeed, data for all the three chains within the limits of statistical accuracy collapse into a single curve showing that, indeed, the electrostatic correction in the relation (50) is independent of chain size.

14.4.5 Comparison with experimental data.

As far as $f(x)$ depends on the net charge ΔQ, the results for a homopolymer chain with equal amounts of basic and acidic groups can be generalized to an arbitrary number of ionizable groups. Results for chains with, say, larger number of basic groups ($N_B > N_A$) are statistically equivalent to results for a chain with an increased number of acidic groups $N'_A = N_B$, but with $0 \leqslant N_A^- \leqslant N_A$. Then, the radius of gyration, $R_{g,\mathrm{prot}}$, for a protein with $N_{r,\mathrm{prot}}$ residues at the x value of interest can be estimated from the simulation results for a model homopolymer chain, say with $N_{r,\mathrm{model}} = 80$ for the same value of x. Then, the relation (50) reduces to

$$R_{g,\mathrm{prot}}\,(N_{r,\mathrm{prot}}) = R_g(N_{r,\mathrm{model}}) \times \left(\frac{N_{r,\mathrm{prot}}}{N_{r,\mathrm{model}}}\right)^{\nu} \tag{52}$$

where $\nu \sim 0.66$ (see Table 14.4). The expression (52) is obtained by assuming that the pre-exponential factor $\tilde{R}_g^0$ in Eq. (51) is not sensitive to sidechain variety and that electrostatic correction in relation (50) depends on x only.

In order to test whether the generic-model calculations can be used for evaluation of dimensions of unfolded proteins, we compared R_g values measured for three proteins

unfolded by different means with the R_g values predicted by the generic-model calculations. Eq. (51) was used for R_g evaluation in cases of chemical denaturation with GdnHCl; otherwise the relation (52) was employed (for reference, see Table 3 in Ref. [56]).

14.4.5.1 Cytochrome c

The sequence of this 104-residue protein has $N_B = 21$ basic and $N_A = 12$ acidic groups. The folding/unfolding studies [74] of cytochrome *c* revealed that the protein at pH 3 is a mixture of native state with $R_g = 13.9$ Å and of compact unfolded state with $R_g = 18.1$ Å. The net charge of cytochrome *c* at this pH is +10 [75]. This corresponds to $x = 0.48$ and the relation (52) gives $R_{g,\text{prot}} \sim 20$ Å. At pH < 2 the cytochrome c is in the fully unfolded state with $R_g = 30.1$ Å [76] and has the net proton charge +17 [75]. Now $x = 0.80$ and the relation (52) gives $R_{g,\text{prot}} \sim 33$ Å. The above experimental data are in good agreement with the results of the model calculations for the polypeptide with excess positive charges (see Fig. 14.14), which suggest that at $x = 0.8$ an unfolded protein has dimensions similar to that of the random-coil polypeptide.

In the study [77], a value of $R_g \sim 31$ Å has been obtained for cytochrome *c* unfolded at pH 7 by GdnHCl. It is known that GdnHCl screens electrostatic interactions in proteins, and therefore R_g for unfolded cytochrome *c* at pH 7 in GdnHCl is greater than R_g at pH 3. Power law (51) for a flexible chain without electrostatic interactions gives R_g^0 values between 33–38 Å, which is in good agreement with the experimental data. The range of values rather than a single value comes from deviations in the pre-exponential factor in relation (51) for different sidechains.

It is worth noting that the value of R_g corresponding to chemically unfolded cytochrome *c* at neutral pH equals that corresponding to pH-induced unfolding. The factors determining R_g for the different unfolding conditions are different. In the first case one assumes screening of charge–charge interactions, which diminishes their contribution. In the case of pH-induced unfolding, electrostatic interactions are far from being negligible, but attraction and repulsion effects are statistically compensated.

14.4.5.2 SNase

Staphylococcal nuclease is a highly charged protein with $N_{r,\text{prot}} = 149$ residues. The sequence comprises $N_B = 28$ basic and $N_A = 20$ acidic groups, which gives a value of $x = 0.29$ when all ionizable groups of SNase are charged. Titration experiments indicate that in solutions of potassium chloride this is achieved at pH 6–8 [78]. Study [79] of the truncated form of SNase with 13 amino acids from SNase carboxyl terminus deleted has yielded a value of $R_g = 21.2$ Å for an unfolded state at pH 7.5 and temperature 20 ºC. That is in good correlation with the value $R_{g,\text{prot}} \sim 24$ Å given for the truncated SNase by the expression (52) (here $N_{r,\text{prot}} = 136$, $N_B = 26$, $N_A = 19$ and $x = 0.27$ correspond to the truncated SNase). Unfolding of SNase at pH 5.5 induced by a pressure jump from 1 000 to 2 400 bar also results in an unfolded state with $R_g = 22$–24 Å [35] and the relation (52) evaluates $R_{g,\text{prot}} \sim 26$ Å for the compact unfolded state of SNase, in good correlation with the experimental observations.

It should be noted that the apparent radius of gyration measured in the study [35] for the temperature-induced unfolded state of SNase, is $R_g \sim 43$ Å at pH 7 and that value seemingly correlates with the range of values 42 Å $\leqslant R_g^0 \leqslant$ 48 Å predicted by the relation

(51) for the EU-state. However, the relation (51) describes polypeptides without electrostatic interactions, whereas the conditions of the experiments (pH value and absence of strong chemical denaturants in solutions) described in the work by Panick *et al.* [35] are not consistent with the suggestion about the insignificance of electrostatic interactions. This apparent contradiction is due to the fact that under the experimental conditions unfolded SNase most likely adopts a two-domain or dumbbell form [35].

14.4.5.3 Hen-egg white lysozyme

This protein contains 129 residues of which there are $N_B = 16$ basic and $N_A = 9$ acidic groups. It has been shown that in the urea-induced unfolded state lysozyme has $R_g = 21.9$ Å at pH 2.9 [80,81] and $R_g = 23.5$ Å at low (0.6 M) concentration of the GdnHCl at pH 5.2 [82]. Those values are consistent with the value $R_{g,\text{prot}} \sim 23$ Å predicted by the Eq. (52) for the x range of the compact unfolded state. However, at pH 2.9 the protein has a net proton charge +13 [78], which gives $x = 0.8$. For these conditions one expects a value of R_g corresponding to a flexible chain without electrostatic interactions. Lysozyme maintains the CU-state in the case of GdnHCl-induced denaturation as well. It has been shown, for instance, that after chemical unfolding (4M GdnHC) [83] the protein has $R_g = 22.9$ Å. Again, this value is much lower than the expected one, $R_g^0 = 39$–43 Å. The authors of the work [83] have attributed this discrepancy to the presence of four disulphide bonds, which restrict the rotational freedom of the polypeptide chain. In support of this idea, the results of R_g measurements were reported in the same work for the carboxymethylated state of the lysozyme where disulphide bonds are reduced. Then, at pH 2.0 $R_g = 37.9$ Å for an unfolded state in 20 mM HCl solution and $R_g = 35.8$ Å for the unfolded state in 4M guanidine hydrochloride [83] in good correlation with the values predicted by the relation (52) at $x = 0.8$ and by the relation (51) for the random-coil chain, respectively.

14.4.6 Concluding remarks

The Monte Carlo studies surveyed in this section suggest that unfolded proteins adopt two distinct states depending on charge content and temperature. In the first one, the compact unfolded state is observed at physiologically relevant temperatures and within a wide pH range around the protein isoelectric point, where attractive charge–charge interactions tend to reduce the radius of gyration to a value close to R_g of native proteins. Thus, charge–charge interactions enforce unfolded proteins to collapse into a compact body, as was suggested earlier [18].

It is interesting to note that unfolded lysozyme with intact disulphide bonds has R_g that equals the corresponding value calculated by Eq. (52) for low values of x (strong electrostatic attraction). The calculations suggest, as well, that about seven salt bridges are expected in the CU-state for a polypeptide chain with a high density of titratable groups. A value of four to five salt bridges has been estimated previously [84]. Combining these two observations, one can assume that the CU-state of proteins is maintained by the salt bridges, which act as cross-links in a similar way that disulphide bonds do.

Beyond the pH range where the CU-state is observed, increasing repulsive charge–charge interactions cause a rapid increase of R_g and unfolded proteins tend to

adopt the expanded denatured state with dimensions similar to those of a polypeptide chain without electrostatic interactions (random-coil protein). Charge–charge interactions in this state are also significant, but the electrostatic repulsion is statistically compensated for by the attraction coming from a remarkably small amount (~ 20 per cent) of opposite charges.

14.5 REFERENCES

1 N. Metropolis and S. Ulam, J. Am. Stat. Assoc., 44 (1949) 335.
2 S.M. Ulam and J. von Neumann, Bull. Amer. Math. Soc., 53 (1947) 1120.
3 G. Buffon, Histoire de l'Acad. Roy. des Sci., 1 (1773) 43.
4 G. Buffon, Essai d'arithmétique morale, 4 (1777) 46.
5 P. Sabri, published on Internet: http://www.geocities.com/CollegePark/Quad/2435/history.html, 2004.
6 N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, J. Chem. Phys., 21 (1953) 1087.
7 K. Binder, Monte Carlo methods in statistical physics, Springer-Verlag, Berlin, 1979.
8 D.P. Landau and K. Binder, A guide to Monte Carlo simulations in statistical physics, Cambridge Univ. Press, Cambridge, 2000.
9 E. Ising, Z. Physik, 31 (1925) 253.
10 A.P. Lyubartsev, A.A. Martsinovski, S.V. Shevkunov and P.N. Vorontsovvelyaminov, J. Chem. Phys., 96 (1992) 1776.
11 A. Irback and E. Sandelin, J. Chem. Phys., 110 (1999) 12256.
12 P.H. McPherson, M.Y. Okamura and G. Feher, Biochimica Et Biophysica Acta, 934 (1988) 348.
13 P. Beroza, D.R. Fredkin, M.Y. Okamura and G. Feher, Biophys. J., 68 (1995) 2233.
14 D. Bashford and M. Karplus, J. Phys. Chem., 95 (1991) 9556.
15 C. Tanford, Physical chemistry of macromolecules, Wiley, New York, 1961.
16 A. Koumanov, H. Ruterjans and A. Karshikoff, Proteins, 46 (2002) 85.
17 G.M. Ullmann and E.W. Knapp, Eur. Biophys. J. Biophys. Lett., 28 (1999) 533.
18 C.N. Pace, R.W. Alston and K.L. Shaw, Protein Sci., 9 (2000) 1395.
19 M. Oliveberg, S. Vuilleumier and A.R. Fersht, Biochemistry, 33 (1994) 8826.
20 M. Oliveberg, V.L. Arcus and A.R. Fersht, Biochemistry, 34 (1995) 9424.
21 Y.J. Tan, M. Oliveberg, B. Davis and A.R. Fersht, J. Mol. Biol., 254 (1995) 980.
22 B. Kuhlman, D.L. Luisi, P. Young and D.P. Raleigh, Biochemistry, 38 (1999) 4896.
23 M. Schaefer, M. Sommer and M. Karplus, J. Phys. Chem. B, 101 (1997) 1663.
24 J. Warwicker, Protein Sci., 8 (1999) 418.
25 A.S. Yang and B. Honig, J. Mol. Biol., 237 (1994) 602.
26 A.H. Elcock, J. Mol. Biol., 294 (1999) 1051.
27 P.J. Kundrotas and A. Karshikoff, Protein Sci., 11 (2002) 1681.
28 P.J. Kundrotas and A. Karshikoff, Phys. Rev. E, 6501 (2002) 011901.
29 H.X. Zhou, J. Phys. Chem. B, 106 (2002) 5769.
30 H.X. Zhou, Biochemistry, 41 (2002) 6533.
31 H.X. Zhou, Proc. Natl. Acad. Sci. USA, 99 (2002) 3569.
32 A. Warshel, S.T. Russell and A.K. Churg, Proc. Natl. Acad. Sci. USA, 81 (1984) 4785.
33 H.B. Bohidar, Biopolymers, 45 (1998) 1.
34 M. Hirai, S. Arai, H. Iwase and T. Takizawa, J. Phys. Chem. B, 102 (1998) 1308.
35 G. Panick, R. Malessa, R. Winter, G. Rapp, K.J. Frye and C.A. Royer, J. Mol. Biol., 275 (1998) 389.
36 M. Beltramini, P. Di Muro, R. Favilla, A. La Monaca, P. Mariani, A.L. Sabatucci, B. Salvato and P.L. Solari, J. Mol. Struct., 475 (1999) 73.
37 G. Panick, R. Malessa and R. Winter, Biochemistry, 38 (1999) 6512.
38 G. Caracciolo, G. Amiconi, L. Bencivenni, G. Boumis, R. Caminiti, E. Finocchiaro, B. Maras, C. Paolinelli and A.C. Castellano, Eur. Biophys. J. Biophys. Lett., 30 (2001) 163.

39 J. Perez, P. Vachette, D. Russo, M. Desmadril and D. Durand, J. Mol. Biol., 308 (2001) 721.
40 M.M. Teeter, A. Yamano, B. Stec and U. Mohanty, Proc. Natl. Acad. Sci. USA, 98 (2001) 11242.
41 H. Jang, C.K. Hall and Y.Q. Zhou, Biophys. J., 82 (2002) 646.
42 S.V. Kazakov, I.Y. Galaev and B. Mattiasson, Int. J. Thermophys., 23 (2002) 161.
43 A. Lazaridou, C.G. Biliaderis, T. Roukas and M. Izydorczyk, Appl. Biochem. Biotechnol., 97 (2002) 1.
44 C. Tanford and J.G. Kirkwood, J. Am. Chem. Soc., 79 (1957) 5333.
45 P.J. Kundrotas and A. Karshikoff, BBA–Proteins and Proteomics, 1702 (2004) 1.
46 H.X. Zhou, J. Am. Chem. Soc., 125 (2003) 2060.
47 P.J. Kundrotas, S. Lapinskas and A. Rosengren, Phys. Rev. B, 52 (1995) 9166.
48 J.Y. Lee and J.M. Kosterlitz, Phys. Rev. Lett., 65 (1990) 137.
49 J.B. Matthew, Annu. Rev. Biophys. Biomol. Struct., 14 (1985) 387.
50 J. Aqvist, H. Luecke, F.A. Quiocho and A. Warshel, Proc. Natl. Acad. Sci. USA, 88 (1991) 2026.
51 K.B. Wong, J. Clarke, C.J. Bond, J.L. Neira, S.M.V. Freund, A.R. Fersht and V. Daggett, J. Mol. Biol., 296 (2000) 1257.
52 D. Sali, M. Bycroft and A.R. Fersht, Nature, 335 (1988) 740.
53 S. T. Whitten and B. Garcia-Moreno, Biochemistry, 39 (2000) 14292.
54 M. Tollinger, J.D. Forman-Kay and L.E. Kay, J. Am. Chem. Soc., 124 (2002) 5714.
55 M. Tollinger, K.A. Crowhurst, L.E. Kay and J.D. man-Kay, Proc. Natl. Acad. Sci. USA, 100 (2003) 4545.
56 P.J. Kundrotas and A. Karshikoff, J. Chem. Phys., 119 (2003) 3574.
57 D. M. Jacobs, A.S. Lipton, N.G. Isern, G.W. Daughdrill, D.F. Lowry, X. Gomes and M.S. Wold, J. Biomol. NMR, 14 (1999) 321.
58 P. Holliger, L. Riechmann and R.L. Williams, J. Mol. Biol., 288 (1999) 649.
59 R.B. Sutton, D. Fasshauer, R. Jahn and A.T. Brunger, Nature, 395 (1998) 347.
60 D. Stigter, D.O.V. Alonso and K.A. Dill, Proc. Natl. Acad. Sci. USA, 88 (1991) 4176.
61 G.N. Ramachandran and V. Sasisekharan, Adv. Prot. Chem., 28 (1968) 283.
62 P. McCaldon and P. Argos, Proteins, 4 (1988) 99.
63 R. Gabler, Electrical interactions in molecular biophysics, Academic Press, New York, San Francisco, 1978.
64 A. Warshel and M. Levitt, J. Mol. Biol., 103 (1976) 227.
65 B.R. Gelin and M. Karplus, Biochemistry, 18 (1979) 1256.
66 R.W. Pickersgill, Protein Eng., 2 (1988) 247.
67 A. Warshel, Photochem Photobiol., 30 (1979) 285.
68 B.E. Conway, O.M. Bockris and I.A. Ammar, Trans. Faraday Soc., 47 (1951) 756.
69 A. Warshel, S.T. Russell and A.K. Churg, Proc. Natl. Acad. Sci. USA, 81 (1984) 4785.
70 J. Mazur and R.L. Jernigan, Biopolymers, 31 (1991) 1615.
71 L. Sandberg and O. Edholm, Proteins, 36 (1999) 474.
72 B.E. Hingerty, R.H. Ritchie, T.L. Ferell and J.E. Turner, Biopolymers, 24 (1985) 427.
73 M. Doi and S. F. Edwards, The theory of polymer dynamics, Oxford University Press, Oxford, 1986.
74 L. Pollack, M.W. Tate, N.C. Darnton, J.B. Knight, S.M. Gruner, W.A. Eaton and R.H. Austin, Proc. Natl. Acad. Sci. USA, 96 (1999) 10115.
75 L. Konermann and D.J. Douglas, Biochemistry, 36 (1997) 12296.
76 Y.O. Kamatari, *et al.,* J. Mol. Biol., 259 (1996) 512.
77 D.J. Segel, A. Bachmann, J. Hofrichter, K.O. Hodgson, S. Doniach and T. Kiefhaber, Biochemistry, 37 (1998) 12443.
78 D.E. Kuehner, J. Engmann, F. Fergg, M. Wernick, H.W. Blanch and J.M. Prausnitz, J. Phys. Chem. B, 103 (1999) 1368.
79 J.M. Flanagan, M. Kataoka, D. Shortle and D.M. Engelman, Proc. Natl. Acad. Sci. USA, 89 (1992) 748.
80 B. Ibarra Molero and J.M. Sanchez Ruiz, Biochemistry, 36 (1997) 9616.
81 L.L. Chen, G. Wildegger, T. Kiefhaber, K.O. Hodgson and S. Doniach, J. Mol. Biol., 276 (1998) 225.
82 D.J. Segel, A. Bachmann, J. Hofrichter, K.O. Hodgson, S. Doniach and T. Kiefhaber, J. Mol. Biol., 288 (1999) 489.
83 M. Hoshino, Y. Hagihara, D. Hamada, M. Kataoka and Y. Goto, FEBS Lett., 416 (1997) 72.
84 V.Z. Spassov, A.D. Karshikoff and R. Ladenstein, Protein Sci., 3 (1994) 1556.

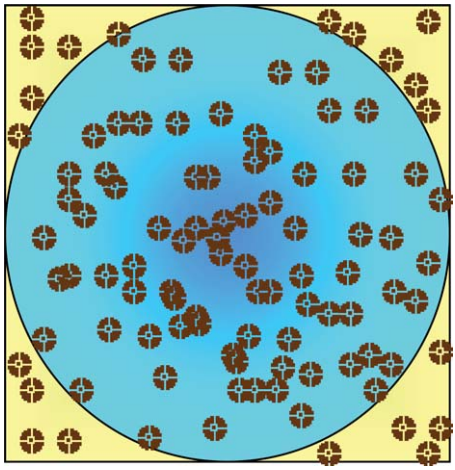


Plate 22. The simple dart board after throwing 100 darts (illustration appears on page 271 of this volume).

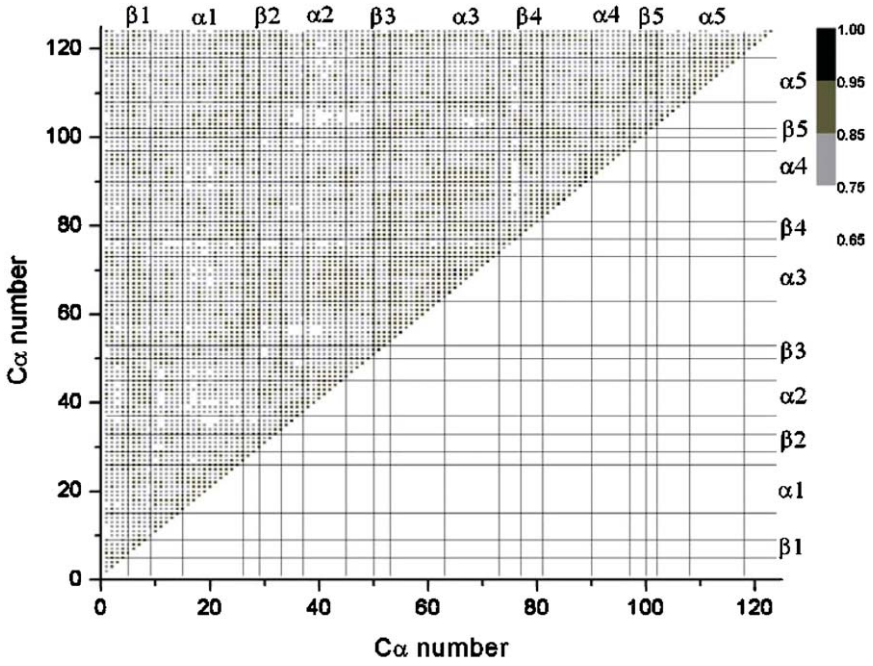


Plate 23. Cross-correlation plot of inter-C_α distances for the first cluster of molecular dynamics conformational snapshots of Spo0F performed in implicit solvent. The shading reflects the degree of correlation attributred to the individual variables. Reprinted from [22] with permission (illustration appears on page 315 of this volume).

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 15

Protein structure generation and elucidation: applications of automated histogram filtering cluster analysis

Heather L. Gordon[1] and Stuart M. Rothstein[2]

[1] *Department of Chemistry and Centre for Biotechnology, Brock University, St. Catharines, Ontario, L2S 3A1, Canada*
[2] *Departments of Chemistry and Physics, Brock University, St. Catharines, Ontario, L2S 3A1, Canada*

Abstract

This chapter reviews automated histogram filtering (AHF), a recently developed approach to cluster analysis, and describes two applications of this methodology, illustrating its power as a tool for analysis of biomolecular simulation data. The first is to cluster low-energy conformations of a model protein, generated by Monte Carlo-simulated annealing. From ensuing cluster memberships, we arrive at distance constraints to serve as input to distance-geometry software. Additional low-energy structures are subsequently generated by the latter software with remarkable efficiency. The second application is to cluster molecular dynamics snapshots of an all-atom model of the response regulator protein Spo0F of *Bacillus subtilis.* Two clusters are uncovered, reflecting our having sampled two distinct basins on the potential energy surface. We subject snapshots for each cluster to AHF cluster analysis, and perform an orthogonal rotation of the dominant so-called 'principal components'. We monitor the loadings of each interatomic distance variable along the rotated principal components. To visualize our results, we construct squared-loadings plots, revealing secondary structural elements that undergo concerted motions along so-called 'collective coordinates'. Dominant collective coordinates are found to be associated with contiguous stretches of amino acid residues implicated in the biological functioning of Spo0F.

15.1 INTRODUCTION

Analysis of high-dimensional data necessitates innovative approaches to data clustering in conjunction with powerful multidimensional data reduction and visualization techniques.

Our interest in this area is to efficiently generate and extract physically relevant information from data that arise in the simulation of biopolymers.

Our data consist of a large number of Monte Carlo configurations or molecular dynamics snapshots, each described by variables chosen such that simular configurations have similar variables. In the case of isolated molecules, the interatomic distances are an effective choice of variables [1].

In this chapter we describe automated histogram filtering (AHF) cluster analysis [2] and provide examples of its successful application in protein structure elucidation. Central to this approach is principal component analysis, a well-known dimensionality-reduction technique for multivariate data analysis. Principal components are linear combinations of the variables, such that the first principal component accounts for as much of their variance as possible, the second as much of the remaining variance, etc. Projections of the variables along the principal component axes are designated their 'scores'.

Clusters are sets of variables that lie close together in multidimensional space. These project together along one or more of the principal component axes. Thus, the presence of clusters is reflected by peaks in histograms of scores plotted along principal component axes. Automated histogram filtering cluster analysis entails virtually constructing and iteratively resolving these peaks.

AHF cluster analysis is a new technique, to date successfully applied to two problems in protein structure generation and elucidation. In the first application, we used AHF-clustering in tandem with Monte-Carlo-simulated annealing and distance-geometry software to generate low-energy conformations with remarkable efficiency for a 69- 'residue' off-lattice model protein [2]. The second application was to an all-atom model of the response regulator protein Spo0F of *Bacillus subtilis* [22]. Here, we AHF-clustered molecular dynamics conformational snapshots. To visualize our results, we constructed squared-loadings plots, revealing secondary structural elements that are undergoing concerted motions along so-called collective coordinates. Collective coordinates exhibiting the greatest conformational variability were, for the most part, found to be associated with contiguous stretches of amino acid residues implicated in the biological function of Spo0F.

This chapter is organized as follows. First we describe the fundamental concepts of automated histogram filtering, Section 15.2. The method rests on principal component analysis, introduced in Section 15.3. This is followed by a discussion of the filtering algorithm and its computer time-complexity, Section 15.4. The first application, to efficiently generate low-energy conformations of a model protein, is described in Section 15.5. The second application, identification of collective coordinates of Spo0F is the focus of Section 15.6. We conclude with an overview and summary, Section 15.7.

15.2 AHF-CLUSTERING: FUNDAMENTAL CONCEPTS

To illustrate the fundamental concepts that underlie AHF clustering, consider the case of 23 'configurations', described by two variables (parameters); shown in Fig. 15.1 (a), top. It is immediately obvious that the configurations are grouped into three 'clusters'. Furthermore, the variables are positively correlated, making it more difficult to extract

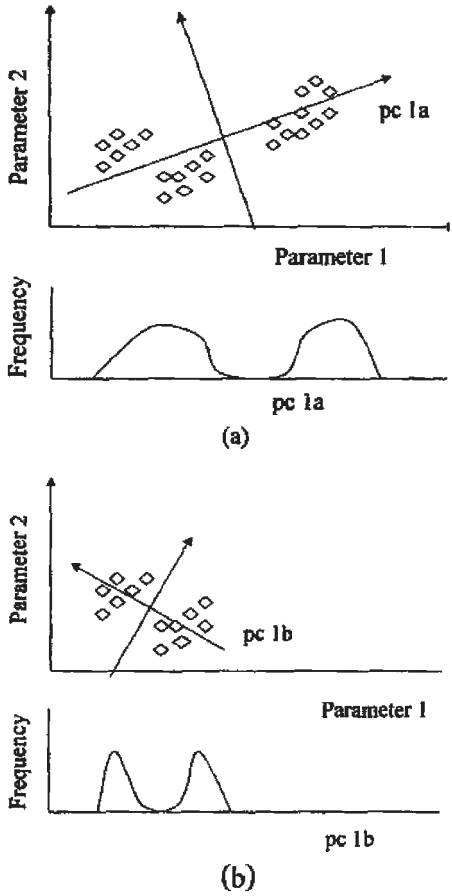


Fig. 15.1. (a) The first principal component (pc 1a) captures as much of the data variation as possible; top. Histogram of scores on the first principal component reveals two peaks; bottom. Filtering the right-hand peak isolates the rightmost cluster of data. (b) Data isolated from the left-hand histogram in (a) are shown at the top. The histogram of scores along the first principal component (pc 1b) is bi-modal; bottom. Filtering these peaks identifies the two leftmost data clusters appearing at the top of (a). Reprinted from [2] with permission.

the clusters *via* algorithms that describe the data values within the original, untransformed basis or coordinate system.

In AHF-clustering, we first subject the data to a principal component analysis (PCA) [3–6], and Section 15.3, below. Here the first principal component, pc 1a, Fig. 15.1(a), captures as much of the data variation as possible, while the second, orthogonal to it, describes the remaining variance.

Now, each variable is assigned 'scores', its projections onto the principal component axes. Here, variable scores along the first principal component accumulate in two segments, reflected by two peaks in the histogram of scores: Fig. 15.1(a), bottom. This illustrates the concept at the core of histogram filtering: *members of clusters have scores*

that in turn congregate in segments along principal component axes. These are reflected by peaks in histograms, plotted along the principal component axes.

We extract or 'filter' the data subset whose scores appear in the right-hand histogram peak. This isolates the rightmost cluster of configurations. The remaining two clusters are identified in a similar manner. A second PCA is performed on the data subset whose scores constitute the left-hand peak of pc 1a's histogram: Fig. 15.1(b), top. Again the scores accumulate in two distinct segments along the first principal component (pc 1b), reflected by a bimodal histogram of scores: Fig. 15.1(b), bottom. Filtering the data constituting these two peaks isolates the remaining two data clusters.

15.3 OVERVIEW OF PRINCIPAL COMPONENT ANALYSIS

Automated histogram filtering cluster analysis exploits the power of principal component analysis to detect subtle effects in highly correlated multivariate data. The data consist of a large number of snapshots, taken along a molecular dynamics trajectory, or a large number of Monte Carlo configurations. For isolated molecules, the best choice of variables on which to perform a cluster analysis are the inter-residue distances: inter-bead distances for simple model proteins, and inter-C_α distances for all-atom models [1]. In this section, we present some basic equations for principal component analysis, following the approach of [6].

Principal component analysis transforms a set of n correlated variables $\mathbf{x}$ to n new, uncorrelated principal components, or 'factors' $\mathbf{y}$, such that

$$\mathbf{y} = \mathrm{x}\mathbf{E} \tag{1}$$

Here $\mathbf{E}$ is the matrix of normalized eigenvectors, whose j th column corresponds to the j th largest eigenvalue λ_j of the variance–covariance matrix of $\mathbf{x}$.

This equation means the coordinates of any point measured on a principal component axis are the coordinates of that point measured on the variable axis, multiplied by the matrix of eigenvectors. The y coordinates are the component 'scores', projections of each inter-residue distance onto the principal component axis. These are the quantities for which histograms are constructed and virtually inspected for peaks. For example, variable projections along principal component 1a (Fig. 15.1(a)) accumulate within two non-overlapping line segments, giving rise to a bimodal histogram, plotted along the principal component axis.

The principal components, $\mathbf{y}$, and the variables, $\mathbf{x}$, are normalized to unit variance, yet the components have variances described by $\mathbf{\Lambda}$, the matrix of eigenvalues. The matrix of component 'loadings' gives the weighted relationship of the principal components to the original variables:

$$\mathbf{L} = \mathbf{E}\mathbf{\Lambda}^{1/2} \tag{2}$$

The element l_{ij} of this matrix is the correlation coefficient between the ith variable, x_i, and the jth principal component, y_j. These component or factor loadings can be interpreted as projections of the unit variable vectors onto the principal components. Furthermore, the

squared-loading, l_{ij}^2, represents the proportion of the variance of the ith variable explained by the jth principal component:

$$\sum_{i=1}^{n} l_{ij}^2 = \lambda_j \tag{3}$$

In our approach to relating dominant variances to structural elements in proteins, principal component vectors are subjected to orthogonal rotation about the origin until each is maximally co-linear with a distinct cluster of variable vectors. Thus, *we sacrifice components which maximize the total variance for ones having separate groups of highly intercorrelated variables.*

Among the possibilities in the literature, we choose the varimax rotation, where the transformed loadings maximize the following quantity [4]:

$$\text{VMAX} = \sum_{i=1}^{n} \sum_{j=1}^{p} l_{ij}^4 - \frac{1}{\text{p}} \sum_{i=1}^{n} (\sum_{j=1}^{p} l_{ij}^2)^2 \tag{4}$$

where p is the number of factors entering the analysis, sufficient to account for the bulk of the sample variance.

15.4 AHF CLUSTERING: THE ALGORITHM

In automated histogram filtering, principal components (PCs) are considered sequentially until more than one peak is detected in the histogram of interatomic distance variable scores plotted along the principal component axes. When multiple peaks are detected, the data subsets (molecular dynamics snapshots or Monte Carlo configurations) that constitute each peak are isolated one at a time. These are again subjected to a PCA, followed by a further round of histogram construction, inspection and filtering. The PC histograms are iteratively resolved (or filtered) in this manner. Thus our scheme is akin to hierarchical clustering with several levels of analysis.

We have found effective any of the following three termination criteria for scanning histograms for peaks: (1) whenever one has processed the 10th PC without finding multiple peaks; (2) the PC under consideration accounts for less than 10 percent of the sum of the previous contributions to the sample variance; or (3) the number of observations remaining to be filtered is less than 15. Each of the data subsets remaining after one of these termination criteria is met constitutes a ‘cluster’.

Histogram construction and inspection for peaks uses an algorithm that rests on user-defined parameters, predetermined by Monte Carlo studies on distributions suggestive of the data at hand [2]. One parameter, Δ, is used to distinguish two overlapping peaks from a single peak: a population drop from the first maximum to the following minimum must be larger than Δ; the subsequent rise from the minimum up to the next maximum must also exceed Δ. Another important parameter is the number of bins used to construct histograms, which depends on the number of conformations entering into the analysis.

Having to optimize a small set of algorithmic parameters is a small price to pay for the convenience and computational power afforded by computer-automating the clustering procedure. The method’s main strength rests on the data set becoming more homogeneous

as one continues to filter. Groups of data with different and subtle inter-variable correlations are thus teased apart from one another.

Efficiency of our AHF computer codes rests exclusively on the speed of the matrix diagonalizations performed in principal component analysis of data subsets. In general, diagonalization of an $(n \times n)$ matrix scales as n^3. Again, our variables are the inter-atomic distances, where for a protein, the inter-C_α distances constitute the most parsimonious choice. Thus, we are diagonalizing $(n_C^2 \times n_C^2)$ matrices. *This underlying n_C^6 scaling of CPU time with system size is the most significant computational challenge facing applications of clustering biomolecular simulation data by AHF.*

Upon implementing ScaLAPACK parallel matrix diagonalization codes [7], AHF-clustering was efficient on distributed memory computers up to at least 64 nodes for a 124 amino-acid-residue protein [8], a system described in Section 15.6, below. Research is in progress on larger systems, where matrix elements are distributed among several processors, rather than residing on each of a large number of individual processors. Clustering, by AHF, of molecular dynamics conformations for a 200 amino-acid-residue protein was performed in this manner; each level of analysis consumed an hour of wall-clock time on 32 400 MHz processors on a shared memory machine [9].

There are important methodological differences between AHF cluster analysis and our previous implementation of classical, agglomerative hierarchical cluster analysis in tandem with *k*-means clustering in order to cluster conformations of a small model protein obtained by Monte Carlo simulated annealing [2]. In agglomerative hierarchical analysis, clusters are formed by grouping variables into larger and larger clusters until all variables are members of a single cluster. The ensuing clustering dendrogram must be visually examined, and from this one gleans the number of clusters present, K. The *k*-means clustering partitions of the same data set, using the current value of K to partition the data. Thus, we obtained a consensus as to the memberships of the resultant clusters of molecular conformations.

The independent use of two classical clustering techniques to reach a consensus was used in [2] to minimize the effects of interpretative subjectivity inherent in both methods. Conversely, AHF clustering runs are reproducible, eliminating the subjectivity inherent in ascertaining the number of clusters from a dendrogram. This is a significant advantage of using AHF cluster analysis. Another essential difference is that AHF identifies clusters from an iterative analysis of the peaks of histograms of individual principal components considered sequentially in order of decreasing importance to the sample variance. The classical approach rests on minimizing the sum of squared distances between cluster members and the centroids of the cluster, where the variables are transformed using several principal components, despite some being much more important in describing the data than others.

In classical methodologies, there are many methods for calculating distances between objects, and many methods for combining objects into clusters [10]. Our approach associates clusters with peaks in score histograms, plotted along principal component axes. Given the complexity of the data arising from simulation of biomolecules, in most cases there are no experimental tests or validations to which the ensuing clusters may be subjected beyond the utility of the ensuing conformational clusters. It is clear, however,

that our approach and classical ones that exploit the dimensionality-reductive power of principal component analysis are equally costly to implement: the n_C^6 scaling of CPU time with number of amino acid residues required to diagonalize a matrix of inter-residue distances is common to both approaches.

15.5 EFFICIENT GENERATION OF LOW-ENERGY PROTEIN CONFORMATIONS

Often it is assumed that the biologically active (native) state of a protein is the conformation that rests at the global minimum on the potential energy surface. It is now believed that the *distribution* of conformations, not just this single one, is required for a more complete understanding of the complexity of the energy landscape and how it relates to protein function [11].

In this subsection we describe an application of automated histogram filtering, used in conjunction with distance-geometry software, where we generated additional low-energy conformations of a simplified protein model, starting with a small set of three-dimensional structures derived from Monte-Carlo-simulated annealing (MC-SA) simulations. This approach makes feasible the efficient generation of the distribution of low-energy structures for all-atom proteins, not solely a single conformation at the global minimum energy.

We investigated a previously unexplored, off-lattice model protein consisting of a chain of 69 beads which are either hydrophobic, hydrophilic, or neutral in character, denoted B, L and N, respectively [2]. It has been shown that its 46-bead analogue exhibits multiple low-energy states below the folding temperature and that its kinetics of folding [12] shares many features of real proteins [13]. The protein model in our case has a highly-frustrated potential energy surface, consisting of a vast number of local minima, presenting a challenge for the characterization of the surface and for the determination of its low-energy conformations. To generate a large number of conformations at local minima for this system, using standard methods such as MC-SA, is prohibitively expensive. Our objective was to arrive at an economical way to generate these structures and to characterize the surface.

We collected MC-SA conformations for the model protein and subsequently subjected them to local, conjugate gradient (CG) optimization to arrive at a small set of low-energy structures, i.e., those within the first tenth percentile of all those energies obtained. We used automated histogram filtering to perform a cluster analysis on these data. Using structural data for these clusters, by employing DGEOM [14], a well-known distance-geometry program, we economically generated a large number of new low-energy structures, up to 40 times faster than by doing additional MC-SA-CG runs. We located conformations at both the apparent global minimum (Fig. 15.2) and the fourth lowest rank in the energy spectrum, neither of which had been obtained from MC-SA-CG. By implementing AHF clustering and a distance-geometry approach to conformational generation, we produced over 33 000 unique structures for this system, substantially augmenting the energy distribution obtained by MC-SA alone.

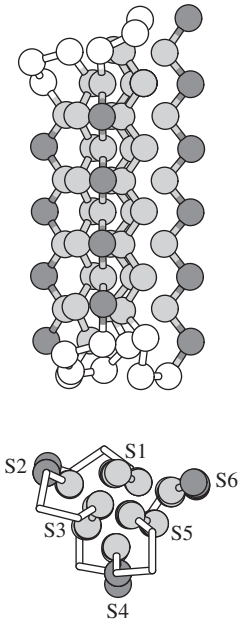


Fig. 15.2. Side and top views of the apparent global minimum structure of the 69 'residue' three-colour or BLN protein model. The white beads mimic neutral, dark grey beads hydrophilic, and light grey beads hydrophobic amino acid residues, respectively. The top view omits spheres for neutral beads at the turns for the sake of clarity. Reprinted from [2] with permission.

15.6 COLLECTIVE COORDINATES FOR AN ALL-ATOM PROTEIN MODEL

Protein backbones and side chains exhibit flexibility on the millisecond to microsecond timescale. It has been proposed that some of these fluctuations relate to protein function [15]. A recent NMR ^{15}N study of the bacterial response regulator protein of *Bacillus subtilis,* Spo0F, is one example in the literature demonstrating significant structural fluctuations of known magnitude and timescale for regions of the protein known to associate with divers binding partners [16].

From a theoretical perspective, functionally-relevant motions occur along the direction of a few so-called 'collective coordinates' [17]. These form the basis of a low-dimensional subspace in which significant internal motion of the protein is expected to occur. This drastic reduction in complexity is being exploited to develop a new generation of molecular modelling and structure-based drug-design tools [18,19]. Furthermore, collective coordinates have proven useful in providing insight into the nature of the protein energy landscape [20], and in improving the sampling of configuration space by molecular dynamics simulations [21].

In collaboration with S. Tanaka, we developed a methodology to extract conformational-basin-specific collective coordinates from MD trajectories [22]. Our application was to an all-atom model of the Spo0F protein (124 amino acid residues). We AHF-clustered inter-residue distances of MD conformational snapshots taken from 1 ns

trajectories, generated both *in vacuo* and using an implicit-solvent representation, starting from the same NMR-derived three-dimensional conformation [16]. In each case we uncovered two clusters, consisting of conformational snapshots reflecting MD sampling of a distinct basin on the potential energy surface.

We performed varimax rotation of the first 100 PCs obtained from inter-residue distances from conformations belonging to each individual cluster, Eq. (4). We computed the loadings for each of the inter-C_α distance variables on each of these 100 PCs, Eq. (2), and found the maximum squared-loading value for each variable. On a $C_\alpha \times C_\alpha$ grid, we plotted the variance described by the PC which contained the maximum squared loading, with colour shading assigned to an individual variable reflecting the amount of variance described. These so-called 'squared-loadings' plots, based on rotated PC vectors, were key to the identification of collective coordinates dominating each of the two clusters.

The power of the AHF clustering/varimax PCA approach is illustrated as follows. The standard approach, construction of inter-C_α distances cross-correlation plots for implicit solvent snapshots within one cluster, appears in Fig. 15.3. The complex and highly dispersed pattern of correlations illustrates the difficulty of locating significant regions of concerted protein motion by this approach.

In our analysis of the same data, squared-loadings plots (Fig. 15.4) depict dominant features by the darkest shading, which suggest variation in inter-residue distances is localized

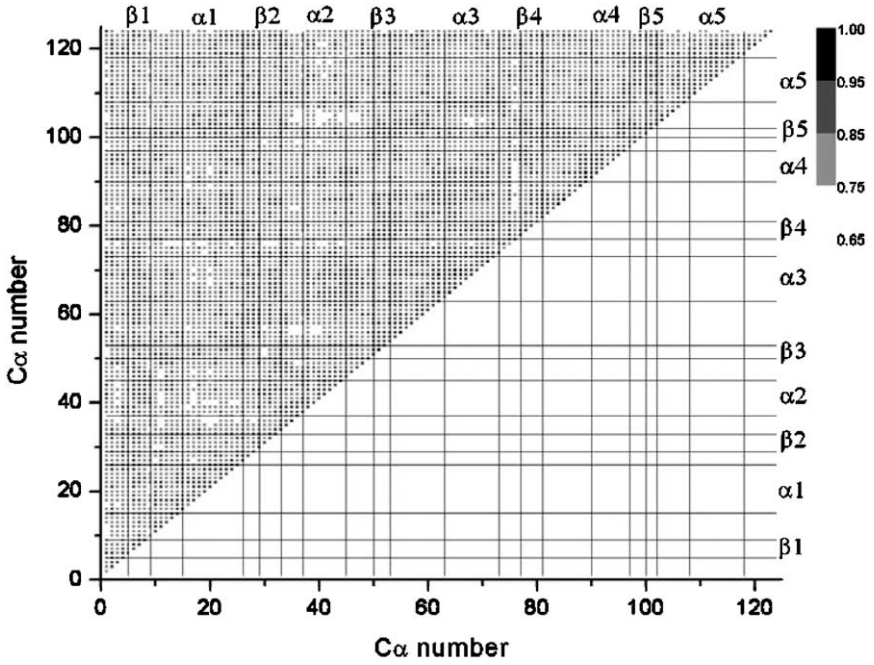


Fig. 15.3. Cross-correlation plot of inter-C_α distances for the first cluster of molecular dynamics conformational snapshots of Spo0F performed in implicit solvent. The shading reflects the degree of correlation attributed to the individual variables. Reprinted from [22] with permission (see Color Plate 23).

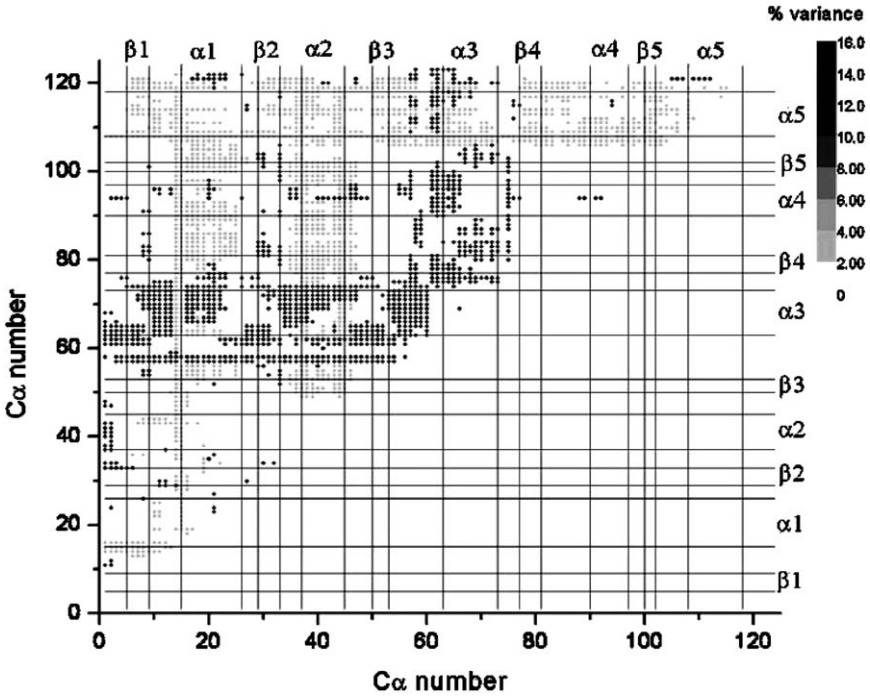


Fig. 15.4. Squared-loadings plot for the first cluster of molecular dynamics conformational snapshots of Spo0F performed in implicit solvent. Varimax rotated squared-loadings greater than or equal to 0.25 within one of the first 100 principal components are shown. The shading reflects the variance described by the principal component that contained the maximum squared loading. Darker points imply that the principal component describes a larger percentage of the total sample variance. No points are plotted for inter-C_α distances whose maximum squared-loading value is either less than 0.25 in any of the first 100 principal components, or if its maximum squared-loading value belongs to a rotated principal component that accounts for less than 2 percent of the total variance. Elements of secondary structure are noted; intervening loops are not labelled, but are designated by horizontal and vertical lines separating adjacent helices and strands. Reprinted from [22] with permission (see Color Plate 24).

to those involving the $\alpha 3$ helix and the rest of the protein. This concerted motion is not detectable by visual inspection of the molecular dynamics trajectory of conformations nor in the convential cross-correlation plot, Fig. 15.3. From this we infer concerted motion of these mobile secondary structure elements along their collective coordinates.

For the simulations of Spo0F employing an implicit-solvent representation, collective coordinates obtained in this fashion were consistent with the location of the Spo0F protein's known active sites and experimentally determined mobile regions [16].

15.7 OVERVIEW AND SUMMARY

Innovative approaches to data clustering in conjunction with powerful multidimensional data reduction techniques are central to meeting the challenge of extracting physically

relevant information from high-dimensional simulation data that emanate from investigations in computational biophysics.

In this chapter we have reviewed applications of automated histogram filtering (AHF) cluster analysis [2,22]. Here the variables of interest are inter-atomic distances in data consisting of molecular dynamics snapshots or a set of Monte Carlo conformations. AHF cluster analysis rests on the fact that clusters of variables project in concert along principal component (PC) axes; these projections are denoted as 'scores'. Scores for a cluster of variables accumulate in segments along one or more principal component axes, reflected by a single peak in the histogram of scores, plotted along the principal component axes. Multi-peaked histograms arise due to the presence of at least two clusters.

AHF-clustering codes construct and virtually inspect histograms for scores along each principal component, in order of the principal component's ability to explain the data. When multiple peaks are detected, the data subsets that constitute each peak are isolated (filtered) one at a time. Each subset is again subjected to a principal component analysis, followed by a further round of histogram construction, inspection and filtering. Each of the data subsets remaining after one of three termination criteria is met (Section 15.4) constitutes a cluster.

We have described how automated histogram filtering cluster analysis, used in conjunction with distance-geometry software, was able to generate, with remarkable efficiency, low-energy structures for a highly frustrated, off-lattice protein model [2]. We have also described how squared-loadings plots are generated and their use in identifying functionally related motions from clustered molecular dynamics conformational snapshots of an all-atom model of the Spo0F protein [22].

Varimax rotation of PCs is independent of data clustering, and is not a new technique. However, to our knowledge, squared-loadings plots of variable scores along rotated PCs are a new development. In research recently concluded in our laboratory, these have proven useful to understand local structural diversity of Boltzmann-distributed conformations of a protein model as it undergoes folding [23].

The methods reviewed here are just the first steps in the development of this emerging technology. Further research, some of which is in progress, is required to more firmly establish the strengths and limitations of automated histogram filtering cluster analysis. To facilitate use of our codes by others, copies suitable for parallel data processing are available from one of us (SMR) upon request.

15.8 ACKNOWLEDGEMENTS

This work was supported, in part, by a Cottrell College Science Award from the Research Corporation and by grants from the Natural Sciences and Engineering Research Council of Canada. Technical assistance was provided by Patricia Wang Pan. Results cited in this review were generated at high-performance computing facilities located at McMaster University (SHARCNET), The University of Alberta (WESTGRID), and Niagara College. Access to these resources is gratefully acknowledged.

15.9 REFERENCES

1 N. Elmaci and R.S. Berry, J. Chem. Phys., 110 (1999) 10606.
2 S.A. Larrass, L.M. Pegram, H.L. Gordon and S.M. Rothstein, J. Chem. Phys., 119 (2003) 13149.
3 L.L. Thurstone, Multiple-factor analysis: a development and expansion of the vectors of the mind, University of Chicago, Chicago, Illinois, 1965.
4 R.J. Rummel, Applied factor analysis, Northwestern University, Evanston, Illinois, 1970.
5 H.H. Harman, Modern factor analysis, University of Chicago, Chicago, Illinois, 1976.
6 S. Daultrey, Principal components analysis, in CATMOG. Concepts and techniques in modern geography, Vol. 8, University of East Anglia, Norwich, UK, Geo Abstracts, 1976.
7 http://www.netlib.org/scalapack
8 J. Imada, P. Chapman and S.M. Rothstein, in 19th International Symposium on High Performance Computing Systems and Applications, IEEE Computer Society, Los Alamitos, CA, 2005, p. 238.
9 J. Imada, private communication.
10 M.J. Norušis, SPSS professional statistics 6.1, SPSS Inc., Chicago, Illinois, 1994, pp. 83–109.
11 C.J. Tsai, S. Kumar, B. Ma and R. Nussinov, Protein Sci., 8 (1999) 1181.
12 Z. Guo and D. Thirumalai, Biopolymers, 36 (1995) 83.
13 S.E. Radford and C.A. Dobson, Phil. Trans. Royal Soc. Lond. Ser. B Biol. Sci., 348 (1995) 17.
14 J.M. Blaney, G.M. Crippen, A. Dearing, J.S. Dixon and D.C. Spellmeyer, Quantum chemistry program exchange, No. 590, Indiana University, Bloomington, Indiana, 1995.
15 V.A. Feher, J.W. Zapf, J.A. Hoch, J.M. Whiteley, L.P. McIntosh, M. Rance, N.J. Skelton, F.W. Dahlquist and J. Cavanagh, Biochem., 36 (1997) 10015.
16 V.A. Feher and J. Cavanagh, Nature, 400 (1999) 289.
17 A. Kitao and N. Go, Curr. Opin. Struct. Biol., 9 (1999) 164.
18 M.L. Teodoro and L.E. Kavraki, Curr. Pharm. Design, 9 (2003) 1635.
19 M.L. Thedoro, G.N. Phillips Jr and K.E. Kavraki, J. Comput. Biol., 10 (2003) 617.
20 A. Kitao, S. Hayward and N. Go, Proteins: Struct. Funct. Genet., 33 (1998) 496.
21 A. Amadei, A.B.B. Linssen, B.L. de Groot, D.M.F. van Aalten and H.J.C. Berendsen, J. Biomol. Struct. Dyn., 13 (1996) 615.
22 P.W. Pan, H.L. Gordon, R.J. Dickson, S.M. Rothstein and S. Tanaka, J. Chem. Phys., 122 (2005) 034904.
23 P.W. Pan, H.L. Gordon and S.M. Rothstein, J. Chem. Phys., 124 (2006) 024905.

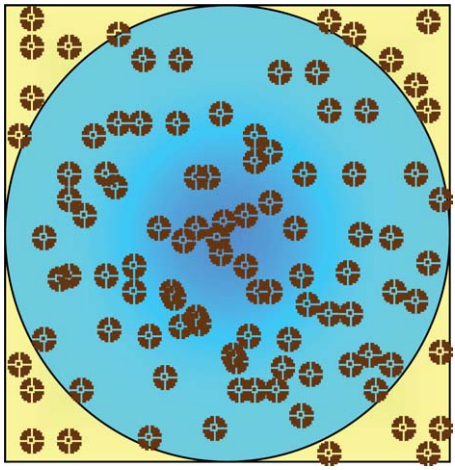


Plate 22. The simple dart board after throwing 100 darts (illustration appears on page 271 of this volume).

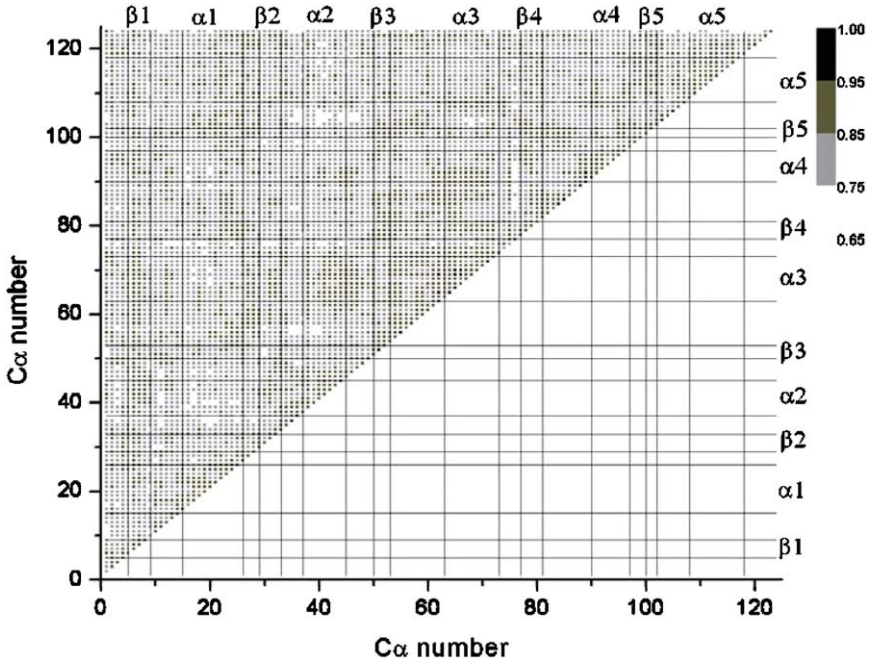


Plate 23. Cross-correlation plot of inter-C_α distances for the first cluster of molecular dynamics conformational snapshots of Spo0F performed in implicit solvent. The shading reflects the degree of correlation attributred to the individual variables. Reprinted from [22] with permission (illustration appears on page 315 of this volume).

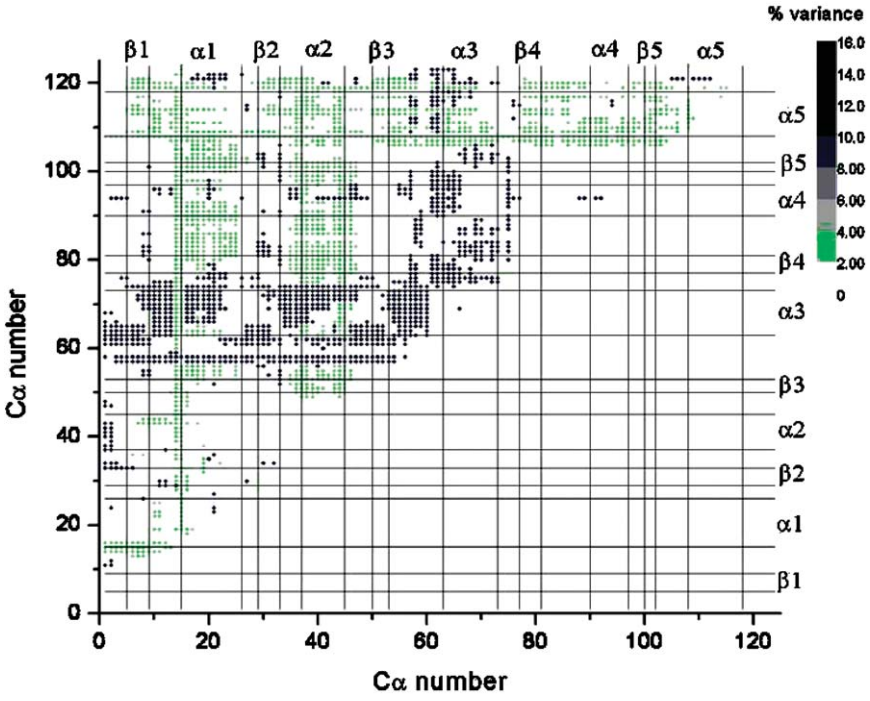


Plate 24. Squared-loadings plot for the first cluster of molecular dynamics conformational snapshots of Spo0F performed in implicit solvent. Varimax rotated squared-loadings greater than or equal to 0.25 within one of the first 100 principal components are shown. The shading reflects the variance described by the principal component that contained the maximum squared loading. Darker points imply that the principal component describes a larger percentage of the total sample variance. No points are plotted for inter-C_α distances whose maximum squared-loading value is either less than 0.25 in any of the first 100 principal components, or if its maximum squared-loading value belongs to a rotated principal component that accounts for less than 2 percent of the total variance. Elements of secondary structure are noted; intervening loops are not labelled, but are designated by horizontal and vertical lines separating adjacent helices and strands. Reprinted from [22] with permission (illustration appears on page 316 of this volume).

Plate 25. Overlay of the native(red) and folded (blue) structures of trp-cage protein [35], the HIV accessory protein [13] and the bacterial ribosomal protein L20 [14] (illustration appears on page 324 of this volume).

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 16

All-atom protein folding with stochastic optimization methods

A. Schug[1], A. Verma[2], K. H. Lee[3] and W. Wenzel[1]

[1]*Forschungszentrum Karlsruhe, Institute for Nanotechnology, P.O. Box 3640, D-76021 Karlsruhe, Germany*
[2]*Forschungszentrum Karlsruhe, Institute for Scientific Computing, P.O. Box 3640, D-76021 Karlsruhe, Germany*
[3]*Korean Institute for Science and Technology, Supercomputing Materials Laboratory, P.O.Box 131, Cheongryang, Seoul 130-650, Korea*

Abstract

We recently developed an all-atom free energy forcefield (PFF01) for protein structure prediction with stochastic optimization methods. We demonstrated that PFF01 correctly predicts the native conformation of several proteins as the global optimum of the free energy surface. Here we review recent folding studies, which permitted the reproducible all-atom folding of the 20 amino-acid trp-cage protein, the 40 amino-acid three-helix HIV accessory protein and the 60 amino-acid bacterial ribosomal protein L20 with a variety of stochastic optimization methods. These results demonstrate that all-atom protein folding can be achieved with present day computational resources for proteins of moderate size.

16.1 INTRODUCTION

Ab initio protein tertiary structure prediction (PSP) and the elucidation of the mechanism of the folding process are among the most important outstanding problems of biophysical chemistry [1,2]. The many complementary proposals for PSP span a wide range of representations of the protein conformation, ranging from coarse-grained models to atomic resolution. The choice of representation often correlates with the methodology employed in structure prediction, ranging from empirical potentials for coarse-grained models [3,4] to complex atom-based potentials that directly approximate the physical interactions in the system. The latter offer insights into the mechanism of protein structure formation and promise better transferability, but their use incurs large computational costs that has confined all-atom protein structure prediction to all but the smallest peptides [5,6].

It has been one of the central paradigms of protein folding that proteins in their native conformation are in thermodynamic equilibrium with their environment [7]. Exploiting this characteristic the structure of the protein can be predicted by locating the global minimum of its free energy surface without recourse to the folding dynamics, a process which is potentially much more efficient than the direct simulation of the folding process. PSP based on global optimization of the free energy may offer a viable alternative approach, provided that suitable parameterization of the free energy of the protein in its environment exists and that the global optimum of this free energy surface can be found with sufficient accuracy [8].

We have recently demonstrated a feasible strategy for all-atom protein structure prediction [9–11] in a minimal thermodynamic approach. We developed an all-atom free-energy force field for proteins (PFF01), which is primarily based on physical interactions with important empirical, though sequence independent, corrections [11]. We already demonstrated the reproducible and predictive folding of four proteins, the 20 amino-acid trp-cage protein (1L2Y) [9,12], the structurally conserved headpiece of the 40 amino-acid HIV accessory protein (1F4I) [10,13] and the 60 amino-acid bacterial ribosomal protein L20 [14]. In addition we showed that PFF01 stabilizes the native conformations of other proteins, e.g., the 52 amino-acid protein A [5,15], and the engrailed homeodomain (1ENH) from *Drosophilia melangaster* [16].

16.1.1 Force field

We have recently developed an all-atom (with the exception of apolar CH_n groups) free energy protein force field (PFF01) that models the low-energy conformations of proteins with minimal computational demand [10,11,17]. In the folding process at physiological conditions the degrees-of-freedom of a peptide are confined to rotations about single bonds. The force field is parameterized with the following non-bonded interactions:

$$V(\{\vec{r}_i\}) = \sum_{ij} V_{ij}\left[\left(\frac{R_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{ij} \frac{q_i q_j}{\varepsilon_{g(i)g(j)} r_{ij}} + \sum_i \sigma_i A_i + \sum_{\text{hbonds}} V_{hb} \tag{1}$$

Here r_{ij} denotes the distance between atoms i and j and $g(i)$ the type of the amino acid i. The Lennard Jones parameters ($V_{ij}\, R_{ij}$ for potential depths and equilibrium distance) depend on the type of atom pair and were adjusted to satisfy constraints derived from a set of 138 proteins of the PDB database [17–19]. The non-trivial electrostatic interactions in proteins are represented via group-specific dielectric constants ($\varepsilon_{g(i)\, g(j)}$ depending on the amino-acid to which atom i belongs). The partial charges q_i and the dielectric constants were derived in a potential-of-mean-force approach [20]. Interactions with the solvent were first fit in a minimal solvent accessible surface model [21] parameterized by free energies per unit area σ_i to reproduce the enthalpies of solvation of the Gly-X-Gly family of peptides [22]. A_i corresponds to the area of atom i that is in contact with a ficticious solvent. Hydrogen bonds

are described via dipole–dipole interactions included in the electrostatic terms and an additional short-range term for backbone–backbone hydrogen bonding (CO to NH) which depends on the OH distance, the angle between N, H and O along the bond and the angle between the CO and NH axis [11].

16.1.2 Optimization methods

The low energy region of the free energy landscape of proteins is extremely rugged due to the comparatively close packing of the atoms in the native structure. Suitable optimization methods must therefore be able to speed up the simulation by avoiding high-energy transition states, adapt large-scale moves or accept unphysical intermediates. Here we report on four different optimization methods, the stochastic tunneling method [23], the basin hopping technique [24,25], the parallel tempering method [26,27] and a recently employed evolutionary technique. The stochastic tunneling method and the basin hopping approach are inherently sequential algorithms, which evolve a single configuration according to a given stochastic process. In contrast, parallel tempering and evolutionary techniques are inherently parallel optimization strategies that are well suited to presently available multiprocessor architectures with low bandwidth connections. Since all-atom protein structure prediction remains a computationally challenging problem it is important to search for suitable optimization methods that are capable of exploiting such architectures, i.e., a high degree of parallelism with very little and optimally asynchronous communication is desirable.

16.1.2.1 Stochastic tunneling method

The stochastic tunneling technique (STUN) [23] was proposed as a generic global optimization method for complex rugged potential energy surfaces (PES). For a number of problems, including the prediction of receptor–ligand complexes for drug development [28,29], this technique proved superior to competing stochastic optimization methods. The idea behind the method is to flatten the potential energy surface in all regions that lie significantly above the best estimate for the minimal energy (E_0). In STUN the dynamic process explores not the original, but a transformed PES,

$$E_{\mathrm{STUN}} = 1\mathrm{n}\left(x + \sqrt{x^2+1}\right) \tag{2}$$

which dynamically adapts and simplifies during the simulation. Here $x = \gamma(E - E_0)$, where E is the energy, and E_0 the best energy found so far. The problem-dependent transformation parameter [23] γ controls the steepness of the transformation. (we used $\gamma = 0.5$ kcal/mol) The transformation in Eq. (2) ameliorates the difficulties associated with the original transformation [23], because $E_{\mathrm{STUN}} \propto 1\mathrm{n}(E/kT)$ continues to grow slowly for large energies. The ficticious temperature of STUN must be dynamically adjusted in order to accelerate convergence. STUN works best if its dynamical process alternates between low-temperature 'local search' and high-temperature 'tunneling phases'. At finite temperature the dynamics of the system then becomes diffusive at energies $E >> E_0$ independent of the relative energy differences of the high-energy

conformations involved. On the untransformed PES, STUN thus permits the simulation to 'tunnel' through energy barriers of arbitrary height.

16.1.2.2 Parallel tempering

The parallel (or simulated) tempering technique (PT) [26,27] was introduced to overcome difficulties in the evaluation of thermodynamic observables for models with very rugged potential energy surfaces and has been applied previously in several protein folding studies [30–32]. Low-temperature simulations on rugged potential energy surfaces are trapped for long times in similar metastable conformations because the energy barriers to structurally potentially competing different conformations are very high. The idea of PT is to perform several concurrent simulations of different replicas of the same system at different temperatures and to exchange replicas (or temperatures) between the simulations i and j with probability:

$$p = \min\left(1, \exp\left(-\left(\beta_j - \beta_i\right)\left(Ei - E_j\right)\right)\right) \tag{3}$$

where $\beta_i = 1/k_\beta T_i$ and E_i are the inverse temperatures and energies of the conformations, respectively. The temperature scale for the highest and lowest temperatures is determined by the requirement to efficiently explore the conformational space and to accurately resolve local minima, respectively. For proteins the temperatures must thus fall in a bracket between approximately 2–600 K. As described elsewhere [12] we have used an *adaptive temperature control* for the simulations: starting with an initial, ordered set of geometrically distributed temperatures we monitored the exchange rate between adjacent temperatures. If the exchange rate between temperature i and $i+1$ was below 0.5 per cent, then all temperatures above t_i were lowered by 10 per cent of $t_{i+1} - t_i$. If the exchange rate was above 2 per cent, then all all temperatures above t_i were increased by the same difference.

To further improve the computational efficiency of PT we also used a *replication step*, in which the best conformation replaces the conformation at the highest temperature every 250 000 simulation steps. This mechanism results in a rapid, large-scale exploration of the folding funnel around the best conformation found near the presently best conformation. The parallel tempering method was implemented in our program using the MPI communication library, which is available on most present day parallel computational architechtures with distributed memory. Since the communication effort is low (only the temperatures and energies need to be exchanged) and communication occurs only every few thousand steps, when replica exchange is attempted, this implementation scales very well with the number of processors.

16.1.2.3 Basin hopping technique

An alternative approach to effectively eliminate high-energy transition states of the PES is used in the the basin hopping technique [24] (BHT), also known as Monte Carlo with minimization. This method simplifies the original potential energy surface by replacing the energy of each conformation with the energy of a nearby local minimum. This replacement eliminates high-energy barriers in the stochastic search that are responsible for the freezing problem in simulated annealing. In many cases the additional minimization

effort to find an associated local minimum is more than compensated for by the increase of efficiency of the stochastic search on the simplified potential energy surface.

For the protein simulations we replace a single minimization step with a simulated annealing run [33]. Within each SA simulation, new configurations are accepted according to the Metropolis criterion. The temperature is decreased geometrically from its starting point to the final value, which must be chosen to be small compared with typical energy differences between competing metastable conformations, to ensure convergence to a local minimum (typically 2–5 K). Depending on the choice of the starting temperature, the SA search can deviate more or less significantly from its starting conformation. The individual relaxation step is thus parameterized completely by the starting temperature, the final temperature and the number of steps. We investigate various choices for the numerical parameters of the method, but have always used a geometric cooling schedule.

At the end of one annealing step the new conformation was accepted if its energy difference to the current configuration was no higher than a given threshold energy ε_T, an approach recently proven optimal for certain optimization problems [34]. Throughout this study we use a threshold acceptance criterion of 1 kcal/mol.

16.1.2.4 Evolutionary strategies

While basin hopping and STUN are essentially sequential algorithms, PT provides some degree of inherent parallelism, which suits present day distributed architectures. In order to use even larger numbers of processors, we implemented an evolutionary strategy, in which the computational work is performed by many independent client computers that request tasks from a master computer. The master maintains a list of open tasks comprising the active conformations of the population. Each client performs an increasingly extensive energy minimization on the conformation it is given. When the client returns a new conformation after completing its task, it may replace an existing conformation following a scheme that balances the diversity of the population and the continued energetic improvement of its members. Specifically the client performs either a Monte Carlo (MC) or a simulated annealing (SA) [33] simulation of specified length on the conformation. Conformations are drawn randomly from the active population. When the client returns a new conformation after completing its task, the result is stored. Additionally, the new conformation replaces the energetically worst conformation in the active population, provided its energy is lower than the highest energy of the population and that it differs by at least 3Å backbone root mean square deviation (RMSB) from all members of the active population. If the new conformation has an RMSB of less than 3Å to some conformation of the population, it replaces this conformation if it is lower in energy.

This approach builds on the strength of the basin hopping technique, which is used for the individual steps. Since stochastic optimization methods never locate the global optimum with certainty, several independent simulations must be undertaken to obtain a relative degree of confidence in the convergence of the methods. In the evolutionary approach the progress a single simulation has made towards the optimum is exploited, because the resulting conformation will become a member of the active population. As such it may replace less optimized conformations and speed up the convergence of the overall population in comparison to a population of uncorrelated replicas.

16.2 RESULTS

16.2.1 The trp-cage protein

Using the PFF01 force field we simulated 20 independent replicas of the 20 amino-acid trp-cage protein [6,36] (pdb code 1L2Y) with a modified version of the stochastic tunneling method [9,23]. Six of 25 simulations reached an energy within 1 kcal/mol of the best energy, all of which correctly predicted the native experimental structure of the protein (see Fig. 16.1 (left)). We found a strong correlation between energy and RMSB deviation to the native structure for all simulations. The conformation with the lowest energy had a backbone root mean square deviation of 2.83 Å.

We also folded this protein with the parallel tempering method [12]. We found that the standard approach, which preserves the thermodynamic equilibrium of the simulated populations, did not reach very low energies even for the low-temperature replicas. We believe that the reason for this convergence failure was the insufficient exchange probability between replicas at different temperatures. We therefore introduced the adaptive temperature control described in the methods section. Fig. 16.2) shows the energies and corresponding temperatures for a simulation using 30 replicas. The temperature adjustment scheme results in a temperature distribution that permits frequent exchange of replicas and significantly speeds convergence. The best final structure associated with the lowest temperature in the simulation with 30 replicas had an RMSB deviation of 2.01 Å. We found convergence of the method using eight to thirty replicas. However, a minimal number of at least eight replicas appears to be required to fold the protein, for lower replica numbers it appears that even the adaptive temperature scheme fails to generate rapid replica exchange while spanning both high and low temperatures required for the speedy exploration of the free energy surface and the refinement of local minima, respectively.

Finally we have folded the trp-cage protein with the basin hopping technique. In comparison with the stochastic tunneling method we note that care must be taken with the parameterization of the basin hopping technique. First of all, very high starting temperatures, above 600 K, are required to permit a sufficient exploration of the free energy surface. This was also observed in the parallel tempering simulations (see Fig. 16.2) bottom

Fig. 16.1. Overlay of the native(red) and folded (blue) structures of trp-cage protein [35], the HIV accessory protein [13] and the bacterial ribosomal protein L20 [14] (see Color Plate 25).

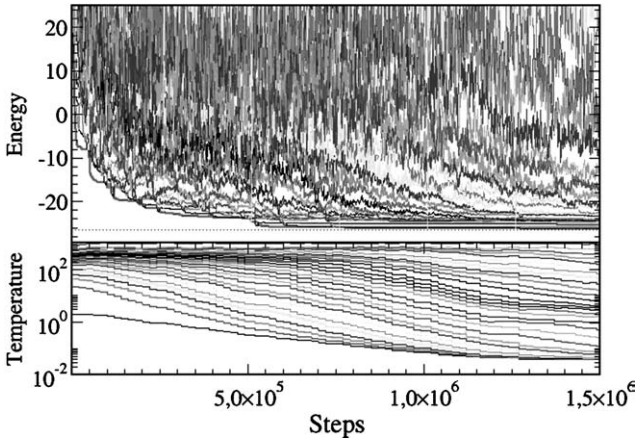


Fig. 16.2. Energies (upper panel) and temperatures (lower panel) of the 30 replica modified parallel tempering simulation of the trp-cage protein reported in the text. The dotted line in the upper panel corresponds to the estimate of the global optimum of the free energy (obtained independently). The lower panel demonstrates a rapid equilibration of the temperatures during the simulation. The upper panel demonstrates the convergence of the energy and the rapid exchange of information between the different replicas as discussed in the text (see Color Plate 26).

panel). As in PT, the lowest temperature had to be chosen in the range of 2–6 K to ensure that local minima are resolved well. In the stochastic tunneling method, the nonlinear transformation of the energy permits large-scale relaxations through thermodynamically forbidden regions, in basin hopping this effect can only be achieved by raising the temperature to unphysical values. We cannot rule out the possibility that basin hopping simulations with low starting temperatures would converge eventually, however, it appears that such an approach would not be computationally competitive. Furthermore, we note that the convergence of the basin hopping method is improved dramatically when the length of the relaxation run is moderately increased with the number of the basin hopping cycle. When we compared simulations comprised of basin hopping steps of constant length mid simulations where the length increased with the square root of the cycle number, we found much lower energies for the latter after investing the same total number of function evaluations in each run.

Using the basin hopping method with a starting temperature of $T_s = 800\ K$ and a final temperature of $T_f = 3\ K$ the lowest six of 20 simulations converged to the native structure. A total of 12 of these simulations approached the native conformation as its estimate of the optimum. The energies and RMSB deviations of all simulations are shown in Fig. 16.3. The plot indicates the existence of a set of structures with 2–3 Å RMSB deviation, which may correspond to the folding funnel, and a competing metastable conformation with about 5 Å RMSB. While all methods correctly identify the folding funnel, the basin hopping approach results in the lowest energies. Note that the second best simulation has an RMSB of only 1.8Å to the native conformation and loses in energy with less than 0.5 kcal/mol.

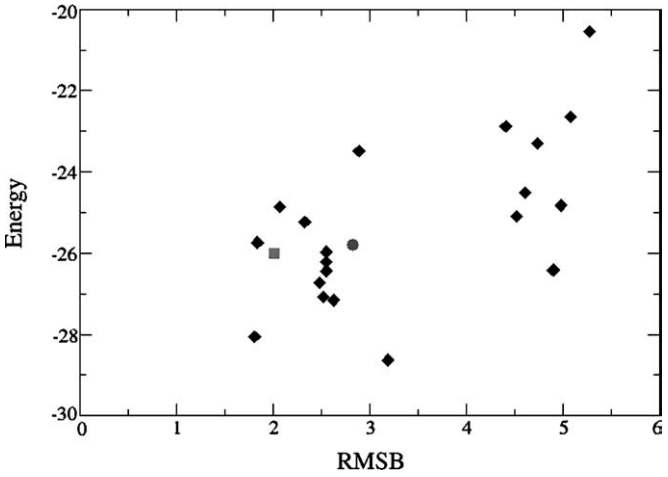


Fig. 16.3. Energy vs. RMSB plot for the final energies of the twenty basin hopping simulations described in the text (diamonds). For comparison we also indicate the best energy result for the STUN method (circle) and for the thirty processor PT simulation (square) (see Color Plate 27).

16.2.2 The HIV accessory protein

Encouraged by this result, we applied the modified basin hopping or Monte Carlo with minimization (MCM) strategy [8,25] to fold the structurally conserved 40 amino-acid headpiece of the HIV accessory protein [10]. We performed 20 independent simulations and found the lowest five to converge to the native structure [14]. The first non-native decoy appears in position six, with an energy deviation of 5 kcal/mol and a significant RMSB deviation. All low-energy structures have essentially the same secondary structure, i.e., position and length of the helices are always correctly predicted, even if the protein did not fold correctly.

The good agreement between the folded and the experimental structure is also evident from Fig. 16.1 (center), which shows the secondary structure alignment of the native and the folded conformations. The good physical alignment of the helices illustrates the importance of hydrophobic contacts to correctly fold this protein. An independent measure to assess the quality of these contacts is to compare the C_β–C_β (which correspond to the NOE constraints of the NMR experiments that determine tertiary structure) in the folded structure to those of the native structure. We found that 66 per cent (80 per cent) of the C_β–C_β distances agree to within one (1.5) standard deviations of the experimental resolution.

We also performed a simulation of the HIV accessory protein using the adapted parallel tempering method [13]. We used 20 processors of an INTEL XEON PC cluster and ran the simulation for a total of 30×10^6 energy evaluations for each configuration, which corresponds to approximately 500 CPU hours on an 2.4 GHz INTEL XEON processor. All simulations were started with random conformations at high temperatures to allow for rapid, unbiased relaxation of the structures and the temperature distribution. The final

conformation with the lowest energy/temperature converged to within 1.23/2.46 Å backbone root mean square (RMSB) deviation to the best known decoy/NMR structure of the HIV accessory protein. The overlay of the experimental and the converged structure (see Fig. 16.1) demonstrates the good agreement between the conformations, the difference in NOE constraints demonstrates that not only short-range, but also long-range distances are correctly predicted. Considering the ensemble of final conformations, we find many structures closely resembling the native conformation. The RMSB deviations of the next four lowest conformations (all within 1.5 kcal/mol of the minimal energy) have RMSB deviations of 3.14/2.23/3.78/3.00 Å, respectively, to the native decoy.

16.2.3 The bacterial ribosomal protein L20

In the course of the simulations on the HIV accessory protein we explored methods to share information between the independent basin hopping simulations in order to improve the overall convergence. For the 60 amino-acid bacterial ribosomal protein L20 (pdb-code 1GYZ) we thus experimented with the evolutionary technique described in the methods section. Starting from a seed population of random structures, we performed the folding simulation in three phases: (1) generation of starting structures of the population, (2) evolutionary improvement of the population, and (3) refinement of the best resulting structures to ensure convergence.

In phase (1) we performed high-temperature (500K) Monte Carlo simulations of 50 000 steps each. In these runs we reduced the strength of the solvent interactions (V_S) by 20 per cent to facilitate the rapid formation of secondary structures. It has been argued that hydrophobic collapse competes with secondary structure formation in protein folding. In the collapsed conformational ensemble, large-scale conformational changes, such as those required for secondary structure formation, occur only rarely. The goal of this simulation phase was the generation of a wide variety of competitive starting conformations for further refinement.

At the end of this simulation we had gathered in excess of 17 000 distinct decoys that were ranked according to their total energy as well as according to the individual energy terms of the force field (V_S, V_{LJ}, V_{HB}, V_C(Sidechain) and V_C(backbone). For each criterion we selected the best 50 conformations and eliminated duplicates to arrive at a population of 266 starting structures for phase (2) of the procedure. This population was relaxed in 14 000 simulated annealing (SA) simulations as described in the methods section. At the end of this step we selected the 50 conformations best in total energy for further refinement. In phase (3) we performed 5500 SA simulations on this subpopulation. The length of the individual relaxation simulations was gradually increased from 10^5 steps per simulation to 2.3×10^6.

The best conformation had approached the native conformation to about 4.6 Å RMSB deviation. In total, six of the lowest ten conformations approach the native structure, while four others misfolded. Note that the selection criterion for the active population (see methods section) precludes the occurrence of the same configuration to within 3 Å RMSB; this dominance of near native conformations of the total ensemble is particularly encouraging.

In order to quantify the overall improvement of native content during the simulation, we defined the native content of the simulated ensemble as a weighted average of the

deviations of the population and the native conformation: for a population of size N we add $100(N-R+1)/N$ for each near-native decoy (RMSB less than 4Å) ranked at position R by energy to the total native score of this population. A score of 100 thus corresponds to a native decoy placed at the top position, while a near-native decoy at the very bottom contributes just unity. Non-native conformations contribute nothing. Using this measure the final population contains in excess of 20 per cent of near-native conformations, its native score exceeds 800, increasing 60-fold during simulation phases (2) and (3).

16.3 CONCLUSION

Since the native structure dominates the low-energy conformations arising in all of these simulations, our results demonstrate the feasibility of all-atom protein tertiary structure prediction for three different proteins ranging from 20–60 amino acids in length with a variety of different optimization methods. The free energy approach thus emerges as a viable trade-off between predictivity and computational feasibility. While sacrificing the folding dynamics, a reliable prediction of its terminus, the native conformation – which is central to most biological questions – can be achieved.

The computational advantage of the optimization approach stems from the possibility of visiting unphysical intermediate conformations with high energy during the search. This goal is realized with different mechanisms in all of the employed stochastic optimization methods. In the stochastic tunneling method, the nonlinear transformation of the PES permits the dynamical process to traverse abritrarily high energy barriers at low termperatures; in basin hopping and parallel tempering, simulation phases at very high temperatures accomplish the same objective.

Our results indicate that the simple basin hopping method is very efficient in the determination of the global optimum of the free energy surface of realistic all-atom protein models. It is encouraging that the same structure was also found for the trp-cage protein using the parallel tempering and the stochastic tunneling method and also for the HIV accessory protein using the parallel tempering method. This finding indicates that the result of the folding approach is not an artefact of the optimization strategy. In direct comparison, however, we find that the basin hopping technique gave the lowest energies. Since it is virtually parameter-free, and very simple to implement, it emerges as a natural work-horse for our approach.

Its one important disadvantage is the fact that different basin hopping simulations are completely independent of one another. Because the underlying optimization problem in all-atom protein folding is very difficult, and the free energy surface is very rugged, several simulations must be undertaken to obtain a relative degree of confidence in the convergence of the approach. From this perspective it is desirable to design an optimization method in which different members of the simulated population can learn from one another. In the evolutionary approach we have explored, here is one particularly simple scheme to ensure that the total computational effort is concentrated on the best conformations that arise at intermediate stages of the simulations. So far it leads to the folding of the largest protein at all-atom resolution, but much work remains to be done to optimize this approach.

16.4 DISCUSSION

This review indicates that all-atom protein structure prediction with stochastic optimization methods becomes feasible with present day computational resources. The fact that three proteins were reproducibly folded with different optimization methods to near-native conformation increases the confidence in the parameterization of our all-atom protein force field PFF01. The presently available evidence indicates that the comparatively straightforward basin hopping routine is a good work-horse to evolve individual conformations. The resolution of several independent basin hopping simulations may be enhanced by the use of evolutionary algorithms, such as the one used for the bacterial ribosomal protein L20 . We note that the master–client model for this strategy is asynchronous and can be implemented outside the simulation program using the standard TCP/IP, FTP or HTTP protocols. Using generic libraries, such as MPI, it can also be easily implemented on MPP architectures. Since the amount of information that needs to be exchanged is very small, while the effort in a single basin hopping cycle is substantial, there is virtually no loss in an asynchronous parallel implementation. This makes the evolutionary approach investigated here suitable for present day GRID architectures. While the present results demonstrate proof of principle, much work remains to be done to arrive at an optimal strategy.

Protein structure prediction with stochastic optimization methods requires two separate key ingredients: an accurate force field and efficient optimization techniques. One cannot overemphasize the importance of the interplay of optimization methods and force field validation. Rational force field development mandates the ability to generate decoys that fully explore competing low-energy conformations to the native state. The success of different optimization strategies depends strongly on the structure of the potential energy surface. As a result, the development of efficient optimization techniques for an all-atom protein structure prediction depends on the availability of a force field that stabilizes native conformations of proteins with appreciable hydrophobic cores. For helical proteins the bottleneck in *ab initio* all-atom structure prediction now lies in the development of optimization strategies that significantly increase the system size and increase the reliability of the predictions. Based on the results reviewed here, it is sensible to investigate improvements of the evolutionary techniques based on the basin hopping method for this purpose.

The application of this methodology to a wide range of proteins will generate large decoy sets of metastable conformations that compete with the native structure of the protein. These decoy sets may in turn be used to improve the parameterization of the force field. By its very nature, the approximation of the free energy of the system mandates the use of implicit solvent models. This implies that interactions with the solvent and intramolecular electrostatic interactions must be parameterized in accurate, yet efficient effective models. Since both effects are highly nontrivial, the free-energy approach can only approximate, but not duplicate, the results of all-atom explicit water simulations. The present evidence indicates that the native conformation is reproduced to 3–4 Å resolution with PFF01, but the results of the free-energy approach could be refined in all-atom simulations that start from a set of low-energy decoys.

16.5 ACKNOWLEDGEMENTS

We acknowledge the BMBF, the Deutsche Forschungsgemeinschaft (grants WE 1863/10-2, WE 1863/14-1) and the Kurt Eberhard Bode Stiftung for financial support. Parts of the simulations were performed at the KIST teraflop cluster.

16.6 REFERENCES

1 D. Baker and A. Sali, Science, 294 (2001) 93.
2 J. Schonbrunn, W.J. Wedemeyer and D. Baker, Curr. Op. Struc. Biol., 12 (2002) 348.
3 N. Go and H.A. Scheraga, Macromolecules, 9 (1976) 535.
4 P. Ulrich, W. Scott, W.W.F. van Gunsteren and A.E. Torda, Proteins, SF&G, 27 (1997) 367.
5 C.D. Snow, H. Nguyen, V.S. Pande and M. Gruebele, Nature, 420 (2002) 102.
6 C. Simmerling, B. Strockbine and A. Roitberg, J. Am. Chem. Soc., 124 (2002) 11258.
7 C.B. Anfinsen, Science, 181 (1973) 223.
8 Z. Li and H. Scheraga, Proc. Nat. Acad. Sci. USA, 84 (1987) 6611.
9 A. Schug, T. Herges and W. Wenzel, Phys. Rev. Lett., 91 (2003) 158102.
10 T. Herges and W. Wenzel, Phys. Rev. Lett., 94 (2004) 018101.
11 T. Herges and W. Wenzel, Biophys. J., 87 (5) (2004) 3100.
12 A. Schug, T. Herges and W. Wenzel, Europhyics Lett., 67 (2004) 307.
13 A. Schug, T. Herges and W. Wenzel, Proteins, 57 (2004) 792.
14 A. Schug, T. Herges and W. Wenzel, J. Am. Chem. Soc., 126 (2004) 16736.
15 H. Gouda, H. Torigoe, A. Saito, M. Sato, Y. Arata and I. Shimanda, Biochemistry, 40 (1992) 9665.
16 U. Mayor, N.R. Guydosh, C.M. Johnson, J.G. Grossmann, S. Sato, G.S. Jas, S.M.V. Freund, D.O.V. Alonso, V. Daggett and A.R. Fersht, Nature, 421 (2003) 863.
17 T. Herges, H. Merlitz and W. Wenzel, J. Ass. Lab. Autom., 7 (2002) 98.
18 R. Abagyan and M. Totrov, J. Molec. Biol., 235 (1994) 983.
19 T. Herges, A. Schug, B. Burghardt and W. Wenzel, Intl. J. Quant. Chem., 99 (2004) 854.
20 F. Avbelj and J. Moult, Biochemistry, 34 (1995) 755.
21 D. Eisenberg and A.D. McLachlan, Nature, 319 (1986) 199.
22 K.A. Sharp, A. Nicholls, R. Friedman and B. Honig, Biochemistry, 30 (1991) 9686.
23 W. Wenzel and K. Hamacher, Phys. Rev. Lett., 82 (1999) 3003.
24 A. Nayeem, J. Vila and H. Scheraga, J. Comp. Chem., 12 (5) (1991) 594.
25 J.P. Doye and D. Wales, J. Chem. Phys., 105 (1996) 8428.
26 G.J. Geyer, Stat. Sci., 7 (1992) 437.
27 K. Hukushima and K. Nemoto, J. Phys. Soci. Japan, 65 (1996) 1604.
28 H. Merlitz and W. Wenzel, Chem. Phys. Lett., 362 (2002) 271.
29 H. Merlitz, B. Burghardt and W. Wenzel, Chem. Phys. Lett., 370 (2003) 68.
30 U. Hansmann and Y. Okamoto, J. Comput. Chem., 18 (1997) 920.
31 U. Hansmann, Eur. Phys. J. B, 12 (1999) 607.
32 C. Lin, C. Hu and U. Hansmann, Proteins, 53 (2003) 436.
33 S. Kirkpatrick, C. Gelatt and M. Vecchi, Science, 220 (1983) 671.
34 J. Schneider, I. Morgenstern and J. Singer, Phys. Rev. E, 58 (1998) 5085.
35 A. Verma, A. Schug, K.H. Lee and W. Wenzel, Basin hopping simulations for all-atom protein folding, J. Chem. Phys., 124 (2006) 044515.
36 J.W. Neidigh, R.M. Fesinmeyer and N.H. Anderson, Nature Struct. Biol., 9 (2002) 425.

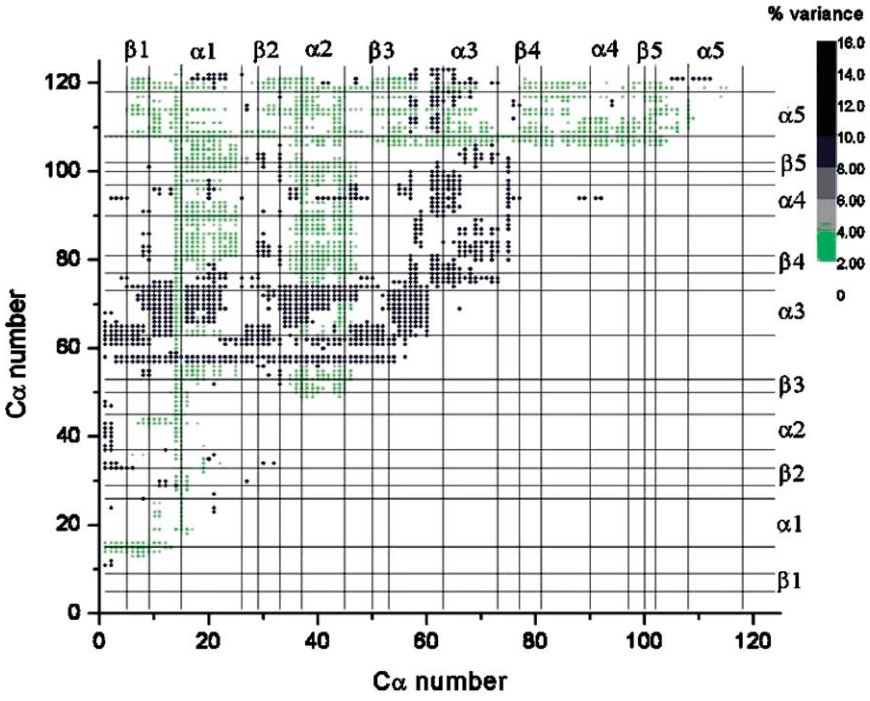


Plate 24. Squared-loadings plot for the first cluster of molecular dynamics conformational snapshots of Spo0F performed in implicit solvent. Varimax rotated squared-loadings greater than or equal to 0.25 within one of the first 100 principal components are shown. The shading reflects the variance described by the principal component that contained the maximum squared loading. Darker points imply that the principal component describes a larger percentage of the total sample variance. No points are plotted for inter-C_α distances whose maximum squared-loading value is either less than 0.25 in any of the first 100 principal components, or if its maximum squared-loading value belongs to a rotated principal component that accounts for less than 2 percent of the total variance. Elements of secondary structure are noted; intervening loops are not labelled, but are designated by horizontal and vertical lines separating adjacent helices and strands. Reprinted from [22] with permission (illustration appears on page 316 of this volume).

Plate 25. Overlay of the native(red) and folded (blue) structures of trp-cage protein [35], the HIV accessory protein [13] and the bacterial ribosomal protein L20 [14] (illustration appears on page 324 of this volume).

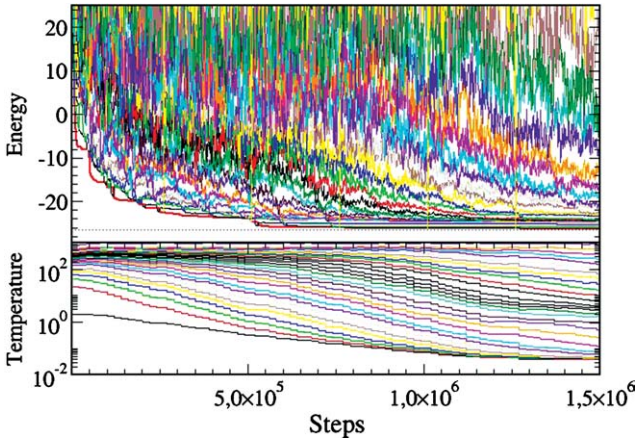


Plate 26. Energies (upper panel) and temperatures (lower panel) of the 30 replica modified parallel tempering simulation of the trp-cage protein reported in the text. The dotted line in the upper panel corresponds to the estimate of the global optimum of the free energy (obtained independently). The lower panel demonstrates a rapid equilibration of the temperatures during the simulation. The upper panel demonstrates the convergence of the energy and the rapid exchange of information between the different replicas as discussed in the text (illustration appears on page 325 of this volume).

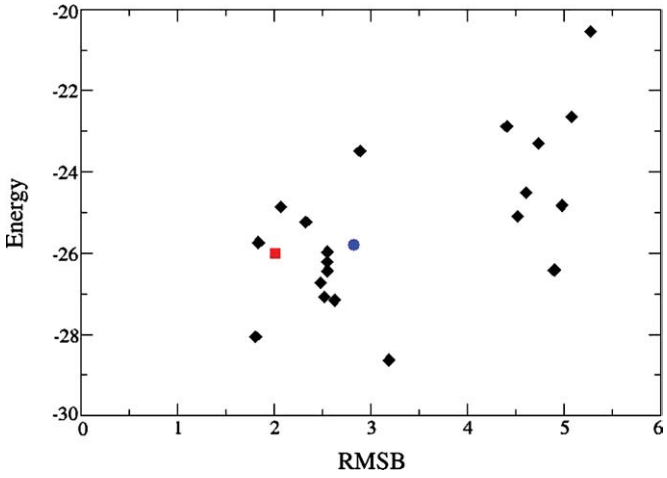


Plate 27. Energy vs. RMSB plot for the final energies of the twenty basin hopping simulations described in the text (diamonds). For comparison we also indicate the best energy result for the STUN method (circle) and for the thirty processor PT simulation (square) (illustration appears on page 326 of this volume).

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

Chapter 17

Simple models for nonlinear states of double-helix DNA

Raimo A. Lohikoski[1] and Aatto Laaksonen[2]

[1]*University of Jyväskylä, Department of Physics, FIN - 40014 University of Jyväskylä, Finland*
[2]*Division of Physical Chemistry, Arrhenius Laboratory, Stockholm University, S 106 91, Stockholm, Sweden*

Abstract

Review type introduction is given to the simple modeling of DNA. Intrinsically simple modeling aims at understanding or explaining of some "core" phenomena, not giving any all-embracing model of the underlying system. As a consequence the amount of this type of models and their versions is large. Here we have restricted our contemplation to the most important lines in the path of theoretical understanding of DNA melting or denaturation which is one of the important phases occurring during DNAs replication and transcription processes. The model "line" initiated by Peyrard and Bishop shows the richness these simple models can have.

17.1 INTRODUCTION

This chapter contains a short evolutionary survey of simple models used for DNA to understand the phases and aspects of the denaturation processes related to the transcription and replication of this fundamentally so important molecule of life. We begin with an introductory review covering the early work in developing these models. Later we will take a look at the current state in this branch of modeling. The word simple refers to an ultimate or near ultimate simplicity of the model itself. This does not necessarily mean that further mathematical or even numerical processing would be simple.

A typical simple model contains only one or a few parameters, unlike, e.g., the classical molecular dynamics simulation method, in which all atoms can be made movable and the number of degrees-of-freedom is therefore orders of magnitude higher. Basic motivations to use simplified, simple models are either the purpose to find the most relevant factor or factors affecting the object system, or an ambition to get theoretical explanations for the system properties or behavior even with quite 'inadequate' methods and, maybe, when more accurate methods are too laborious to work with.

Simple models for dsDNA are traditionally created by physicists whose aim seems to be mainly to understand dsDNA's intrinsic behaviors in the biological events where dsDNA has been observed to participate. These simple models possess nonlinearity as their key element. This is a clear consequence of the observation revealing that biomolecules, like DNA, have large amplitude nonregular motions not possible to describe with basic harmonic interactions. These models typically have their origin in mechanics and largely in the basic principles of classical physics. Interaction potentials are created, and the chemistry is hidden in the parameters of the potential functions, though the functional forms are also important in this respect. Specifically, the transcription, replication and their related processes have been in focus, obviously, because the experimentally observed data are mostly available for these phenomena.

Currently, the conduction properties of dsDNA and dsDNA together with other molecules in molecular or matter environments form a very active research area where there also exists a collection of simple models. However, this field is different because quantum properties are needed to make the models have different characters. This indicates the usual feature of simplified models: models are often quite phenomenon-limited and this is true not only for DNA but also for many other molecules and events.

This chapter gives a short chronological representation of the simple models related or still relating to the general topic of dsDNA opening. This is by no means a perfect view of all the models found in the literature, but more an introduction to the subject and concentration on another 'line of work' proving to be very fruitful both in its qualitative ability for dsDNA as well as in its alluringness among researchers. For those readers who wish to have a detailed introduction to this area we refer to comprehensive monographs by Yakushevich [1], and Dauxois and Peyrard [2].

17.2 ENGLANDER MODEL – BASE TWIST AROUND THE BACKBONE

The history of simple nonlinear DNA models goes back to the year 1980 and the hydrogen–deuterium exchange measurements of dsDNA molecules in an equilibrium solution performed by Englander and co-workers [3]. These experiments showed a dynamical existence of open states in the helical structure of dsDNA, where the open states were moving from site to site. The open state was a base pair or a sequence of adjacent base pairs where the hydrogen bonding in the base pair is broken or at least stretched so much that the canonical dsDNA structure could not exist anymore.

Owing to lack of direct information about the conformation, it was assumed that several adjacent base pairs could be simultaneously open. Further it was assumed that the open segment existed as a mobile unit free to move along the double helix. Based on the experimental data a guesstimate was made that the average length of the opened regime was roughly ten base pairs, leading to a proposition of thermal soliton excitations in the dsDNA. A soliton is a 'packet' which is usually related to a vibrational motion. It is formed as a combination of both linear and nonlinear properties of a system, and has the characteristic property of being radiationless or compact coherent entities over long periods of time or large distances. By definition, a slightly milder form of a soliton is a solitary wave which is allowed to deform slowly.

In a biological context, where large amplitude motions are typical, and consequently nonlinear interactions effective, solitons or solitary excitations were considered as a realistic possibility. One of the first proposed models for the open conformation of dsDNA is shown in Fig. 17.1. It consists of two parallel rods having pendula (bases) pointing towards each other in a minimum energy configuration. The assumed excitation was thought to be a twist or rotation of the bases around one of the two backbone strands. Physically this model is a pure mechanical device. Each base is a mass point attached via a rod to the related backbone. The backbone itself can be described by a spring resisting the twist. Base–base stacking interactions in the helix direction and a limited torsional movement around the P–O bonds are described by a (stiff) torsional elasticity constant K. Short-range attractive interactions come from base-pair hydrogen bonds and solvation. Owing to a mathematical convenience this potential was represented in a gravitational form $V(\phi_n) = mgr(1-\cos\phi_n)$, where ϕ_n is a deviation of the nth rod from its equilibrium angle position. The Hamiltonian for this first DNA model was given as:

$$H = \sum_n \left\{ \frac{1}{2} mr^2 \dot{\phi}_n^2 + \frac{1}{2} K \left(\phi_n - \phi_{n-1} \right)^2 + mgr \left(1 - \cos\phi_n \right) \right\} \tag{1}$$

This Hamiltonian can be treated following the normal procedure of mechanics [4] by deriving the equations of motion and solving them. However, in attempts to solve the equations these early and simple modeling works relied on a fairly drastic approximation that the motion from one base pair to the next was assumed to be slow enough to allow the use of the continuum limit approximation. This then made it possible to obtain analytical mathematical solutions, often even with other simplifying pre-suppositions.

The Englander model is another case of the sine-Gordon model, which is one of the well-known nonlinear equations producing soliton solutions. Because this is a very crude model in describing the actual dsDNA molecule in a given state we will

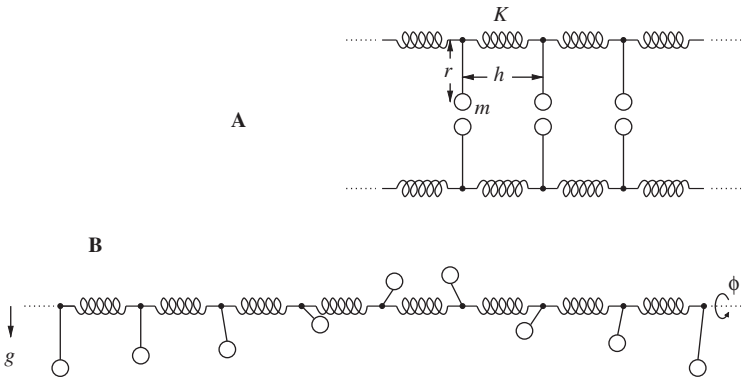


Fig. 17.1. The Englander model. (**A**) The equilibrium state of a dsDNA. The sugar–phosphate backbones are shown by two horizontal springs having pendula to describe the bases. (**B**) A single strand, capable or torsional oscillations is shown with an idealistic soliton excitation.

not go into the details of the model but refer to the original paper by Englander and co-workers [3].

17.3 YOMOSA MODEL

A more realistic but still a highly simple model was developed by Yomosa [5,6]. This model was later given a descriptive name: 'the dynamic plane-base rotator (DPBR) model' by Takeno and Homma (see the next sub-section). Yomosa's Hamiltonian [6] can be expressed as:

$$H = \sum_n \left\{ \frac{1}{2} I \left(\dot{\phi}_{a,n}^2 + \dot{\phi}_{b,n}^2 \right) + A \left(1 - \cos \phi_{a,n} \right) \right.$$
$$+ A \left(1 - \cos \phi_{b,n} \right) + B \left(1 - \cos \phi_{a,n} \cos \phi_{b,n} \right) \qquad (2)$$
$$\left. + S \left(1 - \cos \left[\phi_{a,n} - \phi_{a,n-1} \right] \right) + S \left(1 - \cos \left[\phi_{b,n} - \phi_{b,n1} \right] \right) \right\}$$

where a and b denote the two strands. The first term is for the rotational kinetic energy of the bases around their point of reference on the backbone and in the plane perpendicular to the helical axis, I is the corresponding moment of inertia. The second, third and fourth terms describe the H-bonding between paired bases. The mathematical form is obtained from the double Fourier expansion of rotational interaction energy of paired bases. Terms five and six form the sum of stacking and torsion energies in both strands, respectively.

Using fields of rotational angles in the continuum limit the Euler–Lagrange equations can be written. The parameter A close to zero is the solvent condition, where the hydrogen bonding energy between the bases in a pair nearly equals that between a base and water molecules of the surrounding solvent. Cases $A = 0$ and $A \neq 0$ were studied separately. The former gives solutions where motions of the base twist angles are uncoupled between the strands. They have either constant solutions or soliton solutions of a sine-Gordon equation. Also coupled in-phase and out-of-phase π-soliton solutions were found. The importance of these π-solitons was seen in the fact that some of these made the supposed rewinding of the dsDNA unnecessary during the duplication and transcription processes–the motion of this kind of self-sustaining entity of coupled solitons would open and close the zip formed by the paired bases. In the latter case similar solutions were found. In addition to the constant form ground state solutions there appeared uncoupled 2π soliton solutions, and coupled in-phase 2π soliton solutions of an extended sine-Gordon equation. Relevant parameter values and soliton number density (to compare with the experiments) were evaluated using methods of statistical mechanics. The model was extended later to include nonlinear collective modes of base rotations and to introduce coupling interactions between rotational motions and longitudinal and transverse vibrational motions into the plane base-rotator model [7].

17.4 TAKENO-HOMMA MODEL

Takeno and Homma continued the methodological development of the DPBR model [8,9]. Their main concern in the 1983 paper was the expression for the inter- and intra-strand potentials. After generalizing the model they also found that in the continuum limit the dsDNA structure may carry topological solitons (also called kinks and anti-kinks). Moreover, the original discrete model was found to have commensurate, incommensurate and chaotic phases in addition to a 'clean' Watson–Crick form of the DNA. This shows a coupling to 'chaos physics' which is one specific branch of nonlinear physics. The DPBR model is a version of the Frenkel–Kontrova model studied largely in nonlinear (mathematical) physics. The matured Hamiltonian in this line of dsDNA modeling can be found from Ref. [10]:

$$\begin{aligned} H = \sum_n \Bigg\{ & \frac{1}{2}\left(I_{a,n}\,\phi_{a,n}^2 + I_{b,n}\,\phi_{b,n}^2 \right) \\ & + K_{a,n}\left(1 - \cos\left[\phi_{a,n+1} - \phi_{a,n}\right]\right) \\ & + K_{b,n}\left(1 - \cos\left[\phi_{b,n+1} - \phi_{b,n}\right]\right) + V\left(\phi_{a,n},\phi_{b,n}\right)\Bigg\} \end{aligned} \tag{3}$$

Here the first term is the rotational energy of the bases around a fixed point within the plane perpendicular the backbone, I is the corresponding moment of inertia, the K-terms come from base stacking in both strands, each K is related to the nearest-neighbour interaction constant, and potential energy is of the form:

$$\begin{aligned} V\left(\phi_{a,n}, \phi_{b,n}\right) = & h_n\left(1 - \cos[\phi_{a,n} - \phi_{b,n}]\right) \\ & + g_{a,n}\left(1 - \cos\phi_{a,n}\right) + g_{b,n}\left(1 - \cos\phi_{b,n}\right) \\ & + d_n\left(1 - \cos_{a,n}\cos\phi_{b,n}\right) \end{aligned} \tag{4}$$

Here the first and the fourth terms of V correspond to the hydrogen-bonding energy and the dipole-dipole interaction energy of the nth base pair, while the second and third terms describe the local field energy. $\phi_{a,b}$ angles are depicted in Fig. 17.2. Parameters $g_{a,n}$ and $g_{b,n}$ represent the dipole-induced–dipole interaction energy between neighbouring bases in the same strand, so the choice $g_{a,n} = g_{b,n} = g_n$ is made and the potential energy becomes

$$\begin{aligned} V\left(\phi_{a,n}, \phi_{b,n}\right) = & h_n\left(1 - \cos\left[\phi_{a,n} - \phi_{b,n}\right]\right) \\ & + g_n\left(2 - \cos\phi_{a,n} - \cos\phi_{b,n}\right) + d_n\left(1 - \cos\phi_{a,n}\cos\phi_{b,n}\right) \end{aligned} \tag{5}$$

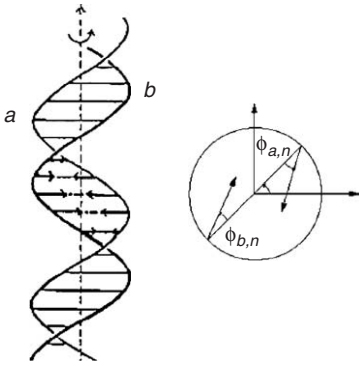


Fig. 17.2. The plane base rotator model idea, and rotation angles for Takeno's Hamiltonian (from Ref. [11]).

Analysis of the model reveals that the in-phase self-localized anharmonic base-pair rotation may be a preliminary step in the denaturation of DNA even at normal physiological temperatures.

17.5 ZHANG MODEL

In 1987, Zhang [12] contributed to the soliton methodology by studying the effects of dipole–dipole interactions on the soliton excitations in the dsDNA conformation. This study was a continuation of the Yomosa model so the Hamiltonian is not shown here. Furthermore Takeno in his 1990 paper mentions that the model considered there was an improvement of the Yomosa and Zhang works. This study also confirmed a possibility for solitons.

In 1990, Zhang continued the work with dsDNA [13] by studying a model where the stretching of the hydrogen bonds in the complementary bases was thought to be a relevant variable and proposed the following Hamiltonian:

$$H = \sum_n \left\{ \frac{1}{2} m \left(\dot{r}_{a,n}^2 + \dot{r}_{b,n}^2 \right) + \frac{1}{2} S \left(\left[r_{a,n} - r_{a,n-1} \right]^2 + \left[r_{b,n} - r_{b,n-1} \right]^2 \right) + \left(A y_n^2 + B y_n^3 \right) \right\} \tag{6}$$

Here again the order of the terms is typical: first the kinetic energy of the bases followed by the base–base interaction in the helix direction and finally the interaction of the base pairs. The symbols are: $r_{a,n}$ and $r_{b,n}$ are the displacements of the bases of the base pair n,m is the effective mass of a base, S is a parameter for the base–base stacking interaction, and the third term in Eq. (6) is a nonlinear potential energy function for the hydrogen bonds of the base pairs ($y_n = r_{a,n} - r_{b,n}$).

In the continuum limit the emerging equations of motion take the form of the Klein–Gordon equation. Owing to the nonlinear H-bond potential, the solutions contain solitary waves. In this model these are a kind of hole covering separated base pairs along a regime of several base pairs. Numerical solution of the discrete equations of motion shows that the solitary wave is unmovable and unstable. In other words it is a localized dampening vibrator. An estimate for the energy of the solitary vibrator is about 0.2–0.3 *eV*. This is much lower than, for example, the free energy of ATP hydrolysis or thermal energy, indicating that this kind of excitation is able to function as a promoter in the strand separation processes of dsDNA.

Zhang and Chou proposed in 1995 still another new simple model [14]. In this model the idea was that the stretching of the hydrogen bonds between the complementary bases also induces a rotation of bases around the backbone. This chest-expander-like model is illustrated in Fig. 17.3. It can be seen now that modeling is coming closer to the real DNA geometry. Namely, the bases have a more realistic connection to the backbone, and base–base direction in a pair seems to mimic the observed one. The Hamiltonian for this model is given as:

$$H = \sum_n \left\{ 2 \cdot \frac{1}{2} I \phi_n^2 + 2 \cdot \frac{1}{2} K \left(\phi_n - \phi_{n-1} \right)^2 + D \left(1 - \mathrm{e}^{-\alpha y_n} \right)^2 \right\} \tag{7}$$

where the first term is the rotational kinetic energy of the identical bases, I is their corresponding moment of inertia, the second term is for the stacking energy, and the last term describes the H-bond potential as a function of the angular displacement of the bases from their equilibrium direction and position. Displacement of the bases from their equilibrium distance, y_n, can be obtained geometrically from Fig. 17.3. In the limit $y_n = 0$ the equation of motion again reduces to a Klein–Gordon equation. This has a plane-wave solution allowing the plane waves to progress through the molecule. Depending on the boundary conditions a standing wave may also be found.

For a small amplitude motion the equation of motion extracted from the Hamiltonian reduces into a form of nonlinear Klein–Gordon equation. This again has a solitary wave solution, i.e., it is unstable or can be destructed by the normal phonons.

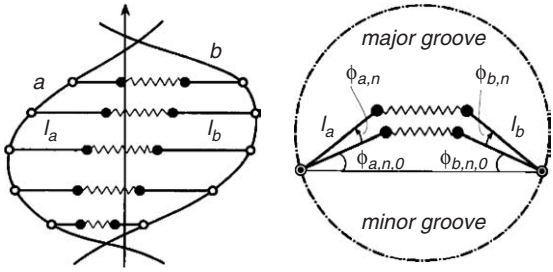


Fig. 17.3. The chest-expander model of Zhang *et al.* [14]. The left picture is the side view; The right picture depicts the dsDNA from the axial direction.

Numerical estimates of the amplitude and the width of the solitary wave give values of about 5°, and only one base-pair distance in the helix axis direction. These values are so small that they may disappear among the 'normal' fluctuations under dynamical stability of standard physiological conditions, discussed, e.g., in the context of the all-atom molecular dynamics simulations of dsDNA in an aqueous environment [15]. Also, owing to the smallness of the values, Zhang *et al*. [14] had already called them 'self-fluctuations'. Base rotation occurs in the major groove side. It is also mentioned that those fluctuations are too small to break the hydrogen bonds between the bases.

17.6 YAKUSHEVICH MODEL

Fundamentally dsDNA can be simplified to be formed of two long, elastic and weakly interacting rods, intertwined in the double helical form of dsDNA. This ingenious model was presented in 1989 by Yakushevich [16] as a model for dsDNA torsion dynamics. It is composed of two parallel rods where nucleotides are connected with harmonic springs in both strands and between the strands. A twist is the effect thought to cause dsDNA opening. The Hamiltonian can be given as:

$$H = \sum_{i,n} \left\{ \frac{1}{2} I_i \phi_{i,n}^2 + \frac{1}{2} K_i \left(\phi_{i,n+1} - \phi_{i,n} \right)^2 \right\} + \sum_n \left\{ \frac{1}{2} k (y_n)^2 \right\} \tag{8}$$

where the symbols are: $\phi_{i,n}$ are for the torsion angles of bases in both strands, I_i is the moment of inertia (all bases are assumed to be similar), K_i is the elasticity constant along the chains a and b, k is the transverse elasticity, and y_n is the stretch of the bases of the complementary pair n caused by the rotation of bases or points on the surface of a disc-like nucleotide.

An analysis of the Hamiltonian follows standard lines. Using principles of mechanics the equations of the motion have been derived. Then the continuum limit was assumed. This includes an assumption of small angles and small transverse elongations. First, the linear limit solution has been searched with the help of an assumption of a plane wave solution. This gives the dispersion relations composed of acoustic and optical branches. Then the general solution of the continuum limit equations of the motion has been searched under the assumption of running waves. This leads to an equation having a general solution like a sine-Gordon equation, double sine-Gordon equation, and a sine-Gordon-type equation. Based on this, dsDNA could be expected to have kink and anti-kink types of solitons.

This dsDNA model was later made more realistic by Gaeta [17–20]. The helicity was included in the model by one additional term in the Hamiltonian:

$$V_H = \frac{1}{2} \sum_{i,n} K_H \left(1 - \cos \left[\phi_{n+h}^{e(i)} - \phi_n^i \right] \right) \tag{9}$$

which describes the 'through space' non-bonded interaction in the helix direction between the nucleotide n and nucleotide $(n + h)$ (half pitch distance of the helix in the equilibrium conformation). Soliton solutions and scaling laws have been derived [21].

In 2002 Yakushevich *et al*. [22] published an investigation of the intramolecular mobility of a more detailed dsDNA model. This work continued the 'traditional' soliton line, but with a renewed, asymmetric helix model.

Now the model Hamiltonian was:

$$H - \sum_n \left\{ \frac{1}{2} I_{a,n} \phi_{a,n}^2 + \frac{1}{2} I_{b,n} \phi_{b,n}^2 + S_{a,n} \sin^2 \left(\frac{1}{2} \left[\phi_{n+1,1} - \phi_{a,n} \right] \right) + S_{b,n} \sin^2 \left(\frac{1}{2} \left[\phi_{n+1,2} - \phi_{b,n} \right] \right) + V_{ab} (\phi_{a,n}, \phi_{b,n}) \right\} \tag{10}$$

Here the first two terms are for the rotational kinetic energy of the n^{th} base pair. $I_{m,n}$ are the moments of inertia of the bases $m = a, b$. Rotations are assumed to occur around the backbones. The third and fourth terms are base–base interactions along the backbone of the dsDNA. Symbol $S_{m,n}$ is the interaction energy of the nth base with $(n + 1)$th base (approximated to be greater than 60 kJ/mol). The last term in the Hamiltonian is for the interaction of the bases of the hydrogen-bonded base pair. This is described by the potential:

$$V_{ab} (\phi_{a,n}, \phi_{b,n}) = K_{ab} \left(r_a \left[r_a + r_b \right] \left[1 - \cos \phi_{a,n} \right] + r_b \left[r_a + r_b \right] \left[1 - \cos \phi_{b,n} \right] - r_a r_b \left[1 - \cos \left(\phi_{a,n} - \phi_{b,n} \right) \right] \right) \tag{11}$$

Here r_a and r_b are the distances of the bases from the backbone. The interaction strength was estimated to be $K_{\text{AT}} = (2/3) \times K_{\text{GC}} = 0.234$ N/m. The dispersion equations have also been derived. Using these and the parameter estimates from the literature an estimate for the interaction strength K can be derived using a different method. The new estimated value, $K_{\text{AT}} = 4.744$ N/m is much higher than the previous value. The equations of motion obtained from previous Hamiltonians are too complicated to be solved analytically. Therefore a variation technique and numerical methods have been used to solve the dynamical system.

As a result of this work three types of topological solitons imitating the localized states with open base pairs are extracted from the model. These are able to move along the chain. How the movement occurs depends on the details of the helix. These solitons seem to be stable with respect to thermal 'disturbances'; the intramolecular phonons are not able to destroy them.

17.7 PEYRARD–BISHOP MODEL OF DNA DENATURATION

In 1989 Peyrard and Bishop (PB) published a paper connected to the same category as the previously described soliton models, but raising a completely different aspect [23]. While the previously touched simple models concentrated on 'hunting' for something soliton-like in dsDNA structure, PB were relinquished from the soliton obsession. Their initial idea, based on the widely spread consensus of the existence of local denatured

regions in native dsDNA, was in theoretical studies of the thermal denaturation, and possibly later in the application of the impending results in understanding the processes where dsDNA denaturates or melts completely.

An important fact emphasized for the first time was the discrete character of the molecule and the importance of having (and saving) this feature in the modeling as well. The earlier simple models had assumed the continuum limit and analyzed properties most often under this assumption. In their study Peyrard and Bishop used the transfer integral methodology. The analysis was composed of the statistical mechanical properties of the model and the determination of strand separation as a function of temperature.

The Hamiltonian of the model is given as:

$$H = \sum_n \left\{ \frac{1}{2} m \left(\dot{r}_{a,n}^2 + \dot{r}_{b,n}^2 \right) + \frac{1}{2} K \left(\left[r_{a,n} - r_{a,n-1} \right]^2 + \left[r_{b,n} - r_{b,n-1} \right]^2 \right) + V \left(r_{a,n} - r_{b,n} \right) \right\} \tag{12}$$

This model was made as simple as possible. Only a single mass parameter m for the extended bases, as well as only a single spring constant K for the whole strand was used. The on-site potential $V(r_{a,n} - r_{b,n})$ was chosen to have the form of Morse potential:

$$V(r_{a,n} - r_{b,n}) = D \left(\mathrm{e}^{-a(r_{a,n} - r_{b,n})} - 1 \right)^2 \tag{13}$$

Instead of directly using variables $r_{a,n}$ and $r_{b,n}$, fixed to the equilibrium strand positions, the in-phase and out-of-phase variables, x_n and y_n were defined by

$$x_n = (r_{a,n} + r_{b,n})/\sqrt{2}, \quad y_n = (r_{a,n} - r_{b,n})/\sqrt{2} \tag{14}$$

This change of variables factorizes the Hamiltonian H into x and y parts. The x-part is the Hamiltonian for the usual harmonic chain, while the nonlinear Morse-potential remains completely in the y-part. For large systems, when the number of bases or particles $N \to \infty$, the partition function of the y-part can be evaluated exactly in the thermodynamical limit. The use of the transfer integral methodology makes it possible to compute the free energy from the model. The same method can also be applied in the evaluation of the mean stretching, $\langle y_m \rangle$, of the hydrogen bonds.

Using the mean values for the H-bonds N–H ... N and N–H ... O found in the base-pairs, $D = 0.33$ eV and $a = 1.8$ Å^{-1}, $\langle y \rangle$ versus. T curves can be plotted. Figure 17.4 shows three such curves. The rapid increase in $\langle y \rangle$ starting at a certain temperature is clearly visible, which also agrees with experimental findings. What the experiments do not show but what comes out of this simple model is the dependence of the denaturation on the elasticity of the strands or parameter K, see again Fig. 17.4. An increase of K raises the denaturation temperature. The value of $K \approx 3 \times 10^{-3}$ eV/Å^2 gives a denaturation in reasonable agreement with the experiments. This low value was kept as an implication that the discreteness effects are so important that they cannot be omitted, as it was made in connection with many earlier soliton models.

The next question was how the denaturation would actually arise. Considering the equation of motion obtained from the y-part of the Hamiltonian, and concentrating on the

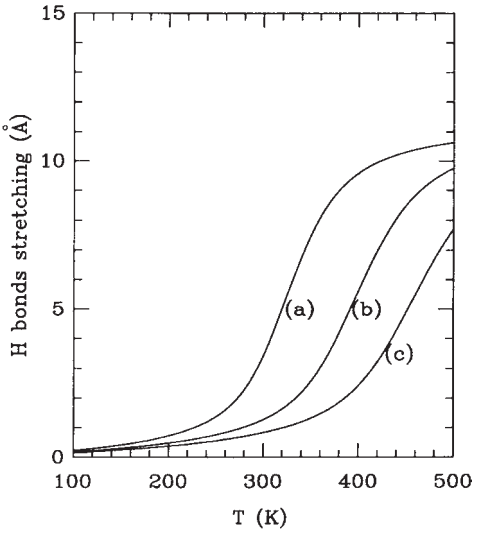


Fig. 17.4. Variation of the mean H-bond stretch <*y*> as a function of temperature for three values of coupling constant *K*. (a) $K = 2 \times 10^{-3}$ eV/A^2, (b) $K = 3 \times 10^{-3}$ eV/A2, (c) $K = 4 \times 10^{-3}$ eV/A^2 (from Ref. [23]).

early phases of the opening events called precursor phenomena, which allow use of small *y*-values and, therefore, series expansion of the $e^{(\ldots)}$-term give an equation of motion for small values of *y*. This can be solved using the multiple-scale expansion. Although the discreteness was found to be important it is much simpler to take first the continuum limit which appears to give qualitative results. As a consequence of this the nonlinear Schrödinger equation, familiar from nonlinear physics, is obtained. Its solutions are breathers; these are the same modes that can be observed in the infrared and Raman spectra of dsDNA. By calculating the mean variance, $\langle y^2 \rangle$ in the same way as the mean stretch, it can be noticed to have a rise of several orders of magnitude in a small temperature range around the denaturation temperature.

In this way this simplified model gave evidence that an energy localization due to the nonlinear effects in the dsDNA chain is a possible precursor, or starting point, for the denaturation process needed in the double helices' life cycles.

17.8 THE PROHOFSKY BRANCH OF THE PEYRARD–BISHOP MODEL

In 1989, Techera *et al*. [24] published their studies of the original PB model. They were searching possible energy transport mechanisms caused by the nonlinearity in the description of the model. The term 'quasi-solitonic' is used here as a synonym for solitary. The equation of motion is presented and the same change of variables is used as in the PB model. Hydrogen bonds are described by the Morse potential, but they point out that any function having the same characteristics could be used. Many other potentials are even mentioned as having been tested although detailed results are not given.

The existence of two separate time scales is given. These arise from the on-site vibrations and from the propagation along the chain. Numerically this leads to a stiff problem where the eigenvalues of the Jacobian matrix of the differential equations system vary largely and may cause problems if not handled properly. The system characterizing values were: base-pair distance, $h = 3$ Å $a = 2.5$ Å^{-1}, Morse well depth 0.4 eV, 125 base pairs, and periodic boundary conditions.

Consideration of the model equation is started by the series expansion of the Morse potential and including the assumption of small amplitude displacements. In the limit of long wavelengths (much longer compared with h) the continuum approximation is used and the familiar nonlinear Schrödinger equation (NLS) is obtained. NLS is a soliton equation whose analytical solution is a cosine wave modulated by a hyberbolic secant [25].

In the completely harmonic case it was noticed that any initial pulse can propagate over thousands of base pairs. If the initial impulse of a pulse is zero it splits into two parts travelling in opposite directions. Instead, if the pulse has initial velocity it determines its progression direction, and the pulse shows slowly emerging behavior, which in the case of this quasi-soliton is a certain kind of a pulse queue born due to a gradual dispersion of an exact soliton.

The general case, where there also exists nonlinearity, divides in two differing dynamic parts: the initial pulse propagates or becomes trapped depending on the relation of the parameters of the model. This (numerical) study reveals that the propagation depends on the balance between the nonlinear force (on a site) and shearing forces (between the sites). Generally within this model the trapping occurs when the nonlinear forces are comparable to the shearing forces or greater.

In the further analysis of the model a semi-discrete approximation using multiple-scale expansion was made [26]. This method appeared to be insufficient and the demand for completely discrete analysis became clear. The model was further enlarged to the cases where the next-nearest-neighbor interactions and the cubic nearest-neighbor interactions were taken into account. The basis for this was found from the fact that dsDNA is a salt in its biologically active state. The cubic term makes the trapping difficult. But on the other hand a sufficiently large nonlinearity in the stacking interaction makes the propagation easier.

In Ref. [27] the inhomogeneities via different masses and hydrogen bonds are discussed within the frame of this model. Owing to the latter, the well depths of the Morse potential change, being 0.2 eV for the A–T and 0.3 eV for the G–C pairs. On some sites the value 0.0 eV is used for open bases. Study of their effects on dynamical properties shows several possibilities depending on the value of stacking constant and permutations of the sequence: pinning pulse, dispersive pulse, and propagating pulse with reflection and transmission, partial reflections and total reflections. (All these are properties of the general NLS equation, met, e.g., in the studies of the discrete NLS equation [28]). In a completely heterogeneous chain, propagation of initially quasi-solitonic pulse was found to be strongly restricted.

17.9 DAUXOIS–PEYRARD–BISHOP DNA DENATURATION MODEL

To come closer to a dsDNA embedded in a cell or similar molecular environment the simple PB model was coupled to a heat bath in 1993 [29]. Attention was paid to the birth and

build-up of the pre-cursor excitations. Technically this was realized by applying molecular dynamics simulations using the Nose–Hoover method to maintain the temperature. When the Nose–Hoover thermostat was combined with the equation of motion of the PB model following the Hoover formulation, the equations of motion were obtained:

$$m\ddot{y}_n = K\left(y_{n+1} + y_{n-1} - 2y_n\right) + 2aD\left(\mathrm{e}^{-ay_n} - 1\right)\mathrm{e}^{-ay_n} - \xi m\dot{y}_n$$

$$\xi = \frac{1}{M}\left(\sum_n m\dot{y}_n^2 - NT\right) \tag{15}$$

Here M is the Nose thermostat's control parameter. Used values of the parameters were: $D = 0.04$ eV, $a = 4.45\ ^{-1}$, $K = 0.06$ eV Å^{-2}, $m = 300$ amu and $M = 1\ 000$.

Equation (15) originates from the y-part of the total Hamiltonian, which is the relevant part with respect to the issue. Owing to the simplifying character of the overall model the results are even better than just qualitative. The numerical simulation results show that in a temperature scan from 200 to 540 K the fluctuations are far from the energy equipartition. For increasing T there appear always larger denaturation bubbles until at high T when the whole system becomes denatured. In this melted state all the bases in the base pairs are at a distance which is on the plateau region of the Morse function.

Examination of the time development of the system at various temperatures shows the dynamical features of the model (see Fig. 17.4 in Ref. [29] for details). The time–site plots show in large scale lines in time direction due to the existence of long-lived excitations. These lines are denser the lower the temperature. A more careful inspection of one of these lines reveals that they are broken in a nearly regular manner. This is caused by the breathing of the localized excitations; at one moment of time the distance between the bases is quite small, i.e., close to the rest state, while at another time point the separation becomes large. This fluctuation is seen in the dotted lines in Fig. 17.4 of Ref. [29]. Another striking feature is the fusion of the local excitations and larger bubbles formed from them. This seems to take place at all temperatures, but at high temperatures it leads to a denaturation. There seem to occur fusion–fission reactions in the course of time when the temperature is under a certain threshold value. Unfortunately this aspect has not been studied in Ref. [29] in more detail. However, the shown figures tell us that at lower temperatures smaller bubbles fuse forming larger ones but they also break in the course of time. At higher temperatures there is a tendency to form separated bases covering the whole chain leading finally to a complete melting. This local excitation and bubble fusion at increasing temperature resembles very much the formation of the base-stack structure seen in the all-atom molecular dynamics simulation of a dsDNA stretching when external force increases continuously [30].

Wave vector-dependent spectral functions can be used to study the molecular dynamics results. Fourier-transform-based power spectrum reveals the frequencies of the system. In this case, comparison of the spectrum at various temperatures shows a shift in the phonon peak from a higher frequency to a lower frequency, i.e., a mode softening. Looking at the Morse function the peak frequency moves from the well bottom towards its plateau.

Transition temperature can be evaluated from the specific heat behavior for increasing temperature. This can be computed from the data of molecular dynamics simulation, or from the transfer integral results. The estimate of it for this model using either way reveals that the denaturation has a broad transition around 500 K, far above the observed value. This shows the quite typical property or danger of the simple models when used for complicated biological systems. Quantitative results are rarely or not easily obtained, while it is always possible to obtain a qualitative description up to a certain level.

Within this model the transfer integral (TI) method does not give analytical expressions for the thermodynamic quantities. Therefore, alternative means have been researched and used. One of these is the self-consistent phonon (SCP) method. Application of the SCP method for this modeling case is described very nicely in the original publication [31], therefore we omit (again) the lengthy mathematical details here. The results are given as free energies. The first-order approximation is good at low temperatures but becomes poor when the temperature is increased. Taking into account the second-order correction improves the situation roughly up to 300 K. At very high temperatures ($T \approx 500 - 600$ K) the free energy coincides with that obtained for the harmonic chain. The SCP method cannot be used to calculate 'simple' unique formulas for the thermodynamic quantities. This depends on the nature of the denaturation containing nonlinear elements which are not accessible by using the perturbation procedure included in the SCP method.

A shortcoming in the original PB model was the lack of an efficient energy localization method. This caused the observed high transition temperatures. A repair to this feature is presented in Ref. [31] by replacing the harmonic stacking interaction energy by the nonlinear form given as:

$$V^{\mathrm{DPB}}\left(y_{n,}\, y_{n-1}\right) = \frac{1}{2}\, K\left(1 + D\mathrm{e}^{-a(y_n + y_{n-1})}\right)(y_n - y_{n-1})^2 \tag{16}$$

The main justification for the choice of the anharmonic interaction comes from the observations indicating that stacking energy is a property of the base pairs themselves rather than the single bases alone. When base pairs break, the electronic structure of their complex will be modified in such a way that the stacking interaction with adjacent bases decreases. In the chosen potential energy function this effect is guaranteed by the usual quadratic pre-factor term $(y_n - y_{n-1})^2$.

This anharmonic potential makes the denaturation transition sharp. It also lowers the transition temperature. In the specific heat a very narrow first-order phase transition-like peak can be seen. The dynamic behavior together with a heat bath is the same as with the harmonic term, and so are the qualitative results of the SCP analysis. But, the overall thermal behavior of the model with the nonlinear stacking interaction agrees much better with the observed denaturation than when using the harmonic model.

17.10 FURTHER ANALYSES OF THE DPB MODELS

Owing to the one-dimensional character of the Dauxois–Peyrard–Bishop (DPB) model and inclusion of only the nearest neighbor interactions, the statistical mechanics of the

model can be treated exactly with the transfer integral method. Standard techniques and Kellog's method have been used to evaluate the essential part of the partition function obtained from the model Hamiltonian, see Ref. [32].

The eigenvalue spectrum of the calculated or computed transfer integral operator has been used to understand the denaturation transition. Moreover, the effect of the on-site potential has been studied carefully with other functional forms. These studies revealed two simultaneous requirements for the model: coupling along the chain, and specific on-site potential. Coupling in the chain direction must not be too tight, however. A looser coupling increases the entropy which drives the transition process. Morse potential form appears also to be very important. It has the plateau, which affects the transfer operator spectrum by giving it a continuum part. Furthermore the transition is possible if the energy increase can be balanced by the corresponding entropy increase to reduce enough the free energy of the denaturated phase (above the transition temperature). For this the Morse plateau is effective in addition to the special nonlinear coupling. Both make the mobile area, or easily allowed space of the bases, larger.

The natural next extension to the PB/DPB models was the inclusion of disorder. In Ref. [33] the PB model has been studied for an inhomogeneous longitudinal coupling constant. The constant K has been replaced by site-dependent constants K_n. The idea behind this is the fact that enzymes direct the function of DNA and an enzyme close to dsDNA may give rise to a perturbation to initiate an energy localization in it. This idea here is similar to what was found earlier in connection to nonlinear lattices with impurities [34]; and follows the known lines of the so called 'butterfly effect'. An impurity, due to its individual modes, can act as a catalyst for fusion of two breathers and generation of a larger excitation. The ability of these types of impurities to function as general sources of energy trapping in biological environments may, after all, be very limited due to the quite selective manner in which they capture modes travelling along their host structures. The search towards non-impurity mechanisms can also be argued, based on the known feature of dsDNA to bend. Bending will change the nucleotide–nucleotide coupling, not their constant physical properties like mass.

Because the emerging discrete equations of motion do not have an analytical solution, the continuum limit case has been examined first. This is made based on Taylor series expansion (recall the involved small amplitude limit!). The obtained equation is transferred into a perturbed nonlinear Schrödinger equation using multiple-scale expansion. Collective coordinate analysis has been applied on this. Application contains a soliton Ansatz and crude approximations. A set of three nonlinear differential equations for the collective variables can be obtained. However, these do not have analytical solutions, so numerical integration is the only tool to extract information out of them. The same, given soliton Ansatz has been used as the initial state of the supposed breather for these computations.

Simulations show the same behavior revealed by the collective coordinate method; the small amplitude breathers are movable while the larger ones are trapped in the chain. The moving vibration shows a narrowing at the defect site for negative perturbation. In other words, it concentrates the energy in a narrower or localized interval on the chain, thereby amplifying the property of base-pair opening. In the case that a moving breather meets a positive perturbation, the energy distribution becomes broader. Sometimes, with certain parameter values, the breather changes its oscillation frequency when it crosses

the perturbing defect. Here the positive perturbation means a larger value of the harmonic coupling constant in a site compared with the other sites, and vice versa.

When more breathers are created in the system and a trapping is obtained with one of them, it often happens that the first trapped breather is pushed away and an incoming one takes its place. Sometimes it also happens that the second breather is reflected from the already trapped one. More detailed analysis revealed that in a collision of one travelling breather with a trapped breather part of its energy transfers to the trapped one, while part of it continues its progress. If the system is periodic, then the same occurs every time the traveling quasi-particle meets the trapped one. If there are more of both trapped and travelling breathers the same will occur in each collision. An energy decrease of the trapped partner has never been observed. This means that the defect site collects the energy and, consequently, a larger mutual fluctuation of the bases on that specific base-pair site becomes possible. This phenomenon cannot be seen with the collective coordinate method due to its origin in the mutual interaction of participating breathers in the presence of a suitable defect.

Therefore, by believing in the prediction power of the simple nonlinear model for the actual biology, this study of variable elasticity in the chain justifies the conclusion that an enzyme or other protein molecule may function as an external stimulus for the energy localization process in the dsDNA, and therefore initiate the transcription process. Further, the extended perturbation in the chain seems to be more effective in the trapping sense than a isolated impurity [34].

Zhang *et al*. [35] tried to find the solutions for the inhomogeneity and partition function problems in connection to the PB model. Inhomogeneity, as said, is needed to describe a natural dsDNA. The limitations of the SCP method applied on a strongly nonlinear system are shown clearly (see Fig. 17.6 of Ref. [35]). The results for harmonic and anharmonic stacking potentials give nearly the same $\langle y \rangle$ versus T curves, while the TI method shows the already found clear difference, and, moreover, a smoother bending in the transition. Boundary effects together with short chains do not seem to have any effect, i.e., the end regions of the helix $\langle y \rangle$ behave exactly in the same way independently of the helix length. On the other hand, the existence of the ends causes a broadening of the transition region. This is larger than what is observed experimentally [36]. The critical temperature point is also shifted as a function of chain length. These may be caused by the nonoptimized parameter values in the model or the limited consideration only to the nearest neighboring interactions.

Applying the statistical mechanics machinery to heterogeneous chains (by taking into account the A–T and G–C base pairs) and plotting $\langle y \rangle$ as a function of temperature for differing G–C contents, it has been proven that the DPB model is able to give the observed T_c trend on the amount of G–C content [37].

Sub-transitions of the melting process are also studied in this work within the DPB model. Observationally these emerge in the ultraviolet radiation absorbance spectra and are visible as sharp peaks having a width of about 0.3–1.0 K. Their origin is the denaturation of base-pair regions of dsDNA composed of homogenous pairs. A reasonably high-precision approximate algorithm is developed to compute the melting profiles of DPB-DNAs with arbitrary sequences. This is based on the ‘extended transfer matrix approach’ (ETMA) method. The computed melting profiles for 40 and 200 base pair long sequences show that the shorter one has a more structured curve than the longer one.

Refined analysis reveals that melting can be divided into the A–T and G–C melting regions. For increasing temperature, the melting of a dsDNA, sequence begins by bubbling of A–T pairs (or segments) followed by gradually strengthening similar behaviour in the G–C segments. The resolution of the ETMA methodology is 0.3 K while in the molecular dynamics studies it stays somewhere around 10 K.

The first study to compare the PB model with exactly the same experimental set-ups appeared in 1998 [38]. Calculated melting curves for 27 and 21 basepair long oligomers were shown. These produce nicely the experimental curves in spite of the sequence and concentration effects (here the concentration was really taken into account!). The complete dissociation of the two single strands, in addition to local opening, in the short chains was discussed.

Singh and Singh [39] extended this analysis to short defected sequences. These defects were mismatched bases to form actual base pairs. In the modeling these were generated by a semi-flat function following the Morse curve at the repulsive part. This function becomes flat at the abscissa value where the Morse function reaches the value it has while approaching infinity. In this way the potential well produced by the hydrogen bonds is taken away. Influence on the electronic structure was also taken into account via a stacking interaction potential parameter. Calculated results for the sequence studied experimentally elsewhere [40] showed that the PB model with modeled defects on various sites of the chain still works well.

In the development of the simple models and potential functions for them it is important to couple the chemistry in the form of the function as well as in the few parameters involved in the mathematical formulas. One such study of the PB model with a base pair describing potential different from the Morse potential is in Ref. [41]. The source of the potentials for this study comes from the super-symmetry formalism. The studied potential is obtained by factorizing the Hamiltonian in a Schrödinger-type equation, and applying the super-algebra formalism to obtain the isospectral potential from the harmonic oscillator. The functional form of obtained potential is:

$$V(y) = ay^2 - \frac{1}{\beta^2 K}\,\frac{\mathrm{d}}{\mathrm{d}y}\left(\frac{\mathrm{e}^{-\sqrt{aK}\beta y^2}}{\Gamma + \int_0^y \mathrm{e}^{-\sqrt{aK}\beta q^2}\,\mathrm{d}q}\right) \tag{17}$$

In this equation the y is the breathing motion of a base pair, and a, K and Γ are parameters, of which K relates to the stacking interaction (spring constant) along the chain, a describes the harmonic part of the base-pairs' interaction potential, and Γ relates to the singularity of this potential. Values for these can be found from the experiments. The qualitative advantage of this potential is its character to couple the intra-site H-bond interaction and inter-site stacking interaction together. Other potentials used with this kind of dsDNA study are not known to have this feature. Unfortunately, the authors compare their melting curve only with the original PB curve. The curve obtained with Eq. (17) resembles those obtained by the DPB model, it has a sharper knee and a more reasonable value of T_c than that which the original model gives.

De Luca *et al*. [42] used the machinery of the nonlinear dynamical systems developing under nonlinear physics to study the original homogeneous PB model, focusing on the energy localization conditions when initially a small group of neighboring base pairs (oscillators by their terminology) is excited and the oscillation amplitude is small

outside this region on a long time-scale. The existence criteria can be found by means of information entropy. Localized excitations exist when the initially oscillating subgroup oscillates with a frequency inside the reactive band of the linear chain. The reactive band is opposite to the dispersive or radiative band, for which it is typical that the localized frequencies of the system which are there lead to quasi-normal mode excitations, which again are most often delocalized in space. With the same parameter values as used with the original PB model, they predicted a robust localization from the PB model justifying the room-temperature localization in this framework. In a way this gave an answer to the questions of Refs [31,43], where the localization was studied specifically in the context of the PB model, but in order to find more general criteria.

Very recently Ares *et al*. [44] have used Monte-Carlo simulations (the Metropolis algorithm) to evaluate the opening steps within the physiological temperature range in the DPB model. They computed the mean profiles $\langle y_n \rangle$ at selected temperatures between 40 and 100 C° ($\approx$ 310 – 375 K) to obtain the fraction of open pairs (f) at each temperature. They used the parameters of Ref. [38]: K = 0.025 eV/Å^2, $\rho = 2$, $\beta = 0.35$ Å^{-1}, $D_{GC} = 0.075$ eV, $a_{GC} = 6.9$ Å^{-1}, $D_{AT} = 0.05$ eV, and $a_{AT} = 4.2$ Å^{-1}, which are obtained by fitting to thermodynamic properties. The threshold value for an open pair was chosen to be 0.5 Å.

These simulations are closely related to the experimental works [45,46], considering exactly the same sequences. Study was concentrated on two cases, bubbles in the middle and bubbles at the end of the oligomers. The former has an A–T rich middle part, while in the latter the A–T pairs are concentrated at the ends. The sequence lengths were 60, 42 and 33 base pairs. The average fractional length $l_{f,ave}$ was used to monitor the simulation results. By definition [44]

$$l_{f,ave} = \frac{\sigma}{1-p} \tag{18}$$

where f is the averaged fraction of open base pairs, p is the averaged fraction of denaturated sequences and $\sigma = (fp)$ is the fraction of bases participating in a bubble state at a given temperature.

The results are excellent. They show that the simple refined nonlinear theoretical model, with only few parameters, can accurately reproduce observations on the dsDNA melting, not just for the melting curve, but also for the formation and role of the bubble states in the pre-melting regime.

17.11 SIMPLE HELICOIDAL MODEL OF DSDNA

Nature has a purpose for all its events, functions, processes, structures, etc. Therefore the helicity of the genotype preserving and transferring molecule cannot be an accidental, arbitrary form of the standard dsDNA molecule. Specifically, the known extra twist accumulating in transcription in dsDNA indicates that the conformational form has to be better realized in the models of dsDNA. While the helix itself would not have a leading role in the processes needing or causing the denaturation of the structure it seems quite obvious that it at least has to be taken into account in the models where a more extensive view of the events is the fundamental aim.

In 1999, Barbi, Cocco and Peyrard introduced a dsDNA model cogniscent or the helicoidal structure of dsDNA [47,48]. While most of the earlier ideas of the three dimensional form of dsDNA were concentrated on demonstrating the existence of the more or less abstract static or even dynamic solitary states in the dsDNA structure, this new evolution was based on the experiences gained by the studies with the Peyrard–Bishop denaturation model and its versions as well as on the pieces of findings cumulated about this biological memory both alone and in interactions with other molecules. As such their model can be kept as an extension of the PB approach towards a more realistic description of biological processes.

The presented model takes into account the twist-opening interactions due to the helicoidal molecular geometry. Each base (actually the whole base–sugar complex, but for short termed as 'base') is considered as a single, nondeformable object. The sugar–phosphate backbone between two base pairs is modeled as an elastic rod. The model idea is represented graphically in Fig. 17.5.

The model has two degrees-of-freedom per base pair: one radial variable related to the motion of the bases along the diameter joining their attachment points to the helicoidal backbone, and a twist angle of each base pair defined by the angle between the diameter and a reference direction, see Fig. 17.5. This twist angle, which increases from one base pair to the next one, creates a helicoidal structure in the model for dsDNA. The twist motion is restricted on the basepairs plane. In the first version of the model and related studies the center-of-mass of each base pair was fixed leading to a symmetric motion of the bases with respect to the helix axis. Later the model has also been used with a rigid backbone and moving base-pair planes [48].

The coordinates of one base pair are: position r_n, and angle ϕ_n of one of the two bases with respect to a fixed reference frame. The Lagrangian for the model is given as:

$$L = \sum_n \left\{ m\left(\dot{r}_n^2 + r_n^2\dot{\phi}_n^2\right) - D\left(\mathrm{e}^{-\alpha(r_n - R_0)} - 1\right)^2 \right.$$

$$\left. - K\left(l_{n,n-1} - l_0\right)^2 - S\left(r_n - r_{n-1}\right)^2 \mathrm{e}^{(-b[r_n + r_{-n-1} - 2R_0])^2} \right\} \tag{19}$$

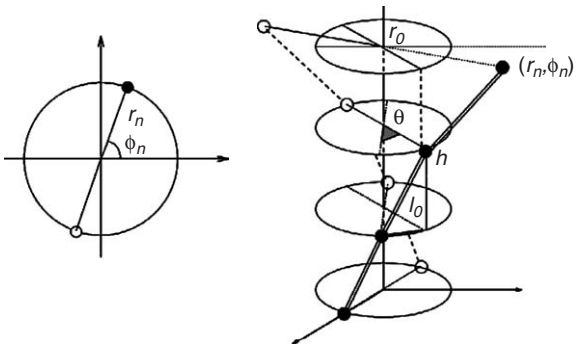


Fig. 17.5. Graphical illustration of the Barbi *et al*. model [49].

where

$$l_0 = \sqrt{h^2 + 4R_0^2 \sin^2(\theta/2)} \tag{20}$$

and

$$l_{n,n-1} = \sqrt{h^2 + r_{n-1}^2 + r_n^2 - 2r_{n-1}r_n \cos(\phi_n - \phi_{-1})} \tag{21}$$

The symbols l_0 and $l_{n,n-1}$ are for the equilibrium and the actual distances between the two bases n and $(n - 1)$ along both of the strands. The first term in the Lagrangian is the kinetic energy. The second term is the base-pair interaction; the Morse potential with equilibrium distance R_0. The quadratic term in $(l_{n,n-1} - l_0)$ in the third component of the model represents the elastic energy of the backbone rods between neighboring base pairs on each strand. The last term models the stacking interaction between neighboring base pairs having the effect of decreasing the stiffness of the open parts of the chain relative to the closed ones and stabilizing the latter with respect to the denaturation of a single base pair. Terms of this type are said to increase the cooperative effects close to the melting transition.

The parameters of the model finally create the desired structure. In dsDNA, the equilibrium helicoidal structure originates from the competition between the hydrophobic effect (which tends to eliminate water molecules from the core of the molecule by bringing the neighboring base-pair planes closer to each other) the electrostatic repulsion between neighboring base planes (which has the opposite effect) and the rigidity of the two strands (that separates the external ends of the base pairs by a fixed length related to the phosphate length). The first two forces are in this model combined in a unique stacking effect that fixes base-plane distance h. As the equilibrium backbone length l_0 is greater than h, it is then necessary to incline the strands in the typical helicoidal structure to minimize the energy. The geometrical parameters h and R_0 have been chosen to match the structure of B-DNA. All other geometrical parameters can in a straight forward manner be fixed according to the available structural data [47,48,50,51]. Much more subtle is the choice of the parameters *b, D, S* and *K* gauging the effective forces. The choice of the parameter *K* can be independently derived from the twist persistence length, see Refs [52,53], while the choice of *b* and *S* is based on a comparison with recent mechanical denaturation experiments [54,55]. The parameter *D*, which sets the main energy scale, has been tuned to reproduce the experimental value of the denaturation temperature $T_c = 350$ K as closely as possible. The full set of parameters for this model according to [49] and [51] is summarized in Table 17.1.

This helicoidal model is first considered in the limit of small displacements. From the Lagrangian, the equations of motion for small displacements with respect to the equilibrium position can be obtained. Analytical (approximated) solutions for the small amplitude nonlinear distortions of the molecule are derived by the method developed in Ref. [50]. A radial breather-like opening is shown to be associated with the angular untwisting of the molecule. Numerical results of dynamical simulations of the model show that these solutions are stable for long periods of time and can move along the chain. In this context the word small needs an explanation. The base deviations from their equilibrium states are of the order of a few picometers, and the angle deviations also are parts per hundred. Energetically, even these small geometrical shifts may bring about large changes.

Table 17.1 The parameters of the helicoidal DNA model

Parameter	Symbol	Value	Units
Morse potential range	a	6.3	$Å^{-1}$
Stacking interaction range	b	0.5	$Å^{-1}$
Morse potential depth	D	0.15	eV
Stacking interaction coupling	S	0.65	eV $Å^{-2}$
Interplane distance	h	3.4	Å
Elastic longitudinal coupling	K	0.04	eV $Å^{-2}$
Equilibrium distance	R	0 10	Å
Twist angle	Θ	0.60707	rad
Base masses	m	300	amu

According to Refs [49,51]

Another size limit was considered by Campa in 2001 [56] by studying large bubble propagation within this model. This study already contained an extension of the model to heterogeneous sequences giving mass variations among the bases. Most of the analyses have then been performed computationally, so the question of descriptiveness and solvability has been handled reasonably in this way. Analyses show that a bubble can form and dynamically preserves its stability, even in the presence of thermal effects, over 1 000 base pairs. It is worth mentioning that this model can handle more than one bubble at the same time in the helix. This is a feature which is not trivially included in other simple models. However, it is a feature, which nature may be using in a multi-transcription process.

This reduced helicoidal model has also been used in understanding the single molecule manipulation results, especially the effects of an external twist, see Ref. [50]. This, of course, adds one more mathematical term in the model to make the external end-torsion possible. It should be possible to use the simple helicoidal model for other types of analyses with small enhancements, for example, for DNA loops.

17.12 CONCLUDING REMARKS

It is quite clear from the above that the nonlinear simple models of double-stranded DNA denaturation are strongly phenomenon-specific. In the case of the intramolecular evolution of the dsDNA during its lifetime, nonlinearities are needed due to the large amplitude intramolecular motions. Consequences arising from the nonlinearities in the system are the related properties.

Roughly 20 years ago it was assumed to be possible to show the existence of solitons (or something like that) in the dsDNA structure. Today the mere word soliton or solitary can only be understood to express the main class name. The underlying observations have to be specified with their own names, which again may have their sub-families. However, what the simple models have revealed from the DNA structure are the terms under the soliton class: kinks, breathers and bubbles. All these are observable static and dynamic 'particles' or quasi-particles emerging in dsDNA.

Another issue this survey through the dsDNA denaturation models shows is the traditional physical approach to model building. In the first phase the model is attempted to be as simple as possible. This means inclusion of only the utmost necessary variables while abandoning such details as the inhomogeneities of nucleic acids, the molecular environment, defects, etc. The purpose in this kind of modeling is to find the most fundamental features governing the object with quite an idealized but descriptive mathematical model. In the light of the history of reductive modeling, when the computers were not so helpful, one important issue influencing the model creation and its processing has been its solvability. Today the models can still be kept simple but finding of analytical solutions is not (or should not be) the fact directing the approximation process. Many simplifications can be left aside when the solution search is given to the computers. An excellent example of a general model progress is the Peyrard–Bishop model line described above. However, what one must not completely forget in the analytical solution attempts is the obvious and maybe strong but often very much hidden wish to find a suitable analytical formula making the future easier and thus avoiding the (less elegant) numerical treatment.

17.13 REFERENCES

1 L.V. Yakushevich, Nonlinear physics of DNA. Wiley physics titles for higher education, second edition, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2004.
2 T. Dauxois and M. Peyrard, Physics of solitons, Nonlinear science and fluid dynamics, Cambridge University Press, Cambridge, 2006.
3 S.W. Englander, N.R. Kallenbach, A.J. Heeger, J.A. Krumhansl and S. Litwin, Proc. Natl. Acad. Sci. USA, 77 (1980) 7222.
4 S.T. Thornton and J.B. Marion, Classical dynamics of particles and systems, fifth edition, Belmont, CA: Thomson-Brooks/Coleo, 2003.
5 S. Yomosa, Phys. Rev. A, 27 (1983) 2120.
6 S. Yomosa, Phys. Rev. A, 30 (1984) 474.
7 S. Yomosa. J. Phys. Soc. Japan, 64 (1995) 1917–1935.
8 S. Takeno and S. Homma. Prog. Theor. Phys., 70 (1983) 308.
9 S. Homma and S. Takeno. Prog. of Theor. Phys., 72 (1984) 679.
10 S. Takeno and S. Homma, J. Phys. Soc. Japan, 59 (1990) 1890.
11 S. Takeno, Phys. Lett. A, 339 (2005) 352–360.
12 C.-T. Zhang. Phys. Rev. A, 35 (1987) 886.
13 C.-T. Zhang. J. Phys. Condens. Matter, 2 (1990) 8259.
14 C.-T. Zhang and K.-C. Chou, Chem. Phys., 191 (1995) 17.
15 R. Lohikoski, J. Timonen, A. Lyubartsev and A. Laaksonen, Mol. Simul., 29 (2003) 47.
16 L.V. Yakushevich. Phys. Lett. A, 143 (1989) 413.
17 G. Gaeta, Phys. Lett. A, 143 (1990) 227.
18 G. Gaeta, Phys. Lett. A, 168 (1992) 383.
19 G. Gaeta, Phys. Lett. A, 172 (1993) 365.
20 G. Gaeta, C. Reiss, M. Peyard and T. Dauxois, Nuovo Cimento Soc. Ital. Fis., D17 (1994) 1.
21 J.A. Gonzales and M. Martin-Landrove, Phys. Lett. A, 191 (1994) 409.
22 L.V. Yakushevich, A.V. Savin and L.I. Manevitch, Phys. Rev. E, 66 (2002) 016614.
23 M. Peyrard and A.R. Bishop, Phys. Rev. Lett., 62 (1989) 2755–2758.
24 M. Techera, L.L. Daemen, and E.W. Prohofsky, Phys. Rev. A, 40 (1989) 6636.
25 P.G. Drazin and R.S. Johnson, Solitons:an introduction: Cambridge University Press, Cambridge, 1989.
26 M. Techera, L.L. Daemen and E.W. Prohofsky. Phys. Rev. A, 40 (1990) 4543.
27 M. Techera, L.L. Daemen and E.W. Prohofsky. Phys. Rev. A, 42 (1990) 1008.

28 R. Lohikoski and J. Timonen, Physica Scripta, T33 (1990) 202.
29 T. Dauxois, M. Peyrard and A.R. Bishop. Phys. Rev. E, 47 (1993) 684.
30 R. Lohikoski, J. Timonen and A.Laaksonen, Chem. Phys. Lett., 407 (2005) 23.
31 T. Dauxois, M. Peyrard and A.R. Bishop, Phys. Rev. E, 47 (1993) R44.
32 T. Dauxois and M. Peyrard, Phys. Rev. E, 51 (1995) 4027.
33 J.J.-L. Ting and M. Peyrard, Phys. Rev. E, 53 (1996) 1011.
34 K. Forinash, M. Peyrard and B. Malomed. Phys. Rev. E, 49 (1994) 3400.
35 Y.l. Zhang,W.-M. Zheng, J.-X. Liu and Y.Z. Chen, Phys. Rev. E, 56 (1997) 7100.
36 R.M. Wartell and A.S. Benight. Phys. Rep., 126 (1985) 67.
37 W. Saenger. Principles of nucleic acid structure, Springer Advanced Texts in Chemistry, Springer-Verlag, New York, 1984.
38 A. Campa and A. Giansanti, Phys. Rev. E, 58 (1998) 3585.
39 N. Singh and Y. Singh. Phys. Rev. E, 64 (2001) 042901.
40 G. Bonnet, S. Tyagi, A. Libchaber and F. R. Kramer. Proc. Natl. Acad. Sci. USA, 96 (1999) 6171.
41 E. Drigo-Filho and J. R. Ruggiero. Phys. Rev. E, 56 (1997) 4486.
42 J. De Luca, E.D. Filho, A. Ponno and J. R. Ruggiero, Phys. Rev. E, 70 (2004) 026213.
43 K. Forinash, T. Cretegny and M. Peyrard, Phys. Rev. E, 55 (1997) 4740.
44 S. Ares, N.K. Voulgarakis, K. Rasmussen and A. R. Bishop. Phys. Rev. Lett., 94 (2005) 035504.
45 Y. Zeng, A. Montrichok and G. Zocchi. Phys. Rev. Lett., 91 (2003) 148101.
46 Y. Zeng, A. Montrichok and G. Zocchi. J. Mol. Biol., 339 (2004) 67.
47 M. Barbi, S. Cocco and M. Peyrard. Phys. Lett. A, 253 (1999) 358–369.
48 M. Barbi, S. Cocco, M. Peyrard and S. Ruffo. J. Biol. Phys., 24 (1999) 97.
49 M. Barbi, S. Lepri, M. Peyrard and N. Theodorakopoulos, Phys. Rev. E, 68 (2003) 061909.
50 S. Cocco, M. Barbi and M. Peyrard. Vector nonlinear Klein-Gordon lattices: General derivation of small amplitude envelope soliton solutions. Phys. Lett. A, 253 (1999)161.
51 S. Cocco and R. Monasson, Phys. Rev. Lett., 83 (1999) 5178.
52 T.R. Strick, J.F. Allemand, D. Bensimon, A. Bensimon and V. Croquette, Science, 271 (1996) 1835–1837.
53 M. Barbi. PhD thesis, University of Florence, 1998.
54 T. Strick, J.-F. Allemand, V. Croquette and D. Densimon. Twisting and Stretching Single DNA Molecules. Prog. Biophys. Mol. Biol., 74 (2000) 115–140.
55 S. Cocco and R. Monasson. J. Chem. Phys., 112 (2000) 10017.
56 A. Campa, Phys. Rev. E, 63 (2001) 021901.

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 18

The effects of bridge motion on electron transfer reactions mediated by tunneling

Spiros S. Skourtis[1], Jianping Lin[2] and David N. Beratan[2]

[1] *Department of Physics, University of Cyprus, Nicosia 1678, Cyprus*
[2] *Departments of Chemistry and Biochemistry, Duke University, Durham, NC 27708, U.S.A.*

Abstract

We review recent developments associated with dynamical effects on electron transfer processes. The regimes of validity for standard rate formulations are discussed and explicit examples of dynamically coupled electron-transfer reactions are presented.

18.1 INTRODUCTION

Since the early formulation of electron transfer rate theories, an increasingly detailed molecular-resolution view has emerged of how structure, environment and dynamics influence reaction rates. Recent developments have brought attention to questions of how fluctuations in molecular structure influence these rates. Dynamical issues in electron transfer are the focus of this chapter. We explore the regimes of validity for standard rate formulations and we explore several specific examples of dynamically coupled electron transfer.

18.2 THE GENERAL NONADIABATIC RATE EXPRESSION

The familiar quantum expression for the nonadiabatic electron transfer (ET) rate arises from Fermi's golden rule [1–12] and is usually developed in a Born–Oppenheimer framework. The ET rate is a thermally weighted sum of energy-conserving transitions between initial $|D;i_D,i_B\rangle$ and final $|A;f_A,f_B\rangle$ vibronic levels (Fig. 18.1). The rate is given by

$$k_{ET}=\frac{2\pi}{h}\sum_{i_D}\sum_{i_B}P_{i_D}P_{i_B}\sum_{f_A}\sum_{f_B}\left|\langle D;i_D,i_B|\hat{T}|A;f_A,f_B\rangle\right|^2\delta\left[U_D^{\min}+\varepsilon_{i_D}+\varepsilon_{i_B}-\left(U_A^{\min}+\varepsilon_{f_A}+\varepsilon_{f_B}\right)\right] \qquad (1)$$

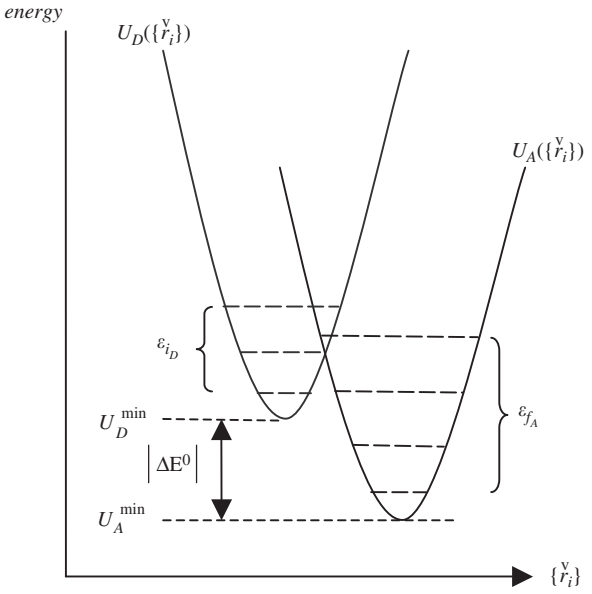


Fig. 18.1. Donor and acceptor potential surfaces and accepting mode vibrational energies.

where $|D;i_D,i_B\rangle = |D\rangle|i_D\rangle|i_B\rangle$ and $|A;f_A,f_B\rangle = |A\rangle|f_A\rangle|f_B\rangle$. $|D\rangle$ and $|A\rangle$ are the donor and acceptor electronic states, and $U_D^{\min}$ and $U_A^{\min}$ are the minimum energies of the donor and acceptor diabatic Born–Oppenheimer surfaces (Fig. 18.1). $|i_D\rangle$ and $|f_A\rangle$ are initial and final vibrational states of the donor and acceptor diabatic Born–Oppenheimer surfaces (vibrational energies ε_{i_D} and ε_{f_A}), and $|i_B\rangle$ and $|f_B\rangle$ are initial and final vibrational states of the bridge (vibrational energies ε_{i_B} and ε_{f_B}). The levels $|i_D\rangle$ and $|f_A\rangle$ are associated with atomic motion that modulates the D and A electronic energies (accepting modes), whereas $|i_B\rangle$ and $|f_B\rangle$ relate to bridge motions that modulate the tunneling barrier (inducing modes) [13]. In Eq. (1), $P_{i_D}P_{i_B}$ is the thermal probability of the initial vibrational state of the system $|i_D\rangle|i_B\rangle$. This probability is a product of the thermal probabilities of $|i_D\rangle$ and $|i_B\rangle$, because the accepting and inducing modes are assumed to be independent (a fair approximation for large bridges).

The matrix element $\langle D;i_D,i_B|\hat{T}|A;f_A,f_B\rangle$ in Eq. (1) is the total initial to final vibronic state coupling. In the Condon approximation it is given by

$$\langle D;i_D,i_B|\hat{T}|A;f_A,f_B\rangle \approx \langle i_D|f_A\rangle\langle D;i_B|\hat{T}|A;f_B\rangle. \tag{2}$$

This matrix element is a product of the overlap between the initial and final accepting mode vibrational levels, and of the bridge-mediated coupling between D and A electronic states, $\langle D;i_B|\hat{T}|A;f_B\rangle$. The latter depends, in general, on the initial and final vibrational levels of the bridge, because the bridge motion modulates the tunneling barrier and interacts with the tunneling electron. The bridge may therefore exchange energy with the electron

(inelastic tunneling). In this case, the tunneling matrix element operator $\hat{T}$ is a function of the electronic and vibrational states of the bridge, and of the electron–phonon interaction associated with bridge nuclear motion. An example of such an operator is presented in Section 18.9. In most of the following discussion, we assume elastic tunneling (a good approximation in the deep tunneling limit).

18.3 SEMI-CLASSICAL RATE EQUATION IN THE FRANCK–CONDON (STATIC BRIDGE) APPROXIMATION

In the Franck–Condon approximation, the dependence of the electronic coupling on bridge motion is ignored (static bridge approximation). $\langle D;i_B|\hat{T}|A;f_B\rangle$ is replaced by $\langle D|\hat{T}|A\rangle\langle i_B|f_B\rangle$, where $\hat{T}$ is a purely electronic tunneling matrix element operator. In the two-level system approximation and the weak-coupling limit

$$\hat{T} = \hat{V}\hat{G}^{(B)}(E_{tun})\hat{V} \tag{3}$$

where

$$\hat{G}^{(B)}(E_{tun}) = \left[E_{tun} - \hat{H}^{B}\right]^{-1} \tag{4}$$

is the bridge Green's function [14–16]. It describes the propagation of the electron via the bridge electronic states, which act as virtual intermediates for the tunneling electron. $\hat{V}$ is the Hamiltonian for the interactions between the D (A) electronic state and the intermediate bridge electronic states. $\hat{H}^{B}$ is the Hamiltonian describing the bridge electronic state energies and inter-state couplings, and E_{tun} is the tunneling energy of the electron.

The static bridge approximation in the nonadiabatic rate of Eq. (1) leads to

$$k_{ET} = \frac{2\pi}{\hbar}|T_{DA}|^2 \rho_{FC} \tag{5}$$

where

$$\rho_{FC} = \sum_{i_D} P_{i_D} \sum_{f_A} |\langle i_D|f_A\rangle|^2 \delta\left[U_D^{min} + \varepsilon_{i_D} - \left(U_A^{min} + \varepsilon_{f_A}\right)\right] \tag{6}$$

is the thermally weighted Franck–Condon density of states. In the high-temperature limit (classical nuclear motion), Eq. (5) produces the Marcus equation

$$\rho_{FC} = (4\pi\lambda k_B T)^{-1/2} \exp[-(\Delta E^0 - \lambda)^2/4\lambda K_B T] \tag{7}$$

where λ is the reorganization energy, ΔE^0 is the reaction free energy, T is the temperature, and K_B is Boltzmann's constant (Fig. 18.2).

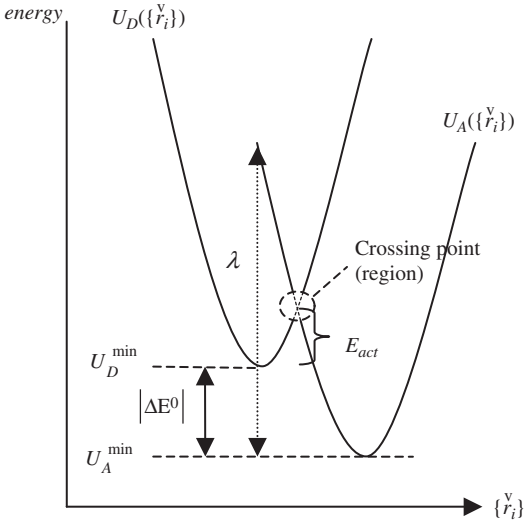


Fig. 18.2. Crossing region, activation energy, and reorganization energy associated with the donor and acceptor potentials.

18.4 DISTANCE DEPENDENCE OF THE ELECTRONIC COUPLING (TUNNELING MATRIX ELEMENT) IN THE STATIC BRIDGE APPROXIMATION

In the weak-coupling limit, the squared electronic coupling $\left(|T_{DA}|^2\right)$ decays approximately exponentially with distance through a periodic 1 D potential [17] with the decay factor β,

$$|T_{DA}|^2 \approx |T^0_{DA}|^2 \exp(-\beta r). \tag{8}$$

For different tunneling bridges linking D to A, β varies over a wide range. Large values of β (2–3 Å^{-1}) are expected theoretically for ET through vacuum for typical binding energies of $\sim$ 5eV [18,19]. Values of β are found to be about 1.0Å^{-1} experimentally [20–23] and theoretically [24–26] for ET via chemical and biological bridges. The β value found for ET through proteins is $\sim$ 0.9–1.4 Å^{-1} [27a,28]. The β value for tunneling through water varies widely based on theoretical estimates [29–31]. Experimentally, for tunneling through glassy water, Gray [27b] found $\beta \sim$ 1.61–1.75 Å^{-1}. Recently, Beratan and coworkers [32] showed that β associated with tunneling between proteins mediated by water can adopt distinct values in different distance regimes.

18.5 METHODS FOR THE COMPUTATION OF TUNNELING MATRIX ELEMENTS

For bridge-mediated ET, superexchange models are used widely to calculate the tunneling matrix element. Beratan, Onuchic and coworkers [18] developed the pathway model

that provides the simplest atomic-scale description of electronic propagation in proteins. Protein structures mediate tunneling by bonded, nonbonded, and H-bonded pathways between donor and acceptor, and the strongest pathway can be used to estimate the *D–A* interactions. The squared electronic coupling is estimated as

$$|T_{DA}|^2 = A^2(\prod_i \varepsilon_{bond}(i))^2(\prod_j \varepsilon_{space}(j))^2(\prod_k \varepsilon_{H-bond}(k))^2 \tag{9}$$

where A is a prefactor. The decay factors are $\varepsilon_{bond}\sim 0.6$ (for through-bond decay), $\varepsilon_{H-bond}\sim 0.36$ (for hydrogen-bond decay), and $\varepsilon_{space}\sim 0.6\times\exp[-\beta_S(R-2.8)]$(for through-space decay). β_s is ~ 1 Å^{-1} for a binding energy of ~ 5 eV.

The packing density approach [33], which is an average medium model for proteins, is also an empirical tunneling model. The electronic coupling square is given by

$$|T_{DA}|^2 = A^2\exp\{-[0.9\rho + 2.8(1-\rho)][R_{DA}-3.6]\} \tag{10}$$

where R_{DA} is the distance between donor and acceptor. ρ is the average packing density. The ρ's are linked to through-bond and through-space distance decay factors. The pathway and packing density models were recently shown to be essentially identical [34].

Quantum chemical approaches are becoming widely used to compute T_{DA}. The perturbation theory approach [14–16] uses Eq. (3) as a starting point. In a nonorthogonal atomic basis set, it gives [15,35,36]

$$T_{DA} = \sum_{i,j} (E_{tun}S_{Di} - H_{Di})\ \{(E_{tun}S - H)^{-1}\}_{ij}\ (E_{tun}S_{jA} - H_{jA}). \tag{11}$$

Here D and A indicate the donor and acceptor; S_{ij} and H_{ij} are the overlap and Hamiltonian matrix elements respectively (i,j are indices of atomic orbitals of the bridge). Other quantum chemical approaches to computing the donor–acceptor interaction include energy-splitting [37–41], tunneling current [42], and generalized Mulliken–Hush methods [43].

18.6 BRIDGE MEDIATED ELECTRON TRANSFER RATES FOR FLUCTUATING BRIDGES

Molecular structures are dynamic; fluctuations in local and global structure arise over a wide range of time scales at thermal equilibrium. Given the large variety of molecular ET bridges, and the diversity of *D*/*A* structure and solvent environments, it is not immediately clear when the static bridge (Franck–Condon) approximation is valid. There are several important questions associated with bridge fluctuations and ET rates: How do bridge fluctuations affect T_{DA} and the ET rate? What structural and dynamical characteristics of a molecular bridge determine the magnitudes and time scales of T_{DA} fluctuations? Do these fluctuation effects differ qualitatively between large (biological) bridges and small molecule bridges?

To address these questions, it is useful to begin with the nonadiabatic rate of Eq. (1). This expression is time independent and fully quantum mechanical. In the following, we use a time dependent version of Eq. (1) with a semi-classical approximation. The time

dependent, semi-classical approach provides intuition, and it makes contact with molecular dynamics methods used to compute ET rates.

We first outline the derivation of the semi-classical rate from the quantum expression. Substituting Eq. (2) into Eq. (1) and using the Fourier representation of the δ-function, leads to [44]

$$k_{\mathrm{ET}}=\frac{1}{\hbar^2}\int_{-\infty}^{+\infty}\mathrm{d}t\ \mathrm{e}^{\mathrm{i}\omega_{\mathrm{DA}}t}\ C_{T_{\mathrm{DA}}}(t)\ C_{\mathrm{FC}}(t) \tag{12}$$

where $\omega_{\mathrm{DA}}=(U_{\mathrm{D}}^{\min}-U_{\mathrm{A}}^{\min})/\hbar$ is the Bohr frequency corresponding to the ET reaction energy gap. k_{ET} is the fourier transform (evaluated at ω_{DA}) of the time dependent Franck-Condon factor $C_{\mathrm{FC}}(t)$ and the electronic coupling correlation function $C_{T_{\mathrm{DA}}}(t)$ product. The former is

$$C_{\mathrm{FC}}(t)=\langle \mathrm{e}^{\mathrm{i}\hat{h}_{\mathrm{D}}t/\hbar}\,\mathrm{e}^{-\mathrm{i}\hat{h}_{\mathrm{A}}t/\hbar}\rangle_{\mathrm{D}}=\sum_{i_{\mathrm{D}}}P_{i_{\mathrm{D}}}\langle i_{\mathrm{D}}\,|\,\mathrm{e}^{\mathrm{i}\hat{h}_{\mathrm{D}}t/\hbar}\,\mathrm{e}^{-\mathrm{i}/\hat{h}_{\mathrm{A}}t/\hbar}\,|\,i_{\mathrm{D}}\rangle \tag{13}$$

where $\hat{h}_{\mathrm{D}}=\sum_{i_{\mathrm{D}}}|i_{\mathrm{D}}\rangle\langle i_{\mathrm{D}}|\varepsilon_{i_{\mathrm{D}}}$ and $\hat{h}_{\mathrm{A}}=\sum_{i_{\mathrm{D}}}|f_{\mathrm{A}}\rangle\langle f_{\mathrm{A}}|\varepsilon_{f_{\mathrm{A}}}$ are the vibrational Hamiltonians of the *D* and *A* diabatic energy surfaces (Fig. 18.1). The thermal average is over the initial (*D*) vibrational states.

The correlation function $C_{T_{\mathrm{DA}}}(t)$ is a thermal average over bridge vibrational states, given by

$$C_{T_{\mathrm{DA}}}(t)=\langle\hat{T}_{\mathrm{DA}}(t)\,\hat{T}_{\mathrm{DA}}(0)\rangle_{\mathrm{B}}=\sum_{i_{\mathrm{D}}}P_{i_{\mathrm{B}}}\langle i_{\mathrm{B}}|\mathrm{e}^{\mathrm{i}\hat{h}_{\mathrm{B}}t/h}\hat{T}_{\mathrm{DA}}(0)\mathrm{e}^{-\mathrm{i}\hat{h}_{\mathrm{B}}t/h}\,\hat{T}_{\mathrm{DA}}(0)/i_{\mathrm{B}}\rangle \tag{14}$$

where $\hat{h}_{\mathrm{B}}=\sum_{i_{\mathrm{D}}}|i_{\mathrm{B}}\rangle\langle i_{\mathrm{B}}|\varepsilon_{i_{\mathrm{B}}}$ is the bridge vibrational Hamiltonian and $\hat{T}$ is the general (electronic-vibrational) tunneling matrix element operator.

In the semi-classical limit, the multidimensional nuclear wave functions of the accepting modes are approximated by atom-centered Gaussian wave packets moving along classical nuclear trajectories. $C_{\mathrm{FC}}(t)$ is the classical thermal average over initial atomic positions and momenta of the time dependent overlap between an initial nuclear wave packet propagated with the electron in the *D* state and the same wave packet propagated with the electron in the *A* state [45,46], i.e.,

$$C_{\mathrm{FC}}(t)=\langle\langle\psi^{\mathrm{D}}(t)|\psi^{\mathrm{A}}(t)\rangle\rangle \tag{15}$$

where the outer bracket denotes a thermal average (Fig. 18.3). In the semi-classical limit the matrix element correlation function is approximated by

$$C_{\mathrm{T_{DA}}}(t)=\langle T_{\mathrm{DA}}(t)\ T_{\mathrm{DA}}(0)\rangle \tag{16}$$

where the $T_{\mathrm{DA}}(t)$ time series is the purely electronic tunneling matrix element (Eqs (3) and (11)), computed for successive bridge conformations along classical bridge atom trajectories $\vec{r}_i(t)$, i.e., $T_{\mathrm{DA}}(t)=T_{\mathrm{DA}}[\{\vec{r}_1(t),....\vec{r}_N(t)\}]$. $\langle..\rangle$ in Eq 16 denotes a classical thermal average over bridge conformations. Substitution of Eqs (15) and (16) into Eq. (12) leads to the time dependent semi-classical expression for the ET rate.

18.7 INFLUENCE OF TUNNELING MATRIX ELEMENT FLUCTUATIONS ON THE ET RATE

Exploring three parameters facilitates a description of bridge fluctuation effects on the rate: the Franck–Condon time, the coherence parameter, and the coherence time. The Franck-Condon time arises from the time dependence of the energy gap between *D* and *A* states. The other two parameters are related to the *D*–*A* electronic-coupling dynamics.

The Franck–Condon time τ_{FC} is the time it takes for the system to move away from the crossing point of the *D* and *A* energy surfaces (i.e., the time to lose initial to final state resonance, see Fig. 18.2). For temperature *T* and reorganization energy, λ, τ_{FC} is given approximately by [74]

$$\tau_{FC} \approx \hbar/\sigma_{\Delta u} = \hbar/\sqrt{\lambda K_B T} \tag{17}$$

where $\sigma_{\Delta u} = \sqrt{\lambda K_B T}$ is the rms fluctuation in $\Delta U = U_A - U_D$ (Fig. 18.3). Semi-classically, τ_{FC} is related to the time evolution of the Franck-Condon factor in Eq. (15). τ_{FC} is the average decay time of the overlap between nuclear wave packets $\psi^D(t)$ and $\psi^A(t)$ that are initially placed at the crossing point between *D* and *A* surfaces. The overlap decays as the wave packets move away from the crossing point (Fig. 18.3).

The coherence parameter [47] of the tunneling matrix element is a measure of the magnitude of T_{DA} fluctuations:

$$R_{coh} = \langle T_{DA} \rangle^2 / \langle T_{DA}{}^2 \rangle = \langle T_{DA} \rangle^2 / \left(\langle T_{DA} \rangle^2 + \sigma^2_{T_{DA}} \right). \tag{18}$$

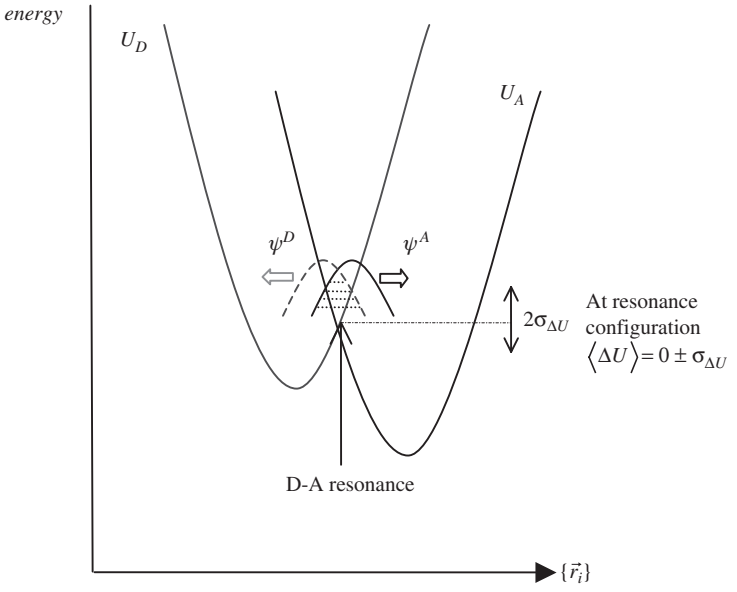


Fig. 18.3. Schematic view of nuclear wave packet motion away from the crossing point of the donor and acceptor potential surfaces.

Here, $\langle..\rangle$ denotes a classical thermal average and $\sigma^2_{T_{DA}}$ is the rms fluctuation in T_{DA}. When the rms T_{DA} fluctuations are large with respect to the average T_{DA}, $\sigma_{T_{DA}} >> \langle T_{DA}\rangle$ in Eq. (18) and $R_{coh} \approx 0$. In the limit of small-rms T_{DA} fluctuations compared with the average, $R_{coh} \approx 1$.

The final parameter is the matrix element coherence time τ_{coh}, the decay time for $C_{T_{DA}}$ (Eq. (16)). τ_{coh} is a measure of how rapidly fluctuations of the tunneling matrix element randomize its value (see Fig. 18.4), i.e.,

$$C_{T_{DA}}(t) \approx \sigma^2_{T_{DA}} e^{-t/\tau_{coh}} + \langle T_{DA}\rangle^2 \tag{19}$$

18.8 QUALITATIVE PICTURE OF TUNNELING MATRIX ELEMENT FLUCTUATION EFFECTS ON THE ET RATE

The parameters defined above can be used to distinguish between distinct ET rate regimes.

Regime I: $\tau_{coh} > \tau_{FC}$ ('slow' coupling matrix element fluctuations). In this regime the tunneling matrix element does not have time to fluctuate while the *D* and *A* electronic states are in resonance. As a result, during each *D–A* energy crossing, *D* and *A* see a static coupling, equal to the value of $T_{DA}(t)$ at the time of the crossing. The rate in this case is proportional to the Franck–Condon weighted density of states, i.e.,

$$k_{ET} \propto \rho_{FC.} \tag{20}$$

In this regime, we distinguish between two limits. The most interesting is the limit of large T_{DA} fluctuations compared with the average T_{DA}, i.e., $\sigma_{T_{DA}} >> \langle T_{DA}\rangle$ ($R_{coh} \approx 0$). In this situation, only fluctuations in T_{DA} can cause significant couplings between *D* and *A*,

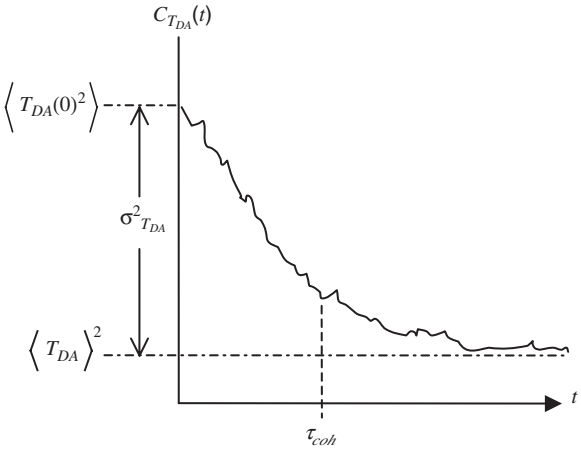


Fig. 18.4. Decay of the donor-acceptor coupling correlation function.

and thus fluctuations determine the ET rate. As such, T_{DA}^2 in Eq. (5) should be replaced by $<T_{DA}^2>$,

$$k_{ET} \propto \langle T_{DA}^2 \rangle \rho_{FC} \tag{21}$$

In the opposite limit of small T_{DA} fluctuations compared with the average, ($R_{coh} \approx 1$), the T_{DA} fluctuations are irrelevant. That is, $<T_{DA}^2> \approx <T_{DA}>^2$, and the rate becomes

$$k_{ET} \propto \langle T_{DA} \rangle^2 \rho_{FC.} \tag{22}$$

In both of the above regimes, an implicit assumption is that $\tau_{coh} << 1/k_{ET}$ [48], so that the electron probability on D shows single exponential decay.

Regime II: $\tau_{coh} \sim \tau_{FC}$ ('fast' matrix element fluctuations). Here, the tunneling matrix element has time to fluctuate while the D and A states are resonant. For each D–A crossing event, D and A see a fluctuating T_{DA}. Here, the Franck–Condon approximation is not valid for every crossing. The rate is more complex than is suggested by Eq. (21) because there is no separation between the relevant time scales (that is, k_{ET} in Eq. (12) does not simplify because the decay times of $C_{FC}(t)$ and $C_{T_{DA}}(t)$ are similar). If $\tau_{coh} \geqslant 1/k_{ET}$, the ET probability does not decay as a single exponential, and the effective rate constant is inhomogeneously broadened (time dependent) [48]. We assume the $\tau_{coh} << 1/k_{ET}$ limit in the following.

To explore how the matrix-element fluctuations generate deviations from the slow-fluctuation regime ($k \propto \rho_{FC}$), Eq. (12) can be expanded in powers of (τ_{FC}/τ_{coh}) [13]. To lowest order in (τ_{FC}/τ_{coh})

$$k_{ET} \approx \langle T_{DA}^2 \rangle \rho_{FC}[1 + \delta k^{dyn}] \tag{23}$$

where $\rho_{FC} = (4\pi\lambda K_B T)^{-1/2} e^{-([U_D^{min} - U_A^{min}] - \lambda)^2/4\lambda K_B T}$ is the classical Franck–Condon factor (the nuclear motion was treated classically), and δk^{dyn} is the fractional dynamical correction to the slow fluctuation rate. The correction depends on the three parameters described above,

$$\delta k^{dyn} \approx \frac{1}{8}\left\{ \frac{(\lambda + (U_A^{min} - U_D^{min}))^2 - 2(\hbar/\tau_{FC})^2}{(\hbar/\tau_{FC})^4} \right\} \times \left(\frac{\hbar}{\tau_{coh}} \right)^2 \times (1 - R_{coh}). \tag{24}$$

The dynamical correction disappears when tunneling matrix element fluctuations are small, i.e., when $R_{coh} \approx 1$. In this case, $<T_{DA}^2> \approx <T_{DA}>^2$ and we recover Eq. (22). When the fluctuations are large but slow, the rate is given by Eq (21) ($R_{coh} \approx 0$, and $(\tau_{FC}/\tau_{coh}) << 1$). As the matrix element fluctuations become more rapid, the rate deviates from Eq. (21). For example, in the case of activationless ET, ($U_A^{min} - U_D^{min} = -\lambda$), the fractional deviation is

$$\delta k^{dyn} \approx -\frac{1}{4}\left\{ \left(\frac{\tau_{FC}}{\tau_{coh}} \right)^2 \times (1 - R_{coh}) \right\} < 0 \tag{25}$$

The correction is negative, showing a reduction of the rate compared with the slow fluctuation limit. This slowing occurs because, for activationless ET, the crossing point between *D* and *A* energy surfaces is at the minimum of the *D* energy surface. As such, at equilibrium, the initial state of the system is in resonance with the final state, and the effect of T_{DA} fluctuations is to move the system out of resonance.

Other analytical studies on the influence of matrix element fluctuations on ET rates demonstrate the existence of different kinetic regimes [13,49–58]. As mentioned above, for slow fluctuations ($\tau_{coh} > \tau_{FC}$) the rate is Marcus–like ($k_{ET} \propto \rho_{FC}$). For fast fluctuations ($\tau_{coh} < \tau_{FC}$) and activationless (or nearly activationless) ET ($\tau_{coh} < \hbar/\left|(U_D^{min} - U_A^{min}) - \lambda\right|$), the rate decreases with decreasing τ_{coh}. For fast fluctuations and large activation energies ($\tau_{coh} < \hbar/\left|(U_D^{min} - U_A^{min}) - \lambda\right|$), the fluctuations can enhance the rate because they reduce the activation energy needed to establish resonance between initial and final states. In general, this rate enhancement is much more pronounced in the inverted region than in the normal region, and its effect is to reduce the drop in rate with increasing energy gap (and to change the temperature dependence of the rate). A temperature dependent enhancement (gating) of the rate can be observed for systems where the coupling is very weak for equilibrium conformations. An increase in the temperature can access bridge conformations with enhanced coupling, thus increasing the rate. Experimentally the anomalous (non-Marcus) inverted region has been observed in small molecule ET [59], and in biological ET [60]. In both cases, it is not clear if this behavior arises from bridge-fluctuation effects.

There have been several studies of ET reactions that explore the time dependence of the tunneling matrix element arising from bridge fluctuations [61–72]. These studies aim to identify bridge conformations that enhance the *D–A* coupling and to compute $\langle T_{DA} \rangle$ and/or $\langle T_{DA}^2 \rangle$. Three studies [73–75] characterize known ET reactions in terms of the different rate regimes. The analysis in [73,74] computes R_{coh}, τ_{coh}, and δk^{dyn} in order to understand how these parameters relate to bridge structure and dynamics. Comparison of the results in [73] and [74] is particularly interesting because [73] explores short–range solvent-mediated ET in small C-clamped molecules, whereas [74] explores long-distance ET in the protein azurin. Both studies compute, R_{coh}, τ_{coh}, and δk^{dyn} using MD simulations and semi-empirical electronic structure analysis via Eqs (11), (18), (19) and (24). Surprisingly, the values of R_{coh}, τ_{coh} and δk^{dyn} are quite similar for the two types of ET reactions. For the C-clamped molecules, $R_{coh} \leqslant 0.1$ $\tau_{coh} \approx 0.1$ps and $\delta k^{dyn} \approx 0.04\%$. For azurin, $R_{coh} \leqslant 0.1$, $\tau_{coh} \approx$ 30–40fs and $\delta k^{dyn} \approx 0.8\%$ The value of R_{coh} indicates that in both ET systems the tunneling matrix element fluctuations are large. However, these fluctuations are 'slow' because, in both cases, τ_{coh} is much greater than $\tau_{FC} \approx \hbar/\sqrt{\lambda K_B T}$. For azurin, $\tau_{FC} \approx$3–4fs, as computed using MD simulations and the semi-classical approximation for $C_{FC}(t)$ in Eq. (15) [76]. The fact that τ_{coh} is at least an order of magnitude larger than τ_{FC} in the two systems leads to very small δk^{dyn}. Therefore, the ET rates are adequately described by Eq. (21) as long as the MD simulations can access those bridge conformations that enhance the *D–A* coupling and thus describe $\langle T_{DA}^2 \rangle$ well. Ref. [75] computes R_{coh} and τ_{coh} for the ET reaction $BPh^- \rightarrow Q_A$ in the bacterial photosynthetic reaction center. This study also computes the non-Condon (dynamical) correction to the ET rate as a function of energy gap and of τ_{coh}, (where τ_{coh} is treated as a parameter), using a form of $C_{T_{DA}}(t)$ that preserves detailed balance. In [75], it is found that $R_{coh} \approx 0.4$ and

$\tau_{coh} \approx$ 60fs for ET from BPh^- to Q_A. The value of R_{coh} indicates that the tunneling matrix element for this ET reaction fluctuates less than it does for C-clamped-molecules and for azurin ET. However, τ_{coh} is of the same order of magnitude as in the other ET reactions. Also, the dynamical correction to the rate is small, as in the other ET reactions. A large dynamical correction is observed only if the system is driven to the highly activated normal and inverted regions.

In the analysis of azurin in Ref. [74], R_{coh} and τ_{coh} are analyzed as a function of temperature, *D–A* distance, tunneling energy, and constraints on bridge motion. It is found that these parameters change little with changes in *D–A* distance or coupling pathway structure. R_{coh} increases as temperature is lowered (reduced fluctuations) but, surprisingly, there is no significant temperature effect on τ_{coh}. Comparison between the results of normal and constrained MD simulations shows that R_{coh} is controlled mainly by dihedral angle fluctuations and global motion, whereas τ_{coh} is largely driven by high-frequency valence angle oscillations. The dependence of τ_{coh} on localized stiff angle motion explains why τ_{coh} does not change with *D–A* distance and with temperature. Further, the analysis suggests that the similarity in the τ_{coh} values for the C-clamped molecules for azurin and for the bacterial reaction center, arises from the similarity of the rapid motion in the different bridges (i.e., rapid structural fluctuations are somewhat 'transferable').

Another notable result of Ref. [74] is that R_{coh} and τ_{coh} do not depend on the tunneling energy (Eqs (3, 4)) for tunneling energies well separated from the bridge electronic state energies. However, when E_{tun} is brought close to bridge resonances, tunneling matrix element fluctations are enhanced and τ_{coh} drops by an order of magnitude to a few fs. Although this shallow tunneling limit is probably not relevant to ET in azurin, it does indicate a general trend. For small *D*(*A*) bridge energy gaps, the tunneling matrix element is very sensitive to bridge fluctuations and it is possible to have $\tau_{coh} \approx \tau_{FC}$. As such, significant dynamical non-Condon effects (δk^{dyn}) may arise in this small gap regime. This situation may also arise for electronically forbidden ET ($\langle T_{DA} \rangle = 0$), e.g., [77–82], where molecular motion lowers the symmetry and leads to large changes in the tunneling matrix element as a function of time. In these cases, computation of the *D–A* coupling as a function of time cannot be based solely on splittings (matrix elements) calculated for frozen bridge snapshot conformations. Suppose that there is a *D–A* surface crossing around time *t* that lasts for a time τ_{FC}. If, during this period, the tunneling matrix element fluctuates, the value that should be associated with the *D–A* coupling at that particular crossing is not the static coupling of the conformation at time *t*. Rather, vibronic non-Condon terms [55,83–87] should be included that contain the coupling derivatives,

$$T_{DA}(t) \rightarrow T_{DA}(t) + \frac{dT_{DA}(t)}{dt} \times \tau_{FC} \ \ldots \text{ etc.} \tag{26}$$

(or equivalently derivatives with respect to bridge atom positions $\vec{\nabla}_{r_i} T_{DA}[\{\vec{r}_1, \ldots \vec{r}_N\}]$).

In all of the above discussion of non-Condon effects, the Born–Oppenheimer approximation was used. However, it is possible that bridge motion leads to a failure of the Born–Oppenheimer approximation (in addition to failure of the Franck–Condon

approximation). As in the case of non-Condon effects discussed thus far, non-Born-Oppenheimer transitions can influence the distance and energy gap dependence of the rate [88–91]. Further, molecular motion that causes rapid change in the *D–A* wave functions (e.g., as in symmetry-forbidden ET) may produce a non-Born–Oppenheimer *D–A* coupling that is dynamic in nature. In Refs [89,90] it is shown that, to lowest order in the nonadiabatic derivative couplings,[1] the non-Born–Oppenheimer contribution to the *D–A* tunneling matrix element is

$$T_{\mathrm{DA}}^{\mathrm{NBO}}(t)=\hbar^2 \sum_{i,j\neq \mathrm{D,A}} \left(\frac{\mathrm{d}V_{\mathrm{D},i}}{\mathrm{d}t}\right)\frac{\mathrm{d}^2 G_{i,j}^{(\mathrm{B})}(E_{\mathrm{tun}})}{\mathrm{d}E_{\mathrm{tun}}^{\,2}}\left(\frac{\mathrm{d}V_{j,\mathrm{A}}}{\mathrm{d}t}\right). \tag{27}$$

Here, i, j index the bridge orbitals, $\mathrm{G}_{i,j}^{(\mathrm{B})}(E_{\mathrm{tun}})$ are bridge Green's function matrix elements, and $V_{\mathrm{D},i}$ ($V_{j,\mathrm{A}}$) are the direct coupling between the D (A) electronic states and the bridge orbitals (all quantities are time dependent). $T_{\mathrm{DA}}^{\mathrm{NBO}}(t)$ has the same form as the time dependent superexchange matrix element $T_{\mathrm{DA}}(t)=T_{\mathrm{DA}}^{\mathrm{su}}(t)=\sum_{i,j\neq \mathrm{D,A}} V_{\mathrm{D},i}\, G_{i,j}^{(\mathrm{B})}(E_{\mathrm{tun}}) V_{j,\mathrm{A}}$ (Eqs (3,4 & 11)) with $G_{i,j}^{(\mathrm{B})}(E_{\mathrm{tun}})\rightarrow\hbar^2 \mathrm{d}^2 G_{i,j}^{(\mathrm{B})}(E_{\mathrm{tun}})/\mathrm{d}^2 E_{\mathrm{tun}}$, $V_{\mathrm{D},i}\rightarrow \mathrm{d}V_{\mathrm{D},i}/\mathrm{d}t$, and $\mathrm{V}_{j,\mathrm{A}}\rightarrow \mathrm{d}V_{j,\mathrm{A}}/\mathrm{d}t$. $T_{\mathrm{DA}}^{\mathrm{NBO}}$ thus depends on the rates of change of $V_{\mathrm{D},i}$ and $V_{j,\mathrm{A}}$ with time rather than on the magnitudes of $V_{\mathrm{D},i}$ and $V_{j,\mathrm{A}}$. The relative magnitude of $T_{\mathrm{DA}}^{\mathrm{NBO}}$ compared to $T_{\mathrm{DA}}^{\mathrm{SU}}$ is of the order

$$\max\left\{\left|\frac{T_{\mathrm{DA}}^{\mathrm{NBO}}}{T_{\mathrm{DA}}^{\mathrm{su}}}\right|\right\}\sim\left[\left(\frac{\hbar\omega}{\Delta E}\right)\times\left(\frac{\delta V}{V}\right)\right]^2 \tag{28}$$

ΔE is the $D(A)$ to bridge energy gap, and V, δV, and ω are the mean value, fluctuation magnitude, and fluctuation frequency, respectively, of the $D(A)$ to bridge electronic couplings ($V_{\mathrm{D},i}$ and $V_{j,\mathrm{A}}$). If $\delta V/V >> 1$ the above ratio can be significant even for $\hbar\omega < \Delta E$. $\delta V/V >> 1$ for symmetry-forbidden ET, where at the equilibrium geometry $V_{\mathrm{D},i}$ and $V_{j,\mathrm{A}}$ are small but they switch to large values because of relatively high-frequency vibrational or torsional motion. This regime may arise for low gap (π-electron) ET systems, such as those with OPV bridges [81,82], and for DNA ET [92].

18.9 QUANTUM NUCLEAR MOTION AND INELASTIC TUNNELING

So far, our discussion of bridge fluctuations has assumed that bridge atoms move classically, and that the tunneling matrix element depends parametrically on the bridge atom trajectories $\{\vec{r}_i(t)\}$, i.e., $T_{\mathrm{DA}}(t) = T_{\mathrm{DA}}[\vec{r}_1(t),....\vec{r}_N(t)\}]$. In this section, we show how to treat the case of quantum bridge vibrational degrees of freedom and their interaction with the tunneling electron.

[1]The nonadiabatic derivative couplings between instantaneous system eigenstates are given by $\langle\Psi_m(t)\,|\,\frac{\mathrm{d}}{\mathrm{d}t}\,\Psi_n(t)\rangle=\sum_i\langle\Psi_m(\vec{r}_1,..\vec{r}_N)|\vec{\nabla}_i\Psi_n(\vec{r}_1,..\vec{r}_N)\rangle.\,\frac{\mathrm{d}}{\mathrm{d}t}\,\vec{r}_i(t)$, where $\vec{r}_i$ denotes atomic position and $|\Psi_n\rangle$ and $|\Psi_m\rangle$ are system eigenstates.

From Eq. (1), the nonadiabatic rate is [13,55,86]

$$k_{ET}=\sum_{i_B}P_{i_B}\sum_{f_B}k_{ET}\{i_B\to f_B\} \tag{29}$$

where

$$k_{ET}\{i_B\to f_B\}=\frac{2\pi}{\hbar}|\langle D;i_B|\hat{T}|A;f_B\rangle|^2\times\rho_{FC}\{i_B\to f_B\} \tag{30}$$

and

$$\rho_{FC}\{i_B\to f_B\}=\sum_{i_D}P_{i_D}\sum_{f_A}|\langle i_D|f_A\rangle|^2\delta[U_D^{min}+\varepsilon_{i_D}+\varepsilon_{i_B}-(U_A^{min}+\varepsilon_{f_A}+\varepsilon_{f_B})]. \tag{31}$$

The total rate in Eq. (29) is a thermally weighted sum of individual rates $k_{ET}\{i_B\to f_B\}$, each associated with a process where an initial bridge vibrational state i_B moves to a final state f_B during the D to A transition. The thermal weights P_{i_B} are Boltzmann populations of initial bridge vibrational states. $k_{ET}\{i_B\to f_B\}$ (Eq. (30)) contains a thermally weighted Franck-Condon factor $\rho_{FC}\{i_B\to f_B\}$ where the thermal average is over initial and final states of the accepting modes. As before, we assume the classical limit for the accepting modes, i.e.,

$$\rho_{FC}\{i_B\to f_B\}=(4\pi\lambda K_BT)^{-1/2}\exp[-([U_D^{min}+\varepsilon_{i_B}-(U_A^{min}+\varepsilon_{f_B})]-\lambda)^2/4\lambda K_BT] \tag{32}$$

$\langle D;i_B|\hat{T}|A;f_B\rangle$ in Eq. (30) is the bridge-mediated tunneling matrix element, a function of the initial and final bridge vibrational states.

An electron tunneling event where the initial and final bridge vibrational states have the same energy ($\varepsilon_{i_B}=\varepsilon_{f_B}$) is elastic tunneling. It is inelastic when the energies are different. For $\varepsilon_{i_B}>\varepsilon_{f_B}$ the electron gains energy from the bridge vibrations, and for $\varepsilon_{i_B}<\varepsilon_{f_B}$ the electron loses energy to the bridge. Eq. (32) shows that the energy gap for elastic tunneling is $U_D^{min}-U_A^{min}$, regardless of the initial (final) bridge vibrational state. For inelastic tunneling, the energy gap is shifted by an amount equal to the overall change in the bridge vibrational energy. That is, given by $U_D^{min}+\varepsilon_{i_B}-(U_A^{min}+\varepsilon_{f_B})=U_D^{min}-U_A^{min}+(\varepsilon_{i_B}-\varepsilon_{f_B})$. An equivalent viewpoint is that inelastic tunneling changes (blurs) the crossing point between the donor and acceptor energy surfaces. For elastic ET, the molecular configuration that gives initial to final state resonance is the collection of atomic positions $\vec{r}_i$ for which $U_D(\{\vec{r}_i\})=U_A(\{\vec{r}_i\})$ (Fig. 18.2). For the inelastic process, the atomic positions are different because the resonance condition is $U_D(\{\vec{r}_i\})+\varepsilon_{i_B}=U_A(\{\vec{r}_i\})+\varepsilon_{f_B}$.

Suppose that an elastic ET reaction is strongly in the inverted regime, with $U_D^{min}=U_A^{min}+2\lambda$. The elastic ET rates are suppressed because the activation energy $E_{act}=(U_D^{min}-U_A^{min}-\lambda)^2/4\lambda=\lambda/4$ (Fig. 18.2) is greater than K_BT. There may be bridge vibrations that couple to the electron and that have frequency $\hbar\omega\approx\lambda/n$ (n being an integer). In this case inelastic tunneling is possible because the electron can deposit energy equal to λ to any initial state of these vibrations $|i_B\rangle$, and excite it to a final state $|f_B\rangle$ with $\varepsilon_{f_B}-\varepsilon_{i_B}=n\hbar\omega=\lambda$. The corresponding rate is activationless because its energy gap is reduced by λ, i.e., $E_{act}=([U_D^{min}+\varepsilon_{iB}-(U_A^{min}+\varepsilon_{fB})]-\lambda)^2/4\lambda=0$. Therefore, in the inverted region inelastic tunneling rates are enhanced, and the overall rate in the sum of

Eq. (29) does not drop by the amount expected for a pure elastic tunneling mechanism. This simple argument explains the earlier statement that bridge fluctuations may lower the activation energy of an ET reaction, thus reducing the drop of the rate in the highly inverted region. The effect is stronger for large bridges where the bridge vibrational spectrum is very dense and, for any value of the energy gap, there are vibrational states that can absorb an equal amount of electronic energy. In the activated normal region, a similar argument can be made where the electron absorbs vibrational energy from the bridge, thus reducing the activation energy. The addition of more inelastic channels (a denser bridge vibrational spectrum) with increasing bridge size has led to the suggestion that the drop in the total inelastic rate with distance will be slower than the drop of the elastic rate [55].

We now explore the tunneling matrix element $\langle D;i_B|\hat{T}|A;f_B\rangle$ that enters the rate $k_{ET}\{i_B\rightarrow f_B\}$ (Eq. (30)). We use a simple example of a bridge containing a single electronic state $|B\rangle$ that interacts with a single bridge vibration (oscillator coordinate y). When the electron is not located on the bridge (i.e., state $|B\rangle$ is not occupied), the equilibrium length of the oscillator is y^0. When the electron occupies state $|B\rangle$, the oscillator feels a constant force F along y that shifts its equilibrium length by δy^0. The eigenstates of the oscillator in the absence of the electron in the bridge are harmonic oscillator eigenstates $|n(y)\rangle$ with energies $(n+1/2)\hbar\omega$. When the electron is on the bridge, the bridge vibrational states are harmonic oscillator eigenstates centered at δy^0, i.e., $|n'(y-\delta y^0)\rangle$, with energies $(n'+1/2)\hbar\omega-(F^2/2m\omega^2)$ (the system is stabilized by electron-oscillator interactions). The Hamiltonian of the bridge is:

$$\hat{H}^{(B)}=\underbrace{U_B^{min}|B\rangle\langle B|}_{el}+\underbrace{\sum(n+1/2)\hbar\omega|n(y)\rangle\langle n(y)|}_{nuc}\underbrace{-F\hat{y}|B\rangle\langle B|}_{el-nuc} \tag{33}$$

This operator contains the electronic and nuclear degrees-of-freedom of the bridge, and their interaction. In the displaced oscillator ($|n'(y-\delta y^0)\rangle$) representation

$$\hat{H}^{(B)}=U_B^{min}|B\rangle\langle B|+\sum_{n'}[(n'+1/2)\hbar\omega-(F^2/2m\omega^2)]|n'(y-\delta y^0)\rangle\langle n'(y-\delta y^0)|. \tag{34}$$

As before, we assume classical accepting modes $\{\vec{r}_i\}$ (i.e., the Franck–Condon factors are given by Eq. (32)). The total Hamiltonian is $\hat{H}=\hat{H}^{(DA)}+\hat{H}^{(B)}+\hat{V}$, where

$$\hat{H}^{(DA)}=U_D(\{\vec{r}_i\})|D\rangle\langle D|+U_A(\{\vec{r}_i\})|A\rangle\langle A| \text{ and } \hat{V}=V_{D,B}|D\rangle\langle B|+V_{A,B}|A\rangle\langle B|+h.c. \tag{35}$$

are the electronic Hamiltonians of the D and A electronic states, and of the $D(A)$-bridge interactions, respectively. U_D,(U_A) are the $D(A)$ Born–Oppenheimer surfaces.

The initial and final bridge vibrational states in $\langle D;i_B|\hat{T}|A;f_B\rangle$ are eigenstates of the unperturbed bridge oscillator, i.e., $|i_B\rangle=|n(y)\rangle$, $|f_B\rangle=|m(y)\rangle$ (since the electron is not localized in the bridge before and after ET). The tunneling matrix element operator is $\hat{T}_{DA}=\hat{V}\hat{G}^{(B)}(E_{tun})\hat{V}$, as in Eq. (3), where $\hat{G}^{(B)}(E_{tun})=[E_{tun}-\hat{H}^{(B)}]^{-1}$. In contrast to Eq. (4),

the bridge Green's function is not purely electronic. It contains the bridge Hamiltonian of Eqs (33), (34). The tunneling energy is the energy of initial and final state resonance:

$$E_{\text{tun}} = U_{\text{D}}(\{\vec{r}_i\}) + \varepsilon_{i_{\text{B}}} = U_A(\{\vec{r}_i\}) + \varepsilon_{f_{\text{B}}} \tag{36}$$

The bridge Green's function is written in the standard way as a sum over bridge eigenstates [86], i.e.,

$$\hat{G}^{(\text{B})}(E_{\text{tun}}) = \sum_{n'} \frac{|B;n'(y-\delta y^0)\rangle\langle B;n'(y-\delta y^0)|}{E_{\text{tun}} - [U_{\text{B}}^{\min} + (n'+1/2)\hbar\omega - (F^2/2m\omega^2)]}, \tag{37}$$

where $|B;n'(y-\delta y^0)\rangle = |B\rangle|n'(y-\delta y^0)\rangle$ is an electronic-vibrational bridge eigenstate with eigenenergy $U_{\text{B}}^{\min} + (n'+1/2)\hbar\omega - (F^2/2m\omega^2)$. Therefore, the inelastic tunneling matrix element becomes

$$\langle D;i_{\text{B}}=n|\hat{T}|A;f_{\text{B}}=m\rangle = V_{\text{D,B}}V_{\text{B,A}} \sum_{n'} \frac{\langle n(y)|n'(y-\delta y^0)\rangle\langle n'(y-\delta y^0)|m(y)\rangle}{E_{\text{tun}} - [U_{\text{B}}^{\min} + (n'+1/2)\hbar\omega - (F^2/2m\omega^2)]} \tag{38}$$

where $E_{\text{tun}} = U_{\text{D}}(\{\overset{\text{v}}{r}_i\}) + (n+1/2)\hbar\omega = U_{\text{A}}(\{\overset{\text{v}}{r}_i\}) + (m+1/2)\hbar\omega$, $(n \neq m)$. For each term in the sum, the numerator is the product $\sqrt{FC_{n,n'}}\sqrt{FC_{n',m}}$, where $FC_{n,n'}$ ($\text{FC}_{n',m}$) is the Franck-Condon factor between the initial (final) eigenstate of the unperturbed bridge oscillator and an eigenstate of the perturbed (displaced) bridge oscillator. The resonance condition Eq. (36) is realized when $U_{\text{D}}(\{\overset{\text{v}}{r}_i\}) - U_{\text{A}}(\{\overset{\text{v}}{r}_i\}) = (m-n)\hbar\omega$.

For elastic ET, the matrix element is

$$\langle D;i_{\text{B}}=n|\hat{T}|A;f_{\text{B}}=n\rangle = V_{\text{D,B}}V_{\text{B,A}} \sum_{n'} \frac{|\langle n(y)|n'(y-\delta y^0)\rangle|^2}{E_{\text{tun}} - [U_{\text{B}}^{\min} + (n'+1/2)\hbar\omega - (F^2/2m\omega^2)]} \tag{39}$$

where $E_{\text{tun}} = U_{\text{D}}(\{\overset{\text{v}}{r}_i\}) + (n+1/2)\hbar\omega = U_{\text{A}}(\{\overset{\text{v}}{r}_i\}) + (n+1/2)\hbar\omega$ (resonance condition $U_{\text{D}}(\{\overset{\text{v}}{r}_i\}) = U_{\text{A}}(\{\overset{\text{v}}{r}_i\})$).

Equations (38) and (39) show that the tunneling matrix elements differ for elastic and inelastic tunneling, and that they depend on Franck–Condon factors of the bridge oscillators that couple to the tunneling electron. Therefore, a tunneling pathway may be forbidden because the Franck–Condon factors of the vibrations to which it couples are small.

This analysis can be used to understand the influence of the bridge vibrations on the interference among coupling pathways [86]. Consider a bridge with two electronic states $|B1\rangle$ and $|B2\rangle$, each interacting with a different bridge oscillator (Fig. 18.5).

The case of two bridge modes produces a total vibrational wavefunction that is a product of vibrational states of the two oscillators, i.e., $|i_{\text{B}}\rangle = |n(y_1)\rangle|m(y_2)\rangle$ with

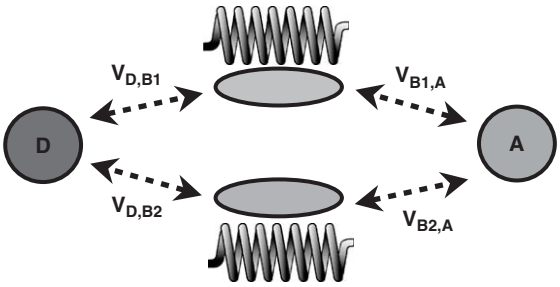


Fig. 18.5. Sketch of two parallel tunneling pathways each coupled to a local oscillator.

$\varepsilon_{i_B} = (n+1/2)\hbar\omega_1 + (m+1/2)\hbar\omega_2$, and $|f_B\rangle = |p(y_1)\rangle\,|q(y_2)\rangle$ with $\varepsilon_{f_B} = (p+1/2)\hbar\omega_1 + (q+1/2)\,\hbar\omega_2$. The bridge Green's function is

$$\hat{G}^{(B)}(E_{tun}) = \sum_{n',m} \frac{|B1; n'(y_1 - \delta y_1^0) m(y_2)\rangle\langle B1; n'(y_1 - \delta y_1^0) m(y_2)|}{E_{tun} - [U_{B1}^{min} + (n'+1/2)\hbar\omega_1 - (F_1^2/2m_1\omega_1^2) + (m+1/2)\hbar\omega_2]}$$

$$+ \sum_{n',m'} \frac{|B2; n(y_1) m'(y_2 - \delta y_2^0)\rangle\langle B2; n(y_1) m'(y_2 - \delta y_2^0)|}{E_{tun} - [U_{B2}^{min} + (n+1/2)\hbar\omega_1 + (m'+1/2)\hbar\omega_2 - (F_2^2/2m_2\omega_2^2)]} \quad . \qquad (40)$$

The first term describes tunneling through the $B1$ pathway where the electron interacts only with oscillator 1 (the second term is for the $B2$ pathway). The states $|n'(y - \delta y_1^0)\rangle$ in the first term are the eigenstates of displaced oscillator 1, with eigenenergies $(n'+1/2)\hbar\omega_1 - (F_1^2/2m_1\omega_1^2)$ (F_1 is the force on the oscillator when the electron occupies state $|B1\rangle$). The second term contains the eigenstates $|m'(y - \delta y_2^0)\rangle$ and energies of the oscillator that interacts with the electron on $|B2\rangle$.

The Green's function above suggests that it is possible to select pathways, or equivalently to erase interferences between pathways, by depositing vibrations in one of the two local oscillators [86]. For example, consider the elastic tunneling case where the initial (final) bridge vibrational state is the ground state $|i_B\rangle = |f_B\rangle = |0(y_1)\rangle|0(y_2)\rangle$. The corresponding tunneling matrix element is

$$\langle D; i_B | \hat{T} | A, f_B \rangle = V_{D,B1} V_{B1,A} \sum_{n'} \frac{|\langle 0(y_1) | n'(y_1 - \delta y_1^0)\rangle|^2}{E_{tun} - [U_{B1}^{min} + (n'+1/2)\hbar\omega_1 - (F_1^2/m_1\omega_1^2) + (\hbar\omega_2/2)]}$$

$$+ V_{D,B2} V_{B2,A} \sum_{m'} \frac{|\langle 0(y_2) | m'(y_2 - \delta y_2^0)\rangle|^2}{E_{tun} - [U_{B2}^{min} + (\hbar\omega_1/2) + (m'+1/2)\hbar\omega_2 - (F_2^2/m_2\omega_2^2)]} \qquad (41)$$

where $E_{tun} = U_D(\{\vec{r}_i\}) + \hbar\omega_1/2 + \hbar\omega_2/2 = U_A(\{\vec{r}_i\}) + \hbar\omega_1/2 + \hbar\omega_2/2$. The elastic tunneling matrix element is the sum of two terms that correspond to two coherently interfering coupling pathways ($B1$ and $B2$).

Consider the inelastic case where the initial vibrational state is the ground state $|i_B\rangle = |0(y_1)\rangle\,|0(y_2)\rangle$, but during the tunneling event the energy $p\hbar\omega_1$ is deposited in oscillator 1. The final vibrational state is $|f_B\rangle = |p(y_1)\rangle\,|0(y_2)\rangle$, and the tunneling matrix element is

$$\langle D;i_B|\hat{T}|A,f_B\rangle = V_{D,B1}V_{B1,A}\sum_{n'}\frac{\langle 0(y_1)|n'(y_1-\delta y_1^0)\rangle\langle n'(y_1-\delta y_1^0)|p(y_1)\rangle}{E_{tun}-[U_{B1}^{min}+(n'+1/2)\hbar\omega_1-(F_1^2/m_1\omega_1^2)+(\hbar\omega_2/2)]} + 0 \tag{42}$$

where $E_{tun}=U_D(\{\vec{r}_i\})+\hbar\omega_1/2+\hbar\omega_2/2=U_A(\{\vec{r}_i\})+(p+1/2)\hbar\omega_1+\hbar\omega_2/2$. The second (zeroth) term in Eq. (42) corresponds to the electron-transfer pathway through $B2$ (the second term in Eq. (40)). It is zero because it contains the combined overlap $\langle 0(y_1)|n(y_1)\rangle\langle n(y_1)|p(y_1)\rangle$ in the sum over $n(y_1)$. Therefore, the inelastic tunneling matrix element has only a contribution from electronic pathway $B1$, whose bridge vibrations were excited during ET, and there are no pathway interferences. This is a molecular analog of Young's double slit experiment. The selective vibrational excitation of a pathway labels the route traversed. For example, if elastic ET is forbidden, i.e., the electron transfer pathways destructively interfere when the bridge vibrations are in their ground states, the rate may be enhanced by vibrationally exciting a part of the molecule that contributes to one of the interfering paths [86].

18.10 TUNNELING THROUGH A BRIDGE WITH LARGE CONFORMATIONAL FREEDOM [93]

Since its invention [94,95], the scanning tunneling microscope (STM) has become a key instrument in surface science. Recently, single-molecule experiments using STM have become accessible. Several groups have employed pulling, pushing, and sliding manipulations with an STM tip [96–99]. For biological systems, Gaub *et.al*. [100] stretched a single DNA molecule with an AFM (atomic force microscope) and probed the structural changes, and Cai *et al.* [101] measured the conductivity of single DNA molecules using conducting AFM. In the last 20 years, many groups [102–110] have developed theoretical methods to simulate and analyze tunneling currents and STM images; they have been relatively successful at predicting magnitudes of currents and qualitative features of STM images. Relatively limited attention has been paid to details of molecular conformational dependence of tunneling currents. We designed a simple STM model [93] at the tight-binding level that can be used to explore the conformational dependence of tunneling currents while pulling flexible molecules. Fig. 18.6 shows the pulling process; the molecule is octane dithiol. We found that the tunneling current decays approximately exponentially with the sulfur-to-sulfur distance with a decay factor or about 1.0 $Å^{-1}$. Figure 18.7 shows the result (the predicted tunneling current is proportional to the conformational averaged electronic coupling ($<|H_{t,kS}|^2>$).

The computed value of coupling decay is consistent with a body of experimental [20–23] and theoretical [24–26] data for ET through saturated molecules. For each sulfur-to-sulfur distance, we show the strongest, weakest, and second weakest couplings from the ensemble of structures in Fig. 18.8 (the coupling versus sulfur-to-sulfur

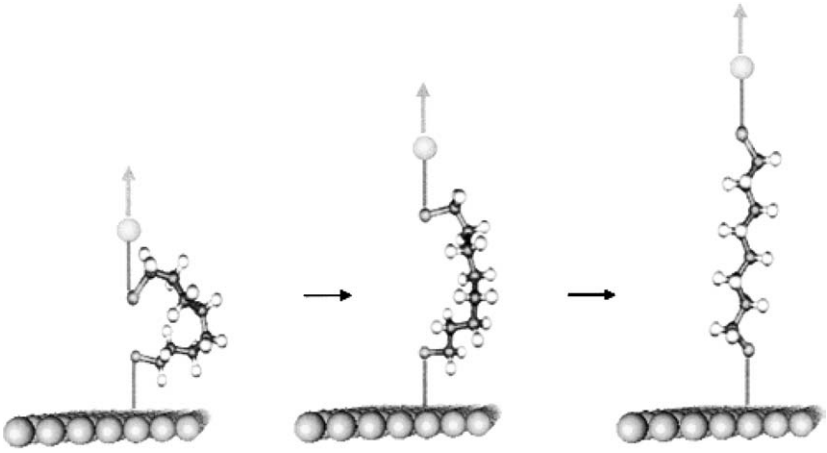


Fig. 18.6. Typical molecular configurations for three molecular extension distances while pulling an octane dithiol molecule using an STM tip.

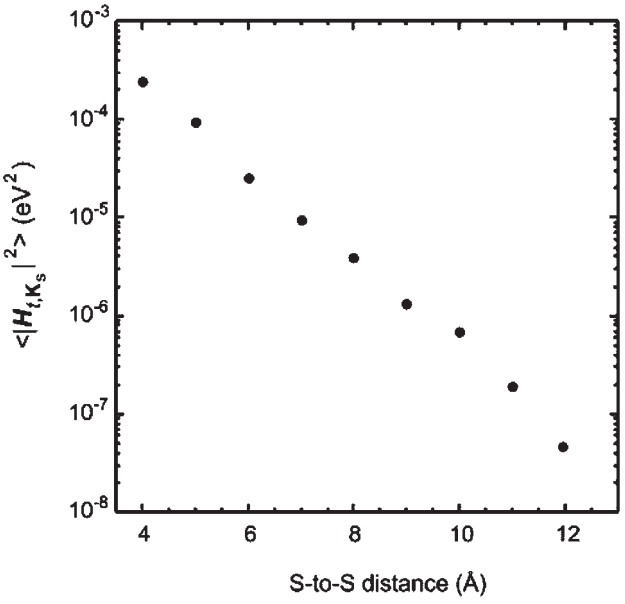


Fig. 18.7. Computed dependence of the tunneling on the extension of an octane dithiol molecule, as a function of sulfur-to-sulfur distance.

distance). We see that the strongest coupling decays approximately exponentially with the sulfur-to-sulfur distance; the decay factor is ~ 1.0 $Å^{-1}$. The weakest and second-weakest couplings do not decay exponentially with sulfur-to-sulfur distance. This result indicates that the strongest conformer dominates the electronic tunneling decay.

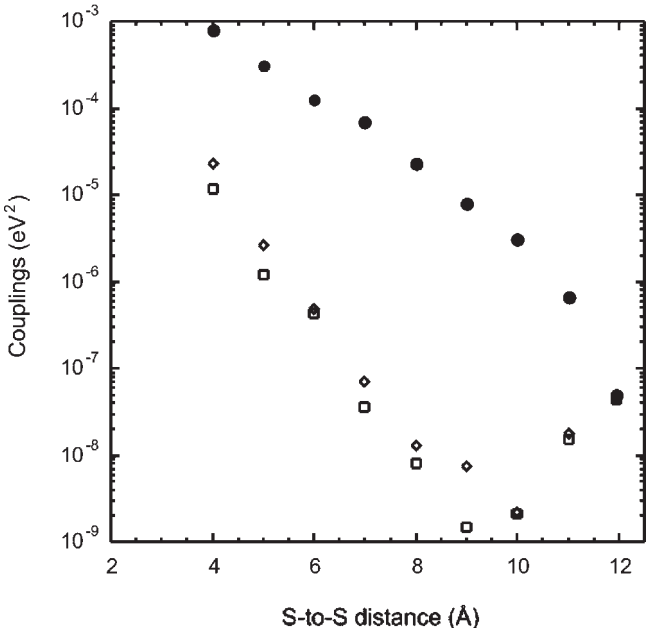
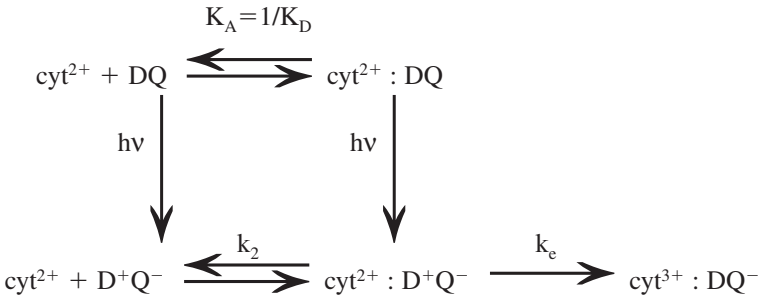


Fig. 18.8. The strongest coupling value for one molecular conformation at each end-to-end distance versus the sulfur-to-sulfur distance. ● represents the strongest coupling, □ represents the weakest coupling, and ◊ represents the second-weakest coupling.

18.11 BIMOLECULAR ET KINETICS: EXPLORING NUMEROUS PROTEIN–PROTEIN CONFORMATIONS [111]

Another instance of conformationally-coupled ET is intermolecular ET. Experimentally, electron transfer between cytochrome c_2 and the photosynthetic reaction center has been studied extensively [112–117]. The reaction scheme below summarizes the excitation, association, and electron transfer events (cyt = cytochrome c_2, DQ = reaction center) [117].



The ET kinetics between the native reaction center and the cytochrome c_2 shows two phases in the experiments [117]. The first-order phase presumably arises from the cytochrome c_2 bound reaction center complex, and the second–order phase likely arises from diffusion-limited ET. Okamura and co-workers [117] found that mutations strongly affect the second-order ET rate for the photosynthetic reaction center (from *Rhodobacter sphaeroides* [118]) /cytochrome c_2 (from *Rhodobacter capsulatus* [119]) system. Early theoretical studies of Onuchic and co-workers [120,121] probed the effects of protein dynamics and docking on electron transfer in *Rhodobacter sphaeroides*. It is challenging to carry out detailed simulations on interprotein electron transfer because of the size of the proteins involved and the range of accessible geometries. We used Brownian dynamics simulations combined with a single-exponential decay model for ET to study this system theoretically [111]. Figure 18.9 shows the comparison between calculated and experimental rates [117]. Most of the calculated bimolecular rates agree with the experimental results within a factor of 10. Both the calculated and experimental results show that mutations have a strong effect on the second-order rates and can change the rates by about three orders of magnitude. Figure 18.10 shows that one of the mutations can switch the ET mechanism from the diffusion-limited regime to the activation-limited regime. In the diffusion-limited regime, the second-order ET rate is independent of the ET pre-exponential factor k_{et}^0, the rate at donor-acceptor contact [111]. In the activation-limited regime, the second-order ET rate increases linearly with k_{et}^0). Recently, a two-step mechanism for the protein–protein binding was described [122–124]: the first step involves protein diffusional association to form an encounter complex, the second step rearranges

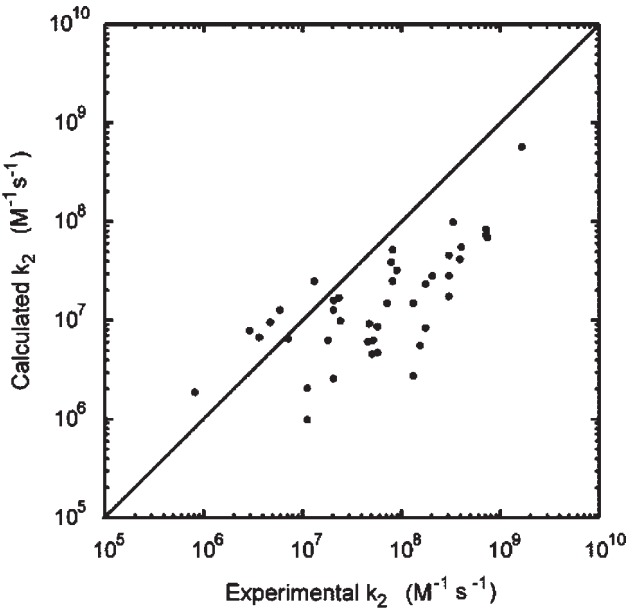


Fig. 18.9. The correlation of calculated and experimental second-order rate constant (k_2) [117] for the native RC/cyt c_2 and double mutants (one on RC, the other one on cyt c_2).

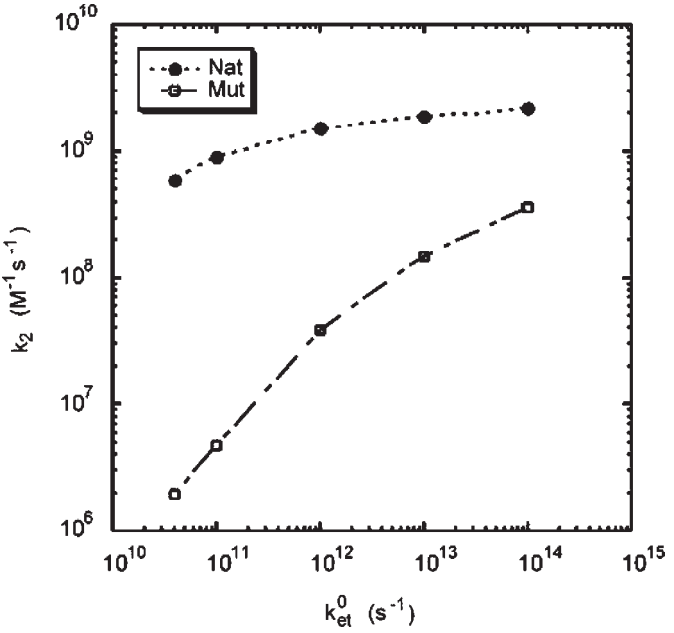


Fig. 18.10. The second-order ET rate for native system and DK(L261)/KE(C99) mutant as a function of ET pre-exponential factor k^{e}_{et}.

the complex to form a more strongly bound ET active complex. Okamura and co-workers proposed adding a third state that involves further complex reorganization [122] for photosynthetic ET. The Brownian dynamics simulations describe the sequence of events in multi-step models under the approximation of frozen protein surface structure. As such, the BD model used here will be incomplete if fluctuations of the protein surface are essential for ET. Nonetheless, the simple theoretical model captures some of the key interprotein geometries that participate in ET and predicts that mutations may switch the mechanism from ET to diffusion control.

18.12 AQUEOUS COUPLING PATHWAYS BETWEEN PROTEINS [32]

Water is expected to influence ET activation free energies as well as electron tunneling barriers [6]. We have examined how the inter-protein electronic coupling interactions change as two solvated redox proteins approach a reactive geometry [32]. The analysis combines molecular dynamics (MD) sampling of conformations for fixed donor–acceptor distances with extended-Hückel analysis of interactions [70,125]. Specifically, we modeled cytochrome b_5 self-exchange. For each MD snapshot, we calculated T_{DA} using extended-Hückel [70,125], pathways [18], and packing density methods [33]. Each MD snapshot was analyzed with and without interfacial water. The conformational averaged squared coupling, $<T^2_{DA}>$, was computed. Figure 18.11 shows $<T^2_{DA}>$ versus the

minimum distance between the two porphyrin rings (the donor/acceptor distance). At the extended-Hückel level without water present, $<T^2_{DA}>$ decays approximately exponentially with distance. With water, $<T^2_{DA}>$ displays a multi-exponential distance dependence with three distinct tunneling regimes. At short distance, water has little effect on interprotein ET and the tunneling is dominated by the protein (direct-contact). At intermediate distances, $<T^2_{DA}>$ is nearly distance independent and water becomes a key tunneling mediator. This is the structured-water regime. At larger distances, water dominates the tunneling and $<T^2_{DA}>$ again decays exponentially with distance. The existence of multiple tunneling regimes for interprotein ET in an aqueous environment provides insight into several recent experimental and theoretical observations. Winkler, Gray and co-workers [126] found that interprotein ET in protein crystals through three water molecules has a similar rate constant to unimolecular ET over the same distance. Klinman and co-workers found that the copper-to-copper ET over about 7 Å in the hydroxylating domain of peptidylglycine α-amidylating monooxygenase shows an unusually large electronic coupling mediated by water, rather than by the protein or the substrate [127]. Pathways analysis of Onuchic and co-workers showed that water molecules mediate the dominant ET coupling pathways between cytochrome c_2 and the photosynthetic reaction center [128]. Cave and co-workers also showed that water between model donors and acceptors can significantly enhance intermolecular ET rates [31b]. These observations all support the conclusion that structured water molecules between the donor and the acceptor can be strong coupling mediators and may enhance ET rates substantially.

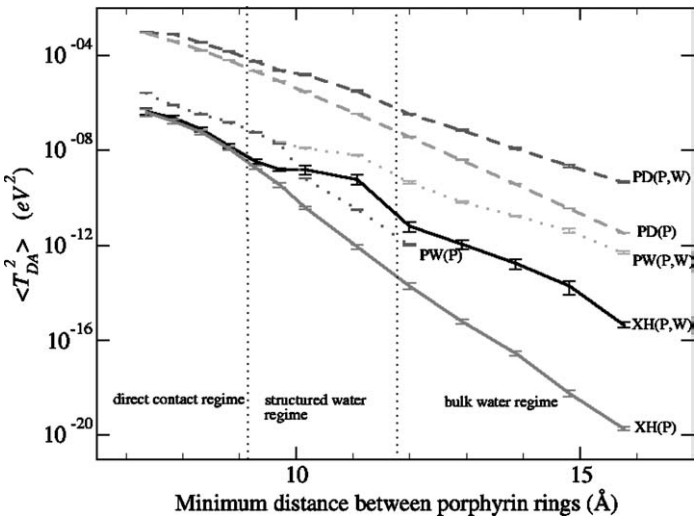


Fig. 18.11. The mean square electronic coupling dependence on the distance between the two porphyrin rings: XH(P,W) - proteins and water, extended-Hückel (XH) calculations; XH(P) - proteins only, XH calculations; PW(P,W) - proteins and water, *Pathways* model analysis; PW(P) - proteins only, *Pathways* model analysis; PD(P,W) - proteins and water, packing density estimates; PD(P) - proteins only, packing density estimates. The three distance ranges correspond to the direct contact regime, structured water mediated regime, and bulk water mediated regime, respectively. The error bars show the sampling error estimated using a renormalization group-based method [129].

18.13 PROSPECTS

The theoretical framework and specific applications outlined here show a richness that emerges as bridge dynamical effects are added explicitly to considerations of electron transfer kinetics. Quantum interferences can be manipulated by inelastic tunneling effects, reaction mechanisms can be switched, and bridge-mediated tunneling can be dominated by minority-population species. Moreover, the familiar timescale separations used in the reaction rate formulations in common use can fail under some circumstances. Further theoretical and experimental examination of these rich dynamical effects suggest a bright future for further dynamical studies of electron transfer reactions.

18.14 ACKNOWLEDGEMENT

S.S. Skourtis was supported by University of Cyprus research funding. D.N. Beratan was supported by National Institutes of Health (GM-048043).

18.15 REFERENCES

1 R.A. Marcus and N. Sutin, Biochim. Biophys. Acta, 811 (1975) 265.
2 N.S. Hush, Trans. Faraday Soc., 57 (1961) 557.
3 R.A. Marcus, J. Chem. Phys., 43 (1965) 679.
4 V.G. Levich and R.R. Dogonadze, Dokl. Acad. Nauk SSSR, 124 (1959) 123.
5 N.R. Kestner, J. Logan and J. Jortner, J. Phys. Chem., 78 (1974) 2148.
6 A.M. Kuznetsov, Charge transfer in physics, chemistry and biology, Gordon and Breach, Amsterdam, 1995.
7 J. Jortner and M. Ratner (Eds.), Molecular electronics, Blackwell Science, Oxford, 1997.
8 A.M. Kuznetsov and J. Ulstrup, Electron tansfer in chemistry and biology, Wiley, Chichetser, 1999.
9 J. Jortner and M. Bixon (Eds.), Electron transfer from isolated molecules to biomolecules, Adv. Chem. Phys., 106–107 (1999).
10 V. May and O. Kühn, Charge and energy transfer dynamics in molecular systems, Wiley-VCH, Weinheim, Germany, 2003.
11 V. Balzani, P. Piotrowiak, M.A.J. Rodgers, J. Mattay, D. Astruc, H. B. Gray, S. Fukuzumi, T.E. Mallouk, Y. Haas, A.P. de Silva and I.R. Gould (Eds.), Electron transfer in chemistry, Vols I–V, Wiley-VCH, Weinheim, 2001.
12 G.B. Schuster (Ed.), Topics in current chemistry, Vols 236–237, Long-range charge transfer in DNA, Springer, New York, 2004.
13 A. Troisi, A. Nitzan and M.A. Ratner, J. Chem. Phys., 119 (2003) 5782.
14 S. Larsson, J. Am. Chem. Soc., 103 (1981) 4034.
15 S.S. Skourtis and D.N. Beratan, Adv. Chem. Phys., 106 (1999) 377.
16 M.D. Newton, in: V. Balzani (Ed.), Electron transfer in chemistry, Vol. I, Wiley VCH Verlag GmbH, Weinhein, Germany, 2001, p. 3.
17 D.N. Beratan and J.J. Hopfield, J. Am. Chem. Soc., 81 (1984) 5753.
18 D.N. Beratan, J.N. Betts and J.N. Ounchic, Science, 252 (1991) 1285.
19 J.N. Onuchic and D.N. Beratan, J. Chem. Phys., 92 (1990) 722.
20 G.L. Closs, L.T. Calcaterra, N.J. Green, K.W. Penfield and J.R. Miller, J. Chem. Phys., 90 (1986) 3673.
21 J.F. Smalley, S. W. Feldberg, C.E.D. Chidsey, M.R. Linford, M.D. Newton and Y.P. Liu, J. Phys. Chem., 99 (1995) 13141.
22 C.Miller, P. Cuendet and M.J. Graetzel, J. Phys. Chem., 95 (1991) 877.

23 H.O. Finklea and D.D. Hanshew, J. Am. Chem. Soc., 114 (1992) 3173.
24 M.D. Newton, Chem. Rev., 91 (1991) 767.
25 M.N. Paddon-Row, M.J. Shephard and K.D. Jordan, J. Phys. Chem., 97 (1993) 1743.
26 M.B. Zimmt and D.H. Waldeck, J. Phys. Chem. A, 107 (2003) 3580.
27 (a) H.B. Gray and J.R. Winkler, Q. Revs. Biophys., 36 (2003) 341; (b) A. Ponce, H.B. Gray and J.R. Winkler, J. Am. Chem. Soc., 122 (2000) 8187.
28 D.S. Bendall, Protein electron transfer, BIOS Scientific Publishers Limited, Oxford, 1996.
29 S. Larsson, J. Phys. Chem., 88 (1984) 1321.
30 M.D. Newton, J. Electroanal. Chem., 438 (1997) 3.
31 (a) I. Benjamin, D. Evans and A. Nitzan, J. Chem. Phys., 106 (1997) 6647. (b) N.E. Miller, M.C. Wander and R. J. Cave, J. Phys. Chem. A, 103 (1999) 1084.
32 J. Lin, I. A. Balabin and D.N. Beratan, Science, 310 (2005) 1311.
33 C.C. Page, C.C. Moser, X. Chen and P.L. Dutton, Nature, 402 (1999) 47.
34 M.L. Jones. I.V. Kurnikov and D.N. Beratan, J. Phys. Chem. A, 106 (2002) 2002.
35 S. Priyadarshy, S.S. Skourtis, S.M. Risser and D.N. Beratan, J. Chem. Phys., 104 (1996) 9473.
36 I.A. Balabin and J.N. Onuchic, J. Phys. Chem., 102 (1998) 7497.
37 C. Liang and M.D. Newton, J. Phys. Chem., 96 (1992) 2855.
38 C.A. Naleway, L.A. Curtiss and J.R. Miller, J. Phys. Chem., 95 (1991) 8343.
39 K.D. Jordan and M.N. Paddon-Row, J. Phys. Chem., 96 (1992) 1188.
40 D.J. Katz and A. A. Stuchebrukhov, J. Chem. Phys., 109 (1998) 4960.
41 T. Kawatsu, T. Kakitani and T. Yamato, J. Phys. Chem. B, 106 (2002) 5068.
42 A.A. Stuchebrukhov, J. Chem. Phys., 194 (1996) 8424.
43 R.J. Cave and M.D. Newton, Chem. Phys. Lett., 249 (1996) 15.
44 M. Lax, J. Chem. Phys., 20 (1952) 1752.
45 E. Neria and A. Nitzan, J. Chem. Phys., 99 (1993) 1109.
46 O.V. Prezhdo and P.J. Rossky, J. Chem. Phys., 107 (1997) 5863.
47 I.A. Balabin and J.N. Onuchic, Science, 290 (2000) 114.
48 M. Bixon and J. Jortner, Russ. Jo. Electrochem., 39 (2003) 5.
49 J.N. Onuchic and A.A.S. da Gama, Theor. Chim. Acta, 69 (1986) 89.
50 K.V. Mikkelsen, J. Ulstrup and M.G. Zakaraya, J. Am. Chem. Soc., 111 (1989) 1315.
51 R.F. Goldstein, S. Franzen and W. Bialek, J. Phys. Chem., 97 (1993)11168.
52 A.M. Kuznetsov and M.D. Vigdorovich, J. Ulstrup, Chem. Phys., 176 (1993) 539.
53 J. Tang, J. Chem. Phys., 98 (1993) 6263.
54 I.A. Goychuk, E.G. Petrov and V. May, J. Chem. Phys., 103 (1995) 4937.
55 E.S. Medvedev and A.A. Stuchebrukhov, J. Chem. Phys., 107 (1997) 3821
56 J.L. Laio and G.A. Voth, J. Chem. Phys., 116 (2002) 9174.
57 A. Milischuk and D. Matyushov, J. Chem. Phys., 118 (2003) 5596.
58 S. Jang and M.D. Newton, J. Chem. Phys., 122 (2005) 24501.
59 D. Rehm and A. Weller, Isr. J. Chem., 8 (1970) 259.
60 G.A. Mines, M.J. Bjerrum, M.G. Hill, D.R. Casimiro, I.-J. Chang, J.R. Winkler and H.B. Gray, J. Am. Chem. Soc., 118 (1996) 1961.
61 J. Wolfgang, S.M. Risser, S. Priyadarshy and D.N. Beratan, J. Phys. Chem. B, 101 (1997) 2986.
62 I. Daizadeh, E.S. Medvedev and A.A. Stuchebrukhov, Proc. Natl. Acad. Sci. USA, 94 (1997) 3703.
63 L.W. Ungar, M.D. Newton and G.A. Voth, J. Phys. Chem. B, 103 (1999) 7367.
64 N.E. Miller, M.C. Wander and R.J. Cave, J. Phys. Chem. A, 103 (1999) 1084
65 E.W. Castner, D. Kennedy and R.J. Cave, J. Phys. Chem. A, 104 (2000) 2869
66 Q. Xie, G. Archontis and S.S. Skourtis, Chem. Phys. Lett., 312 (1999) 237.
67 S.S. Skourtis, Q. Xie and G. Archontis, J. Chem. Phys., 115 (2001) 9444.
68 M.D. Newton, Int. J. Quantum Chem. 77 (2000) 255.
69 A. Troisi and G. Orlandi, J. Phys. Chem. B, 106 (2002) 2093.
70 T. Kawatsu, T. Kakitani and T.J. Yamato, J. Phys. Chem. B, 106 (2002) 11356
71 M.B. Zimmt and D.H. Waldeck, J. Phys. Chem. A, 107 (2003) 3580.
72 H. Nishioka, A. Kimura, T. Yamato, T. Kawatsu and T. Kakitani, J. Phys. Chem. B, 109 (2005) 1978.

73 A. Troisi, M.A. Ratner and M.B. Zimmt, J. Am. Chem. Soc., 126 (2004) 2215.
74 S.S. Skourtis, I.A. Balabin, T. Kawatsu and D.N. Beratan, Proc. Natl. Acad. Sci. USA, 102 (2005) 3552.
75 H. Nishioka, A. Kimura, T. Yamato, T. Kawatsu and T. Kakitani, J. Phys. Chem. B, 109 (2005) 15621.
76 D.M. Lockwood, Y.-K. Cheng and P.J. Rossky, Chem. Phys. Lett., 345 (2001) 159.
77 J.R. Reimers, N.S. Hush, D.M. Sammeth and P.R. Callis, Chem. Phys. Lett., 169 (1990) 622.
78 J.R. Reimers and N.S. Hush, Chem. Phys., 146 (1990) 105.
79 W.T. Yip, D.H. Levy, R. Kobetic and P. J. Piotrowiak, J. Phys. Chem. A., 103 (1999) 10.
80 G.A. Jones, M.N. Paddon-Row, B.K. Carpenter and P.J. Piotrowiak, J. Phys. Chem. A., 106 (2002) 5011.
81 W.B. Davies, W.A. Svec, M.A. Ratner and M.R. Wasielewski, Nature, 396 (1998) 60.
82 W.B. Davies, M.A. Ratner and M.R. Wasielewski, J. Am. Chem. Soc., 123 (2001) 7877.
83 M. Galperin, A. Nitzan and I. Benjamin, J. Phys. Chem. A, 106 (2002) 10790.
84 For a review: A. Nitzan, Annu. Rev. Phys. Chem., 52 (2001) 681.
85 A. Troisi, M.A. Ratner and A. Nitzan, J. Chem. Phys. 2003 (2003) 6072.
86 S.S. Skourtis, D.H. Waldeck and D.N. Beratan, J. Phys. Chem. B, 108 (2004) 15511.
87 M. Abu-Hilu and U. Peskin, J. Chem. Phys., 112 (2005) 021103.
88 D.N. Beratan and J.J. Hopfield, J. Chem. Phys., 81 (1984) 5753.
89 S.S. Skourtis, Chem. Phys. Lett., 372 (2003) 224.
90 A.Teklos and S.S. Skourtis, Chem. Phys. 319 (2005) 52.
91 S. Medvedev and A.A. Stuchebrukhov Chem. Phys., 296 (2004) 181.
92 A.A. Voityuk, J. Chem. Phys., 123 (2005) 034903.
93 J. Lin and D.N. Beratan, J. Phys. Chem. A, 108 (2004) 5655.
94 G. Binnig and H. Rohrer, Helv. Phys. Acta, 55 (1982) 726.
95 G. Binnig and H. Rohrer, Rev. Mod. Phys., 56 (1987) 615.
96 T.A. Jung, R.R. Schittler, J.K. Gimzewski, H. Tang and C. Joachim, Science, 271 (1996) 181.
97 L. Bartels, G. Meyer and K.–H. Rieder, Phys. Rev. Lett., 79 (1997) 697.
98 X. Bouju, C. Girard, H. Tang, C. Joachim and L. Pizzagalli, Phys. Rev. B, 55 (1997) 16498.
99 J.A. Stroscio and D.M. Eigler, Science, 254 (1991) 1319.
100 R. Krautbauer, H. Clausen-Schaumann and H.E. Gaub, Angew. Chem. Int. Ed., 39 (2000) 3912.
101 L.T. Cai, H. Tabata and T. Kawai, Nanotechnology, 12 (2001) 211.
102 V. Mujica, M. Kemp and M.A. Ratner, J. Chem. Phys., 101 (1994) 6849.
103 V. Mujica, M. Kemp and M.A. Ratner, J. Chem. Phys., 101 (1994) 6856.
104 S.N. Yaliraki, M. Kemp and M.A. Ratner, J. Am. Chem. Soc., 121 (1999) 3428.
105 S. Datta, Electron transport in mesoscopic systems, Cambridge University Press, Cambridge, 1995.
106 N.D. Lang, Phys. Rev. B, 52 (1995) 5335.
107 E.G. Emberly and G. Kirczenow, Phys. Rev. B, 58 (1998) 10911.
108 P. Sautet and C. Joachim, Phys. Rev. B, 38 (1988) 12238.
109 H. Ou-Yang, B. Källebring and R.A. Marcus, J. Chem. Phys., 98 (1993) 7565.
110 F. Wang, D.C. Sorescu and K.D. Jordan, J. Phys. Chem. B, 106 (2002) 1316.
111 J. Lin and D.N. Beratan, J. Phys. Chem. B, 109 (2005) 7529.
112 R.C. Prince, R.J. Cogdell and A.R. Crofts, Biochim. Biophys. Acta, 347 (1974) 1.
113 C.C. Moser and P.L. Dutton, Biochemistry, 27 (1988) 2450.
114 G. Venturoli, A. Mallardi and P. Mathis, Biochemistry, 32 (1993) 13245.
115 P.D.M. Tiede, A.C. Vashista and M.R. Gunner, Biochemistry, 32 (1993) 4515.
116 M. Tetreault, S.H. Rongey, G. Feher and M.Y. Okamura, Biochemistry, 40 (2001) 8452.
117 M. Tetreault, M. Cusanovich, T. Meyer, H. Axelrod and M. Y. Okamura, Biochemistry, 41 (2002) 5807.
118 H.L. Axelrod, E. C. Abresch, M.Y. Okamura, A.P. Yeh, D. C. Rees and G. Feher, J. Mol. Biol., 319 (2002) 501.
119 M.M. Benning, G. Weaenberg, M.S. Caffrey, R.G. Bartsch, T.E. Meyer, M.A. Cusanovich, I. Rayment and H.M. Holden, J. Mol. Biol., 220 (1991) 673.
120 A.J.A. Aquino, P. Beroza, J. Reagan and J.N. Onuchic, Chem. Phys. Lett., 275 (1997) 181.
121 A.J.A. Aquino, O. Beroza, D.N. Beratan and J.N. Onuchic, Chem. Phys., 197 (1995) 277.
122 O. Miyashita, J.N. Onuchic and M.Y. Okamura, Biochemistry, 42 (2003) 11651.
123 R.R. Gabdoulline and R.C. Wade, Curr. Opin. Struct. Biol., 12 (2002) 204.

124 A.J. Camacho and S. Vajda, Curr. Opin. Struct. Biol., 12 (2002) 36.
125 R. Hoffmann, J. Chem. Phys., 39 (1963) 1937.
126 F.A. Tezan, B.R. Crane, J.R. Winkler and H.B. Gray, Proc. Natl. Acad. Sci. USA, 98 (2001) 5002.
127 W.A. Francisco, G. Wille, A.J. Smith, D.J. Merkler and J.P. Klinman, J. Am. Chem. Soc., 126 (2004) 13168.
128 O. Miyashita, M.Y. Okamura and J.N. Onuchic, Proc. Natl. Acad. Sci. USA, 102 (2005) 3558.
129 H. Flyvjerg and H.G. Petersen, J. Chem. Phys., 91 (1989) 461.

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 19

Modelling molecular conduction in DNA wires: charge transfer theories and dissipative quantum transport

Ralf Bulla[1], Rafael Gutierrez[2] and Gianaurelio Cuniberti[2]

[1]*Theoretische Physik III, Universität Augsburg, D-86135, Augsburg, Germany*
[2]*Institut für Theoretische Physik Korrelationen und Magnetismus, Universität Regensburg, D-93040, Regensburg, Germany*

Abstract

Measurements of electron transfer rates as well as of charge transport characteristics in DNA have produced a number of seemingly contradictory results, ranging from insulating behaviour to the suggestion that DNA is an efficient medium for charge transport. Among other factors, environmental effects appear to play a crucial role in determining the effectivity of charge propagation along the double helix. This chapter gives an overview of charge transfer theories and their implication for addressing the interaction of a molecular conductor with a dissipative environment. Further, we focus on possible applications of these approaches for charge transport through DNA-based molecular wires.

19.1 INTRODUCTION

The discovery of long-range electron transfer processes in double stranded DNA [1] considerably attracted the attention of biologists, chemists and physicists. The motivation is threefold: (i) the possible use of DNA molecules in nanotechnology applications [2], (ii) the biological role of electron transfer in, for example, radiation damage and repair [3], and (iii) the potentials of biochemical sensors based on electron transfer in DNA [3].

Despite the intensive experimental efforts, the results for electron *transport* still appear to be contradictory, ranging from metallic conduction [4,5] to insulating behaviour with very large bandgaps [6,7]. We refer the reader to Ref. [8] for a recent review on this topic. The measurements of electron *transfer*, on the other hand, appear to be much better controlled and earlier discrepancies on the distance dependence of the electron transfer rate are now attributed to the different experimental set-ups [8].

Theoretically, several classes of factors have meanwhile been identified, which considerably determine the effectivity of charge propagation along the double helix. They can be roughly classified as being related to (i) static disorder associated with the random or quasi-random sequence of bases in DNA oligomers [9–11], (ii) dynamical disorder arising from strong structural fluctuations of the molecular frame [12–14], and (iii) environmental effects related to the presence of an aqueous environment and counter-ions [15–21]. While the first two factors can still be addressed in a first approximation by considering only the atomic structure of isolated DNA oligomers, environmental effects require the consideration of the solvation shells and counter-ions and their interaction with the DNA molecules. Though the performance of *ab initio* approaches has considerably improved in recent years, the description of the dynamical interaction of DNA with an environment is still a formidable computational task, involving at least several thousands of atoms. As a consequence, only relatively few first-principle studies addressing this issue have been carried out in the recent years [15–21]. Thus, model Hamiltonian approaches describing charge propagation in the presence of a dissipative environment are very valuable and help to gain some insight into the subtleties of the physical behaviour of a quantum mechanical system interacting with a macroscopic number of degrees-of-freedom.

This chapter gives an overview of different approaches to address charge propagation in a dissipative environment. In the next section, we discuss some results from *ab initio* calculations of DNA oligomers in the presence of an aqueous environment. In Section 19.3 some basic facts on how to model the interaction between an arbitrary quantum mechanical system in interaction with a dissipative environment are introduced. Finally, in subsection 19.3.1, a special application to a DNA model is discussed.

19.2 ENVIRONMENTAL EFFECTS WITHIN *AB INITIO* APPROACHES

For the purposes of illustrating some basic facts concerning the electronic structure of a *dried* DNA oligomer, let us look at a recent band-structure calculation of poly(GC) carried out with the density-functional-based code SIESTA [23]. First-principle results for poly(AT) oligomers have also recently been presented [24–26]. The natural advantage of poly(GC) or Poly(AT) is its periodic structure, which considerably minimizes the computational efforts. In Fig. 19.1, the resulting band structure is shown. From the practical point of view, it is also expected that this kind of periodic structure will have a higher potential applicability in molecular electronics than their disordered counterparts like λ-DNA.

The topmost valence band (HOMO) and the lowest conduction band (LUMO), both having π-character, are separated by a bandgap of $\Delta = 2.0$ eV. The HOMO and LUMO bands are basically derived from the overlap of guanine and cytosine orbitals, respectively. As a consequence, the charge density of HOMO and LUMO bands is confined along the G- and C-strands, respectively. The Fermi level lies between these two bands so that the system appears to be an insulator.

Most striking are the very small bandwidths of the bands close to the Fermi level. The topmost valence band has a width of $W_{\mathrm{H}} = 40$ meV while the lowest conduction band has a somewhat broader bandwidth of $W_{\mathrm{L}} = 270$ meV. Part (b) of Fig. 1 of Ref. [23]

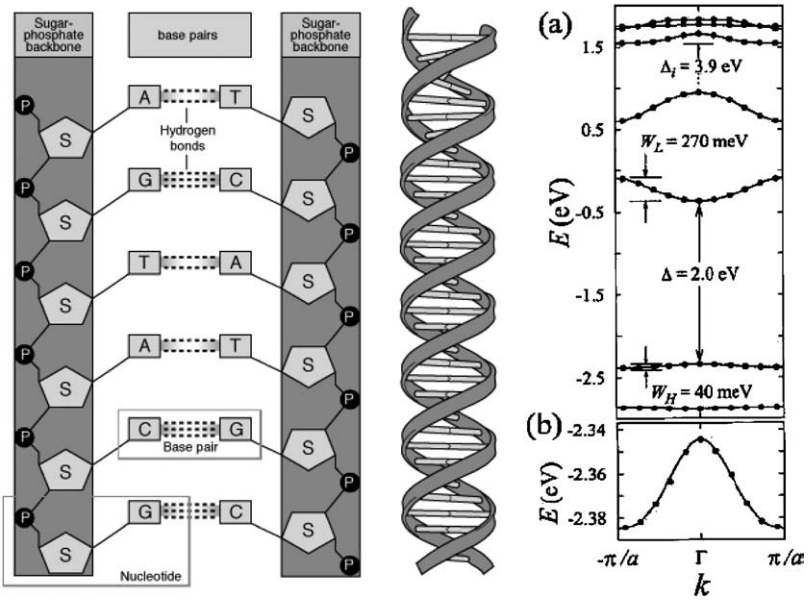


Fig. 19.1. Left: schematic representation of a double-stranded DNA with an arbitrary base-pair sequence. Right: electronic band-structure of poly(GC) DNA within density-functional theory. The figures are reproduced from Refs [22,23] with permission.

(see Fig. 19.1, right panel) shows a tight-binding modelling of the topmost valence band using a single orbital per base pair. The resulting hopping matrix elements are $t = 10$ meV (for nearest neighbour hopping) and $t' = 1.5$ meV (for next-nearest neighbour hopping).

In principle, such a scenario allows for electronic transport mediated by carriers which are introduced by doping either in the top of the valence band or the bottom of the conduction band. The extremely small bandwidths, however, suggest that a Bloch states transport scenario cannot be stable under the various perturbations present in DNA [23].

One of the possible perturbations, which has been studied in some detail, is the role of the environment, necessary to stabilize the double-helix [15–19]. As shown in Ref. [15], the existence of rather different timescales of the environment may have a strong impact on a charge propagating along the DNA molecule. First-principle simulations were performed, including four base pairs of B-DNA in the sequence GAGG, together with Na counter-ions and the hydration shell. It turns out that holes can be gated by the temperature-dependent dynamics of the environment, i.e., there may exist configurations that localize the hole. Dynamical fluctuations of the counter-ions can lead, however, to configurations which support hole motion. A hole thus experiences transitions between quantum-mechanical states that are correlated with different environmental configurations [27]. These results have been partly confirmed by recent *ab initio* simulations in Ref. [19]. The authors have additionally pointed out at a different, proton-mediated mechanism for hole localization, which may be quite effective in poly(GC) DNA.

The *ab initio*-based studies in Refs [16–18] have yielded further insight into the role played by water and counter-ions in modifying the low-energy electronic structure of DNA oligomers. Despite the differences in the DNA conformations (Z- [16] versus B-DNA [17,18]) as well as in computational approaches (different basis sets and approximations for the exchange-correlation potentials), they nevertheless indicate that the environment can introduce midgap states. Though these electronic states do not form truly extended electronic bands, they may support activated charge hopping at high temperatures and thus lead to an enhancement of the conductivity. In this respect, they resemble, to some degree, the defect levels induced by impurities in bulk semiconductors.

We can conclude from this that (i) the appearance of a bandgap is not at all a generic feature for the bandstructure of DNA, and (ii) the extremely small values of the bandwidths do appear to be generic. The general question which arises from that is the relation of the bandwidths W_H and W_L to other typical energy scales due to disorder effects, electron–phonon coupling, and Coulomb correlations. Furthermore, the environment can have a dramatic influence on the electronic structure of the oligomer by inducing defect-like states within the $\pi - \pi^*$ gap. However, as previously stated, the complexity of the problem makes a full *ab initio* treatment rather difficult. This leads us to the issue of how the system – environment interaction can be modelled within a Hamiltonian model approach. Which are the essential ingredients that have to be taken into account?

19.3 MODELLING THE SYSTEM–ENVIRONMENT INTERACTION

The importance of the system–environment interaction has long been recognized in biomolecules (such as proteins), in which electron transfer reactions take place. Within Marcus theory [28], the coupling of the electronic degrees-of-freedom to a reaction coordinate is the first step to a successful description of electron transfer processes. The quantum-mechanical analogue of the reaction coordinate is a phononic degree-of-freedom originating from vibrations of the protein matrix. In general, there might not be one dominating phononic mode; such a breakdown of the standard single reaction coordinate description has been suggested in the context of charge transfer between DNA base pairs [29]. More importantly, even a dominating reaction coordinate is coupled to the fluctuations of the environment, such as surrounding water molecules, so that the resulting spectral function $J(\omega)$ of all relevant phononic modes can be regarded as continuous over a very broad energy range (in theoretical calculations, the low-energy cut-off is typically set to $\omega = 0$).

The coupling of the electronic subsystem (the electron transfered between donor and acceptor site) to the environment leads to a very important effect: when an electron initially localized at the donor site, tunnels to the acceptor site (which typically has a lower energy), the energy difference is dissipated to the environment so that the electron transfer process is irreversible [30]. If this friction were to be too small, or if the electron were to couple to a single phonon mode only, the electron would oscillate between donor and acceptor sites, making the electron transfer process highly inefficient.

In the work of Garg *et al.* [30], the friction term has been modelled quantum mechanically via a coupling to a bath of harmonic oscillators. A minimal model for electron transfer processes, similar to the one proposed in [30], then takes the form:

$$H = \sum_{i=A,D} \varepsilon_i c_i^\dagger c_i - t\,(c_D^\dagger c_A + c_A^\dagger c_D) + \sum_n \omega_n b_n^\dagger b_n + (g_A n_A + g_D n_D) \sum_n \frac{\lambda_n}{2}\,(b_n^\dagger + b_n) \quad (1)$$

The operators $c_i^{(\dagger)}$ denote annihilation (creation) operators for electrons on the donor (i = D) and acceptor (i = A) sites; $n_{A/D}$ is defined as $n_i = c_{i\sigma}^\dagger c_{i\sigma}$. The first two terms of the Hamiltonian, Eq. (1), correspond to a two-site tight-binding Hamiltonian with ε_i the on-site energies and t the hopping matrix element.

The last two terms in Eq. (1) describe the free bosonic bath (with bosonic creation and annihilation operators $b_n^\dagger$ and b_n) and the coupling between electrons and bosons, respectively. Assuming symmetric phonon displacements due to the electronic occupancy at donor and acceptor sites, one can set $g_A = 1$ and $g_D = -1$ see Ref. [31].

The coupling of the electrons to the bath degrees-of-freedom is completely specified by the bath spectral function $J(\omega) = \pi \sum_n \lambda_n^2 \delta(\omega - \omega_n)$. The form of $J(\omega)$ can, in principle, be calculated with molecular dynamics simulations (see, for example, [32]). To study the qualitative influence of the environment, an ohmic bath spectral function $J(\omega) \propto \omega$ (with a suitable high-energy cut-off) is sufficient for most cases. In this description, dominant reaction coordinates lead to additional resonances in the bath spectral function.

Note that such a continuous bath spectral function enforces a *quantum-mechanical* treatment of the phononic degrees-of-freedom, since the temperature range always lies within the continuum of phononic modes.

The Hamiltonian, Eq. (1), can be viewed as a paradigm for modelling the system-environment interaction in biomolecules in which the electronic degrees-of-freedom couple to a dissipative environment. We should add here that the model, Eq. (1), can be exactly mapped onto the well-studied spin-boson model [33,34] for the case of *one* electron in the system. In the spin-boson model description, the state $|\uparrow\rangle$ ($|\downarrow\rangle$) corresponds to the electron localized at the donor (acceptor) site. More complicated situations arise when the spin degree-of-freedom of the electron — not to be confused with the artificial spin in the spin-boson model — is taken into account, see the discussion in Ref. [35].

Calculations for these types of models in the context of electron transfer problems have been presented in [30,35,36]. It is natural to assume that the electron environment interaction plays an equally important role for electron transport through the DNA double helix. The main difference here is that electron transfer/transport occurs over very many sites so that the two-site model, Eq. (1), has to be suitably generalized. One such example is discussed in the following section.

19.3.1 Modelling the system–environment interaction: a DNA-wire in a dissipative bath

The first-principle calculations reviewed in Section 19.1 have shown that the environment in which DNA oligomers are placed may have a non-negligible influence on their

electronic structure. In this section, we will illustrate within an effective model Hamiltonian approach, how the presence of a dissipative environment does affect the low-energy transport properties of a DNA molecular wire [37,38]. Our reference system will be poly(GC) because of its periodic structure, which should make optimal the inter-base electronic coupling along the strands. Moreover, recent experiments [40] on single poly(GC) molecules have shown non-zero current at low bias, which is at variance with the fact that the molecule should have a (rather large) HOMO–LUMO gap [41]. Since these experiments were performed in an aqueous environment and the authors excluded ionic current contributions, one may consider the possibility that the environment is modifying the molecule's electronic structure.

In our model, we will exclusively focus on the low-energy transport, i.e., the charge injection energies are small compared with the molecular bandgap of the isolated molecule (~ 2 – 3 eV). Consequently, only equilibrium transport will be considered and a transmission-like function can still be defined [37,38]. At low energies, only the frontier orbitals (HOMO and LUMO) of the molecule are expected to contribute to transport. As mentioned in section 19.1 these orbitals have π-character and their charge densities extend along the G- and C-strands for the HOMO and the LUMO, respectively. Motivated by this, we have formulated a minimal tight-binding model [37,38,41], where a single electronic π-orbital channel is connected to left and right electrodes, see Fig. 19.2. The sugar–phosphate backbones are assumed to locally perturb the electronic states and lead to the opening of a semiconducting gap for the infinite chain. As a result, the size of the bandgap can be controlled by the strength of this perturbation, given by the parameter $t_\perp$ in Eq. (2) below. The environment is described by a collection of harmonic oscillators which linearly couple to the charge density on the backbone sites. Assuming zero on-site energies (the Fermi level thus lies at $E = 0$), the Hamiltonian reads:

$$\begin{aligned}\mathcal{H} = & -t_\pi \sum_j [c_j^\dagger c_{j+1} + h.c.] - t_\perp \sum_j \Big[b_j^\dagger c_j + h.c. \Big] \\ & + \sum_\alpha \Omega_\alpha B_\alpha^\dagger B_\alpha + \sum_{\alpha,j} \lambda_\alpha b_j^\dagger b_j (B_\alpha + B_\alpha^\dagger) \\ & + \sum_{\mathrm{k}\in \mathrm{L,R},\sigma} \varepsilon_{\mathrm{k}\sigma} d_{\mathrm{k}\sigma}^\dagger d_{\mathrm{k}\sigma} + \sum_{\mathrm{k}\in \mathrm{L},\sigma} (V_{\mathrm{k},1}\, d_{\mathrm{k}\sigma}^\dagger\, c_1 + h.c.) + \sum_{\mathrm{k}\in \mathrm{R},\sigma} (V_{\mathrm{k},N}\, d_{\mathrm{k}\sigma}^\dagger\, c_N + h.c.) \\ = & \ \mathcal{H}_{\mathrm{el}} + \mathcal{H}_{\mathrm{B}} + \mathcal{H}_{\mathrm{leads}} \end{aligned} \tag{2}$$

In the above equation, $\mathcal{H}_{\mathrm{el}} = \mathcal{H}_{\mathrm{c}} + \mathcal{H}_{\mathrm{b}}$ is the Hamiltonian of the π-channel ($\mathcal{H}_{\mathrm{c}}$) and the backbone sites ($\mathcal{H}_{\mathrm{b}}$); $\mathcal{H}_{\mathrm{B}}$ contains both the Hamiltonian of the bath and the mutual interaction of the bath with the electronic degrees-of-freedom at the backbone sites (second row). Finally, $\mathcal{H}_{\mathrm{leads}}$ contains the electrode Hamiltonians as well as the tunnelling Hamiltonian describing the propagation of a charge from the leads onto the π-channel and vice versa. In the absence of coupling to the bath, the eigenstates of $\mathcal{H}_{\mathrm{el}}$ yield two manifolds containing N states each and separated by a bandgap, whose magnitude basically depends on the size of the tranversal coupling $t_\perp$. The bath is completely described by introducing its spectral density as given by [34]: $J(\omega) = J_0\, (\omega/\omega_{\mathrm{c}})\, \exp^{-\omega/\omega_{\mathrm{c}}}$, where

ω_c is a high-frequency cut-off and we assume ohmic dissipation, $J(\omega) \sim \omega$. By performing a unitary transformation, the linear coupling to the bath can be eliminated. However, the transversal coupling terms will be renormalized by exponential bosonic operators [37,38]. Using equation of motion techniques, one can show, to lowest order in $t_\perp$, that the Green function of the wire satisfies the following Dyson equation:

$$\mathbf{G}^{-1}(E) = E\mathbf{1} - \mathcal{H}_c - \Sigma_L(E) - \Sigma_R(E) - t_\perp^2 \mathbf{P}(E) \tag{3}$$

In this expression, the influence of the electrodes is captured by the complex self-energy functions $\Sigma_{L/R}(E)$. The function $P(E)$ is an entangled electron-boson Green function: $P_{jl}(t) = -i\Theta(t) \left\langle [b_j(t)\chi(t), b_l^\dagger(0)\,\chi^\dagger(0)] \right\rangle$ and $\chi = \exp[\Sigma_\alpha(\lambda_\alpha/\Omega_\alpha)\,(B_\alpha - B_\alpha^\dagger)]$. Note that $P(E)$ acts as an additional self-energy and that the influence of the backbones is contained only in this function.

Several coupling regimes to the bath can be analysed [38]. We focus here only on the strong-coupling limit (SCL), defined by the condition $J_0/\omega_c > 1$, which basically means that the timescales of the charge-bath interaction are much shorter compared with typical electronic timescales. We refer the reader to Refs [37,38] for technical details.

The impact of the bath on the electronic structure is twofold [37,38]. On one side, the strong coupling to the bath leads to the emergence of new bath-induced electronic states *inside* the wire bandgap. On the other side, however, these states are strongly damped by the dissipative action of the bath. In other words, the bath completely destroys the coherence of transport through the wire. This effect has also been discussed for transport through molecular chains under the influence of external time-dependent fields, see [42,43]

As a result, the bath-induced states will not manifest as resonances in the transmission spectrum, see Fig. 19.2, left panel. Nevertheless, they induce a *temperature dependent* background, which leads to a (small) finite density of states inside the gap. Charges injected

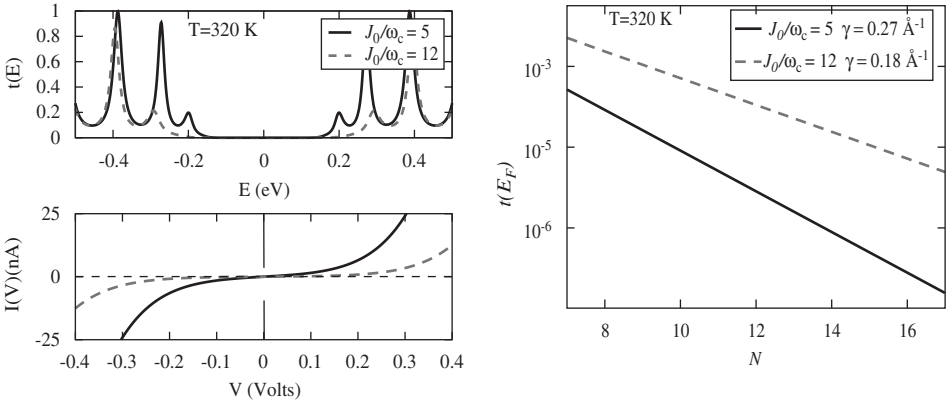


Fig. 19.2. Left panel: transmission and current for different electron–boson coupling strengths. At high temperatures, a small density of states is present at low energies. Right panel: correspnding dependence of the transmission at the Fermi energy on the number of sites N in the wire. A very weak exponential scaling is found, hinting at a strong contribution of incoherent processes.

at low energies will now find states supporting transport at high temperatures and thus, a finite current at low bias may flow. Hence, we call the new gap a pseudo-gap, in contrast to the intrinsic bandgap found in the isolated wire. Note that increasing the interaction with the bath (increasing J_0/ω_c) does not necessarily lead to a global increase of the current, since the frontier orbitals of the wire are strongly damped with increasing coupling.

Signatures of this situation are seen in the length dependence of the transmission at the Fermi energy, see Fig. 19.2, right panel. Tunnelling through typical intrinsic gap would lead to a very strong exponential length dependence $t(E_F) \sim e^{-\gamma N}$ with typical inverse decay lengths $\gamma \sim 1.5 - 2$ Å^{-1} in the case of poly(GC) [40]. We find, however, much smaller values $\sim 0.1 - 0.2$ Å^{-1}. With increasing bath coupling the exponential dependence weakens, reflecting the increase of the density of states in the pseudo-gap and the strong contribution of incoherent processes [44].

19.4 CONCLUSIONS AND OUTLOOK

In conclusion, we have shown that model approaches can deal with experimentally relevant situations. One of the major problems for the theoretical modelling of charge transport in DNA oligomers is the lack of a clear experimental picture of their transport signatures. Thus, focusing on individual factors affecting charge propagation helps to shed light onto the relevant mechanisms controlling the charge dynamics in DNA. We have addressed in this chapter environmental effects in the charge transport through DNA oligomers within a minimal Hamiltonian model approach. Obviously, other factors not treated here, like electronic correlations, static disorder or internal vibrational excitations can also have a non-negligible influence on charge propagation.

19.5 ACKNOWLEDGEMENTS

The authors thank A. Nitzan for useful comments and suggestions. This research was supported by the DFG through SFB 484, by the Volkswagen Foundation under grant Nr. I/78-340 and by the EU under contract IST-2001-38951.

19.6 REFERENCES

1 C.J. Murphy, M.R. Arkin, Y. Jenkins, N.D. Ghatlia, S.H. Bossmann, N.J. Turro and J.K. Barton, Science 262 (1993) 1025.
2 C. Dekker and M. Ratner, Physics World, August 2001.
3 E.M. Boon and J.K. Barton, Curr. Op. in. Struct. Biol., 12 (2002) 320.
4 H.-W. Fink and C. Schönenberger, Nature, 398 (1999) 407.
5 A.Y. Kasumov, M. Kociak, S. Guron, B. Reulet, V. T. Volkov, D.V. Klinov and H. Bouchiat, Science, 291 (2001) 280.
6 D. Porath, A. Bezryadin, S.D. Vries and C. Dekker, Nature, 403 (2000) 635.
7 A.J. Storm, J.V. Noort, S.D. Vries and C. Dekker, Appl. Phys. Lett., 79 (2001) 3881.
8 D. Porath, G. Cuniberti and R. Di Felice, Topics Curr. Chem., 237 (2004) 183.
9 S. Roche, D. Bicout, E. Macia and E. Kats, Phys. Rev. Lett., 228101 (2003).

10 S. Roche, Phys. Rev. Lett., 91 (2003) 108101.
11 D. Klotsa, R.A. Röemer and M. S. Turner, q-bio.GN/0504004, 2005.
12 F. Palmero, J.F.R. Archilla, D. Hennig and F.R. Romero, New J. Phys., 6 (2004) 13.
13 S. Komineas, G. Kalosakas and A.R. Bishop, Phys. Rev. E, 65 (2002) 061905.
14 F.C. Grozema, L.D.A. Siebbeles, Y.A. Berlin and M.A. Ratner, Chem. Phys. Chem., 6 (2002) 536.
15 R.N. Barnett, C.L. Cleveland, A. Joy, U. Landman and G.B. Schuster, Science, 294 (2001) 567.
16 F.L. Gervasio, P. Carolini and M. Parrinello, Phys. Rev. Lett., 89 (2002) 108102.
17 R.G. Endres, D.L. Cox and R.R.P. Singh, Rev. Mod. Phys., 76 (2004) 195.
18 A. Hübsch, R.G. Endres, D.L. Cox and R.R.P. Singh, Phys. Rev. Lett., 94 (2005) 178102.
19 F.L. Gervasio, A. Laio, M. Parrinello and M. Boero, Phys. Rev. Lett., 94 (2005) 158103.
20 Ch. Adessi, S. Walch and M.P. Anantram, Phys. Rev. B, 67 (2003) 081405(R).
21 H. Mehrez and M.P. Anantram, Phys. Rev. B, 71 (2005) 115405.
22 http://www.accessexcellence.org/RC/VL/GG/;http://www.genome.gov/.
23 E. Artacho, M. Machado, D. Sanchez-Portal, P. Ordejon and J.M. Soler, Mol. Phys., 101 (2003) 1587.
24 E.B. Starikov, Phys. Chem. Chem. Phys., 4 (2002) 4523.
25 J.P. Lewis, J. Picus, Th. E. Cheatham III, E.B. Starikov, H. Wang, J. Tomfohr and O.F. Sankey, Phys. Stat. Solidis B, 223 (2002) 90.
26 J.P. Lewis, Th. E. Cheatham III, E.B. Starikov and O.F. Sankey, J. Phys. Chem. B, 107 (2003) 2581.
27 We note that a similar situation is known in electron transfer theories, where solvent reorganization brings into resonance the electronic states of donor and acceptor centres.
28 R.A. Marcus, J. Chem. Phys., (1956) 966; Rev. Mod. Phys., 65 (1999) 599.
29 R. Bruinsma, G. Gruener, M.R. D'Orsogna and J. Rudnick, Phys. Rev. Lett., 85 (2000) 4393.
30 A. Garg, J. N. Onuchic and V. Ambegaokar, J. Chem. Phys., 83 (1985) 4491.
31 V. May and O. Kühn, Charge and Energy Transfer Dynamics in Molecular Systems, WI-VCH, Weinheim, 2004.
32 D. Xu and K. Schulten, Chem. Phys., 182 (1994) 91.
33 A.J. Leggett, S. Chakravarty, A.T. Dorsey, M.P.A. Fisher, A. Garg and W. Zwerger, Rev. Mod. Phys., 59 (1987) 1.
34 U. Weiss, Quantum Dissipative Systems, Vol. 10 of Series in Modern Condensed Matter Physics, World Scientific, 1999.
35 S. Tornow, N.-H. Tong and R. Bulla, Europhys. Lett., 73 (2006) 913.
36 L. Mühlbacher and R. Egger, Chem. Phys., 296 (2004) 193.
37 R. Gutierrez, S. Mandal and G. Cuniberti, Nano Lett., 5 (2005) 1093.
38 R. Gutierrez, S. Mandal and G. Cuniberti, Phys. Rev. B, 71 (2005) 235116.
39 B. Xu, P. Zhang, X. Li and N. Tao, Nano Lett., 4 (2004) 1105.
40 H. Wang, J. P. Lewis and O. Sankey, Phys. Rev. Lett., 93 (2004) 016401.
41 G. Cuniberti, L. Craco, D. Porath and C. Dekker, Phys. Rev. B, 65 (2002) 241314(R).
42 J. Lehmann, S. Kohler, V. May and P. Hänggi, J. Chem. Phys., 121 (2004) 2278.
43 S. Kohler, J. Lehmann and P. Hänggi, Phys. Rep., 406 (2005) 397.
44 D. Segal, A. Nitzan, W. B. Davis and M.A. Ratner, J. Phys. Chem., 104 (2000) 2790; D. Segal and A. Nitzan, Chem. Phys., 281 (2002) 235.

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 20

Electronic structure theory of DNA: from semi-empirical theory of the π-stack to ab initio calculations of the optical conductivity

Robert G. Endres[1,2], Arnd Hübsch[3], Daniel L. Cox[3] and Rajiv R. P. Singh[3]

[1]*Department of Molecular Biology, Princeton University, Princeton, NJ 08544-1014, U.S.A.*
[2]*NEC Laboratories America, Inc., Princeton, NJ 08540 U.S.A.*
[3]*Department of Physics, University of California, Davis, CA 95616, U.S.A.*

Abstract

In addition to DNA's fundamental role in genetics, its self-assembly property based on Watson–Crick base pairing has turned the DNA molecule into a potential component for nanoscale circuits. While the one-dimensional DNA π-stack has long been speculated to support electrical current, a final answer, despite recent multi-disciplinary efforts, has proved difficult. Biological DNA is nothing without its environment; the DNA environment system is a soft poly-electrolyte solution whose structural details and electronic properties depend on humidity, sequence and type of counter-ions. In this chapter we summarize our theoretical approaches [1,2]. We employ a hybrid of semi-empirical Hückel and Slater–Koster theories for describing the electronic π-orbitals with parameters obtained from fitting to *ab initio* density functional theory (DFT) eigenvalues based on the SIESTA package. We use this model to study the dependence of the base-pair couplings on structural parameters and temperature. We then turn to DFT-based calculations of the optical conductivity of realistic solvated DNA segments obtained from molecular dynamics simulations and show that they give a reasonable description of contactless optical measurements of DNA molecules at low and high frequencies.

20.1 INTRODUCTION

Charge transport along DNA molecules has recently attracted a lot of attention due to its potential role in biology, especially in damage and repair processes [3,4], as well as due to its potential use in nanotechnology [5,6]. The excitement surrounding DNA in the

nanotechnology context is fueled by the potential of employing the fundamental Watson–Crick base pairing for self-assembly of electronic circuits, bypassing the need for finding ways to extend the costly microlithography methods which are near the limit of semi-classical physics. Before realization of these ideas, the electronic properties of DNA have to be thoroughly understood, in particular since experiments do not give a conclusive picture of DNA's conductance properties (for a recent review, see [7]). The difficulties can be attributed to the uncertain character of DNA–electrode contacts and the generally unknown conformation and environment of the DNA. However, the electronic properties of DNA are intrinsically complicated as illustrated in Fig. 20.1. Indeed, as a strongly fluctuating molecule with electronic properties and conformation depending crucially upon the environment, DNA is a great system for developing new methods in electronic structure suitable for the nanoworld.

In this article we summarize two quite different approaches to the electronic structure of DNA. The first is based on elementary Hückel theory well known from successfully describing aromatic benzene-type molecules, as well as on Slater–Koster theory which was originally used for calculating the electronic structure of simple crystals. We combine both to understand the base π-orbitals, which were long thought to delocalize along the helix to

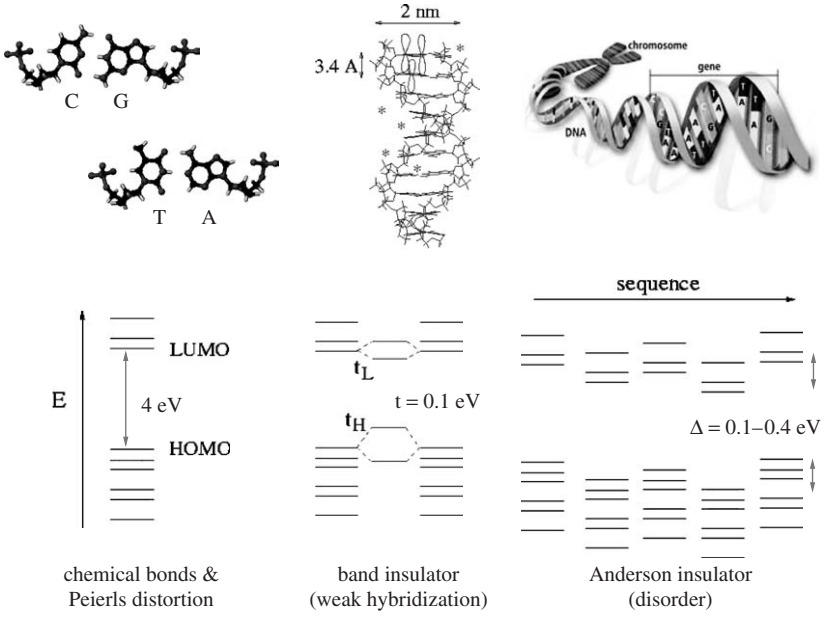


Fig. 20.1. The different length scales of DNA with their characteristic electronic features. At the single base scale, the occupied π and unoccupied π^* orbitals are separated by the HOMO–LUMO gap (~ 4 eV), which is partially caused by the single and double atom–atom bonds reminiscent of the Peierls transition. At the scale of a few base pairs, the base orbitals weakly couple ($t \sim 0.1$ eV) leading to a band insulator picture [8]. In particular, the coupling of the HOMOs (t_H) and of the LUMOs (t_L) is important for hole and electron transport, respectively. At longer scales (10–100 base pairs), the energetic disorder ($\Delta > 0.1$ eV) originating from DNA sequence, backbone, water and counter-ions, and structural distortions cause Anderson localization in the quasi-one-dimensional system [9,10] (see Color Plate 28).

cause wire-like electronic behavior [11]. In particular, we are interested in the vibrational motion of base pairs which, contrary to previous results, causes only a very weak temperature dependence of the insulating energy gap. Our second approach uses *ab initio* density functional theory (DFT) in combination with state-of-the-art molecular dynamics (to generate configurations of the fluctuating molecule) in a study of B-DNA tetramer segments to understand how the negatively charged DNA backbone, the neutralizing counter-cations and water molecules are influencing the electronic states near the Fermi energy. In particular, we compute the frequency-dependent optical conductivity, a contactless probe of the electronic states, and compare with recent measurements displaying an anomalous absorption at low (10–100 meV) energies. While most of the absorption is due to intra-base dipole transitions at and above the π–π^* energy gap as expected, we find that the interesting low-energy absorption is partially caused by water molecules interacting with the open DNA ends (and hence DNA breaks and nicks) possibly causing doping at room temperature. This result is not unexpected, since evidence of thermally activated hopping has been found experimentally [12,13] and discussed previously theoretically [14].

20.2 *AB INITIO* DENSITY FUNCTIONAL THEORY

In order to get a realistic picture of the electronic structure of DNA, we use the localized basis set DFT code SIESTA [15]. Since its aim is efficiency and large systems, it uses Troullier–Martins norm-conserving pseudo-potentials [16] in the Kleinman–Bylander form [17], and a basis set of pseudo-atomic orbitals of double-ζ form optionally including polarization orbitals (DZP). The first-ζ orbitals closely resemble atomic orbitals and are obtained by the method of Sankay and Niklewski [18]. The second-ζ orbitals mainly constitute corrections and are constructed in the split-valency philosophy well known from quantum chemistry [19]. The exchange-correlation functional is described by the generalized gradient approximation (GGA) in the version by Perdew, Burke and Ernzerhof [20]. This semi-local approximation has been shown to describe the important hydrogen bonds sufficiently well.

20.3 MODELING DNA π-ORBITALS WITHIN SLATER–KOSTER–HÜCKEL THEORY

We were motivated to look at the question of fluctuation effects upon the DNA gap by the experimental work of Porath *et al.* [8] who observed a strongly temperature-dependent conduction gap for single DNA molecules contacted to platinum electrodes. This question can be addressed with the semi-empirical modeling of the p states. As in elementary Hückel theory, we restrict ourselves to just atomic p_z orbitals of the base pairs, where the z-axis can be defined locally by its base normal. Generally, two p_z orbitals from different base pairs couple by ppσ and ppπ interactions, which we model with semi-empirical Slater–Koster theory [21,22]

$$V_{ppx} = \eta_{ppx} \frac{\hbar^2}{m d_o^2} \mathrm{e}^{-d/R_c}, \qquad x = \sigma, \pi \tag{1}$$

Note that $\eta_{pp\sigma} > 0$ and $\eta_{pp\pi} < 0$ have opposite signs due to the two possible ways of interaction between the positive and negative lobes of two p_z-orbitals. Parameters d and m are distance and electron mass, and $\hbar^2/(md_o^2) = 7.62$ eV for $d_0 = 1$ Å. This reduces the distance dependence to the exponential factor. The exponential distance cut-off R_c and η_{ppx} are parameters to be determined by fitting. The formula for the atomic interaction matrix elements can be found in Appendix A, Eq. (12). The fitting procedure to realistic *ab initio* calculations can be found directly afterward in Appendix B. Very similar approaches based on Slater–Koster theory have been used in Refs [23,24].

The important inter-base pair electronic couplings t_{12} between two molecular orbitals (MOs) 1 and 2 of adjacent base pairs are obtained as follows. From DFT calculations of single, methylated base pairs with the local DZP basis set we get an expression of the MOs in terms of linear combinations of atomic orbitals (LCAO). We only use the first-ζ coefficients of p_z symmetry. The second-ζ coefficients are generally ~ 80 per cent smaller and are neglected. Let the number of coefficients c_i^1 and c_i^2 be N and M, respectively. If H^{12} is the $N \times M$ Hamilton coupling matrix calculated from Slater–Koster theory, and if we assume orthogonal MOs, we can write

$$t_{12} = \sum_i^N \sum_j^M H_{ij}^{12}\, c_i^1\, c_j^2 \tag{2}$$

A more general formula for non-orthogonal MOs can be found in Ref. [25]. However, from our first-principles calculations, we found the overlap integral S_{12} to be generally small ($0.001 < S_{12} < 0.01$) and we therefore neglected its contributions.

Equation (2) constitutes a straightforward way to calculate inter-base-pair electronic couplings for a variety of spatial and angular separations. Let ϕ and u describe the deviations from equilibrium twist angle and base-pair separation. The matrix H^{12} depends on ϕ and u through the Slater–Koster form of the matrix elements. The electronic coupling between the frontier orbitals as a function of the variables ϕ and u is shown in Fig. 20.2. One variable is held fixed at the equilibrium value, while the other one is varied. Consider first the variation with the angle ϕ (panel (a)). At $\phi = -36°$, i.e., when the two base pairs are aligned perfectly parallel, the coupling is maximal and positive. Let us focus on a GG dimer. At perfect alignment all (19) p_z-orbitals are right on

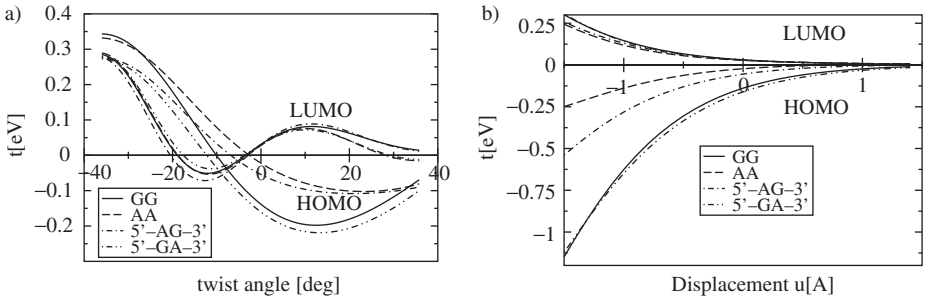


Fig. 20.2. Electronic coupling of base-pair dimer as a function of relative twist angle (a) and separation (b).

top of each other, and the transfer integral is approximately given by $t \sim \sum_i H_{ii}^{12} (c_i)^2$ ($c_i := c_i^1 = c_i^2$) where $H_{ii}^{12} > 0$ is solely due to $pp\sigma$ interaction. Note the interesting property that there are sign changes of $t(\phi)$ with ϕ. These provide a direct explanation for the local minima of $|t|$ observed in earlier *ab initio* studies [26,27]. In B-DNA, essentially all the interaction matrix elements are positive (mainly $pp\sigma$) (see Fig. 8 in Ref. [7]), and this is true for arbitrary twist angles ϕ. Since in the vicinity of the equilibrium, twist angle $\phi \approx 0$ the transfer integral is approximately given by $t \sim \sum_{i \neq j} H_{ij}^{12} \, c_i^1 \, c_i^2$, the sign change of t can only come from the signs of the LCAO coefficients (phases of the p_z-orbitals). This can lead to positive or negative t, or to a complete cancellation.

At finite temperature the base pairs of DNA will fluctuate about their equilibrium positions. We assume that the energy gap $\Delta_{\text{single bp}}$ between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) of a single base pair is temperature independent. Using our model, we are interested in the temperature dependence of the DNA π–π^* band gap

$$\Delta_{\text{DNA}}(T) \approx \Delta_{\text{single bp}} - 2(<t_{\text{HUMO}}>(T) + <t_{\text{LUMO}}>(T)) \tag{3}$$

with $<t>(T)$ describing the temperature-dependent average electronic coupling between adjacent base pairs. For small changes in the variables, one can assume that these two degrees-of-freedom are independent. We assume further classical harmonic potentials with spring constant K and shear constant S for each base-pair dimer. They can be estimated using the equipartition theorem from classical statistical mechanics and estimates for the standard deviation from molecular dynamics [28],

$$0.375^2 \text{Å}^2 = <u^2> = \frac{k_B T}{K} \tag{4}$$

$$\rightarrow K = 0.184 \, \frac{\text{eV}}{\text{Å}^2} = 294.8 \, \frac{pN}{\text{Å}} \tag{5}$$

$$7.5^2 \text{deg}^2 = <\phi^2> = \frac{k_B T}{S} \tag{6}$$

$$\rightarrow S = 4.6 \times 10^{-4} \, \frac{\text{eV}}{\text{deg}^2} \tag{7}$$

With these estimates for K and S the average electronic coupling can be calculated

$$<t>(T) = \frac{\int \text{d}u\text{d}\phi \; t(u, \phi) \; \text{e}^{-\frac{\beta}{2}(Ku^2 + S\phi^2)}}{\int \text{d}u\text{d}\phi \; \text{e}^{-\frac{\beta}{2}(Ku^2 + S\phi^2)}} \tag{8}$$

where $\beta = 1/(k_B T)$. The transfer integral only depends on the differences $u = u_2 - u_1$ and $\phi = \phi_2 - \phi_1$, where u_i and ϕ_i are the displacement and twist angle of the ith base pair.

The results are shown in Fig. 20.3. Two things are remarkable about this result. First, the temperature dependence is very weak (a few meV change over hundreds of K). Second, harmonic displacement along the helical axis alone can lead to an increase in magnitude of the electronic coupling with increasing temperature T, while twisting motion on the other hand has the opposite effect, leading to a cancellation. This can be understood by looking at Fig. 20.2 (a). Since $t(\phi=0)$ is closer to a local maximum (at least for HOMO states), a broadening of the Gaussian distribution $\sim e^{-\beta S\phi^2/2}$ with increasing temperature about $\phi = 0$ will give more contribution from smaller t. On the other hand, a broadening of the distribution for u in the form $e^{-\beta K u^2/2}$ will lead to more contributions from t with larger magnitude, since $|t(u)|$ increases exponentially with decreasing u.

We now return to discuss the experiment by Porath *et al.* who measured I–V curves of short (30 base-pair long) homogeneous poly(G)–poly(C) DNA molecules displaying band-like semiconducting behavior. More interestingly, the authors observed a strong increase of the voltage gap U_c (the minimal bias voltage above which conduction sets in) with temperature (2–4 eV per 300 K) [8]. If we assume that displacements along the helical axis and the twisting motion are the most important degrees-of-freedom which affect the electronic coupling, our results show clearly that this cannot be due to the band gap. The electronic coupling and, hence, the band gap (via Eq. (3)) allow only a very weak temperature dependence and we predict a temperature-independent U_c due to the cancellation of the two modes. Hence, the strongly temperature-dependent conductance gap found experimentally is not accounted for by vibrational modulation of the inter-base hopping between π orbitals.

20.4 ELECTRONIC STRUCTURE AND OPTICAL CONDUCTIVITY OF WET DNA

The observation of small gap (~ 0.1 eV) activated transport in both DC and AC conductivity motivated us to examine potential sources of low-energy excitations in the combined system of DNA plus its surrounding environment of water and counter-ions. The low-energy

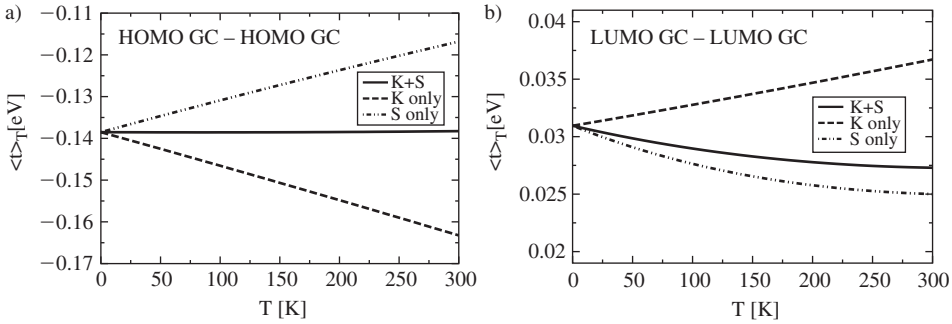


Fig. 20.3. Boltzmann averaged electronic coupling between (a) HOMOs and (b) LUMOs of two G–C base pairs. The effects of a spring (K) and/or a shear (S) constant are shown.

absorption stands in marked contrast to the dominant ultraviolet conductivity peaks. In particular, we are interested in the optical conductivity of DNA which, as a contactless measurement of electronic states removes uncertainties about electrodes. For that purpose, we study several DNA tetramers surrounded by water and counter-ions using a combination of classical molecular dynamics (MD) and DFT based on SIESTA (Section 20.2).

The absorptive part of the optical conductivity is given by the Kubo formula

$$\sigma_{1,\mu}(\omega) = \pi \frac{e^2}{V} \frac{1 - e^{-\beta\hbar\omega}}{\omega} \sum_{n,m} \frac{e^{-\beta\varepsilon_n}}{Z} \delta(\varepsilon_n - \varepsilon_m - \hbar\omega) |\langle n|j_\mu|m\rangle|^2 \tag{9}$$

and depends on the frequency ω. Here, ε_n, ε_m and $|n\rangle$, $|m\rangle$ are the eigenenergies and -states of the many-body Hamiltonian, respectively. V denotes the unit cell volume, and j_μ with $\mu = \perp, ||$ is the component of the current operator perpendicular or parallel to the DNA strand direction. Note that σ_1 is related to the imaginary part of the frequency-dependent dielectric function ε_2 through

$$\varepsilon_{2,\mu}(\omega) = 4\pi \frac{\sigma_{1,\mu}(\omega)}{\omega}$$

In this work, we apply Eq. (9) to the single-particle spectra and eigenfunctions derived from application of SIESTA-based DFT to given DNA configurations; we do not include corrections from time-dependent DFT or scissor operators.

The solvated DNA structures needed for the DFT calculations have been obtained from classical MD simulations with the AMBER7 package [29] (parm98 force field). The DNA tetramers structures (B-DNA 5′-GAAT-3′, 5′-GGGG-3′, 5′-AAAA-3′, and TT-dimer) have been initially charge-neutralized with counter-ions (either by 6 Na or 3 Mg atoms) and solvated with about 600 TIP3P water molecules. We have used a simulation box with periodic boundary conditions of dimensions $38 \times 35.4 \times 25$ Å^3 (z-axis parallel to DNA axis). All simulations have employed the SHAKE method to fix hydrogen-heavy atom distances allowing a 2 fs integration time step. The cut-off for long-range interactions was set to 10 Å. Furthermore, the TT-dimer, a 5′-ATTA-3′ segment from PDB code 1SM5, has been further constrained using the BELLY option to preserve the distorted DNA structure. The simulations were performed at room temperature. After equilibration for 1 ns, the trajectory was recorded every 10 ps over a simulation time of 1 ns, and the resulting 100 snapshots have been used for subsequent analysis.

The structures obtained from the MD simulations were too large for DFT calculations, so we had to reduce the number of atoms. We only kept the first and second solvation shells, i.e., all water molecules within a 4.7 Å radius resulting in approximately 30 water molecules per nucleotide. Finally, we incorporated about 1000 atoms in our DFT calculations. For DNA we have used a double-ζ basis set except for phosphorus and the counter-ions for which the polarization orbitals were also included. For the surrounding water molecules (about 250) we only used the minimal, single-ζ basis set. However, the computations were still very time consuming. In particular, the evaluation of the optical conductivity was very expensive due to the large number of needed dipole matrix elements. To speed up the computations we restricted the number of included unoccupied states which only affect the high-energy spectra above 5 eV. All calculations were done at room temperature.

In order to identify the energetically important states near the Fermi energy, we projected the density of states (DOS) on the atomic p_z-orbitals of the base pairs and on the orbitals of the water molecules. We also calculated the projected DOS of the counter-ions, phosphates, and the sugars. However, we only found contributions of the p_z-orbitals and the water molecules near the Fermi level. A typical projected DOS of structure snapshot of a 5′-GAAT-3′ sequence with Mg counter-ions is shown in Fig. 20.4. As seen in panel (a), there is a π–π^* gap of about 2.7 eV, which is somewhat smaller than the value of a single base. However, if we project the DOS on a single base we find an intra-base π–π^* gap of about 3.7 eV. It turns out that only such intra-base transitions have large dipole matrix elements. Therefore, we observe optical gaps of the same size (compare Figs. 20.5 and 20.7).

Although we observe a rather big π–π^* gap, the actual HOMO–LUMO gap of about 150 meV is very small [combine panels (a) and (b) of Fig. 20.5]. Thus, electrons could be excited from water states below the Fermi energy into unoccupied p_z-orbitals due to thermal fluctuations. Notice that the water states just below the Fermi energy belong to water molecules near the ends of the sequence (see panels (b) and (c) of Fig. 20.4). Therefore, a small gap between HOMO and LUMO only appears if water molecules can enter the DNA structure on damage sites like ends, breaks, or nicks to contact the DNA bases.

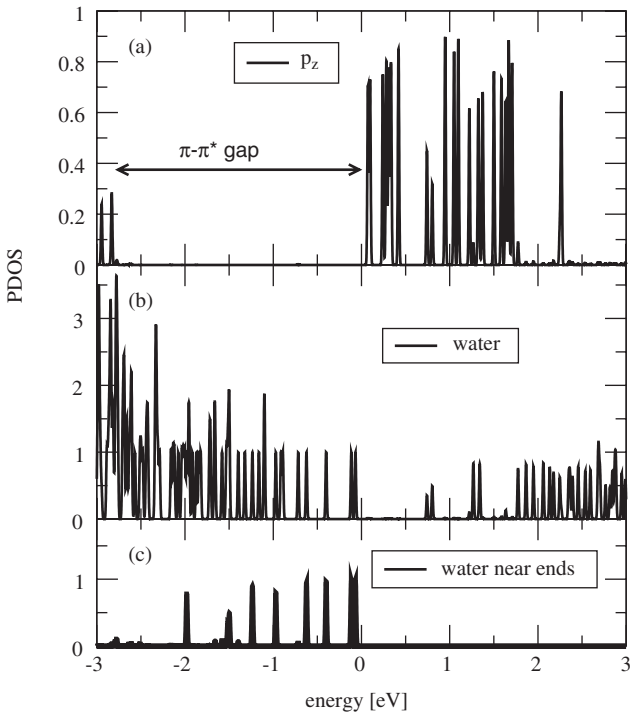


Fig. 20.4. Typical PDOS of a structure snapshot of a 5′-GAAT-3′ sequence with Mg counter-ions. Panel (a) shows the p_z-orbitals of atoms in base pairs. Panel (b) contains the PDOS of all water molecules while panel (c) shows the contributions of water molecules near the sequence ends. The Fermi energy is set to 0.0 eV.

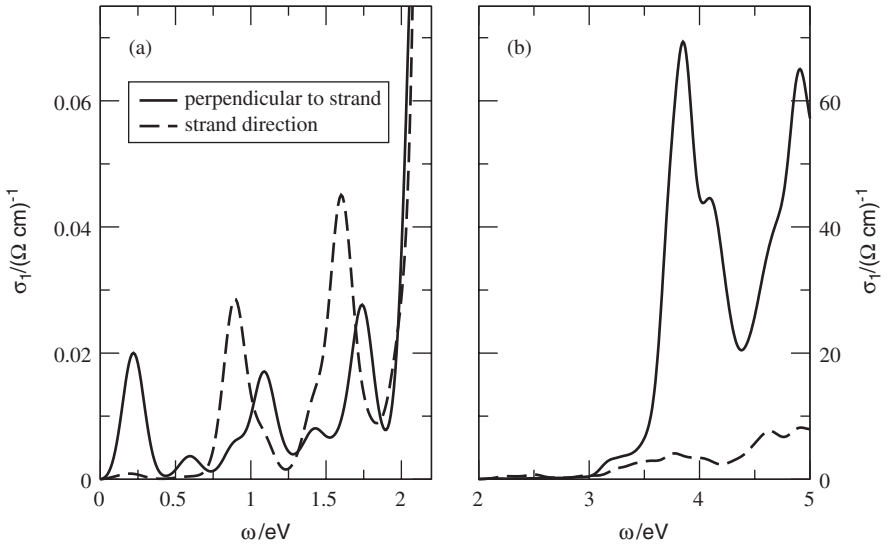


Fig. 20.5. Polarization dependence of the optical conductivity for a structure snapshot of a 5′-GAAT-3′ sequence with Na counter-ions where $\sigma_{1,\perp}(\omega)$ [$\sigma_{1,||}(\omega)$] is plotted with solid (dashed) lines. Panel (a) shows the low-energy features whereas panel (b) presents the range of the π–π^* transition. The theoretical line spectra are broadened with Gaussian functions of width 0.1 eV.

The results of the projected DOS suggest the possibility of a thermally activated doping of the DNA by water states which should also lead to an electronic contribution to the low-frequency absorption. We indeed observe low-energy features in the optical conductivity (Fig. 20.5 (a)). By studying the (fictitious) temperature dependence of these features within SIESTA, we confirm that they are due to transitions between thermally occupied π^* states of the DNA bases. Furthermore, as shown in panel (b), we also observe the pronounced π–π^* transition at 3.8 eV. Whereas the polarization dependence of the low-energy absorption is rather small (see panel (a)), the in-plane π–π^* transition is much more visible if the electric field is perpendicular to the DNA strand direction. A less pronounced anisotropy of the optical conductivity has also been found in Ref. [30].

To take into account the dynamical character of the environment we average the results for the optical conductivity over up to nine structure snapshots from the MD simulations as shown in Fig. 20.6. As discussed above, the low-energy features of the optical conductivity result from a doping of the DNA by water states. Therefore, they are strongly affected by environment fluctuations and we only obtain some smeared intensity in the low-energy range of the averaged conductivity. In contrast, the intrinsic π–π^* gap is much less affected by dynamical fluctuations leading to well-defined structures in that range.

During to the intrinsic character of the π–π^* transition, the averaged optical conductivity in the range 2–5 eV is barely affected by different counter-ions (Fig. 20.7 (a)). To show that the main structure in the spectra around 3.7 eV really has to be interpreted as the intra-base π–π^* transition we also plot the optical conductivity of dry protonated

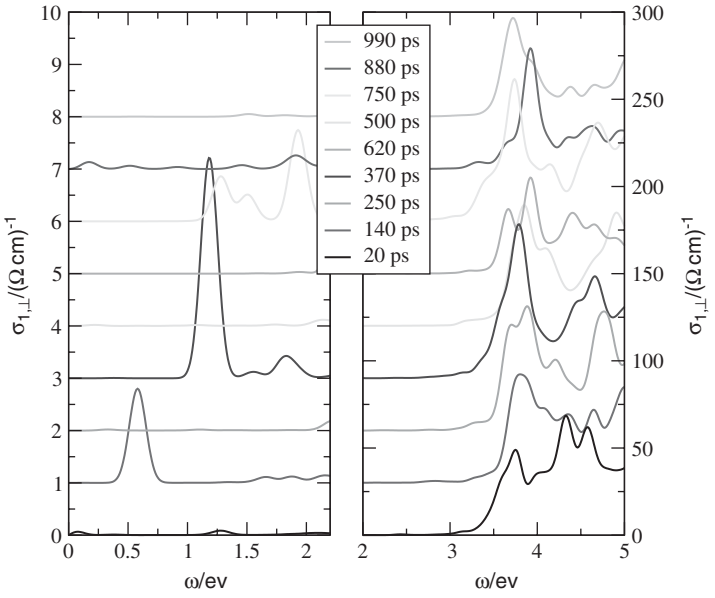


Fig. 20.6. Optical conductivity for several structure snapshots of a 5′-GAAT-3′ sequence with Na counter-ions spread over the MD simulation time of 1 ns (see Color Plate 29).

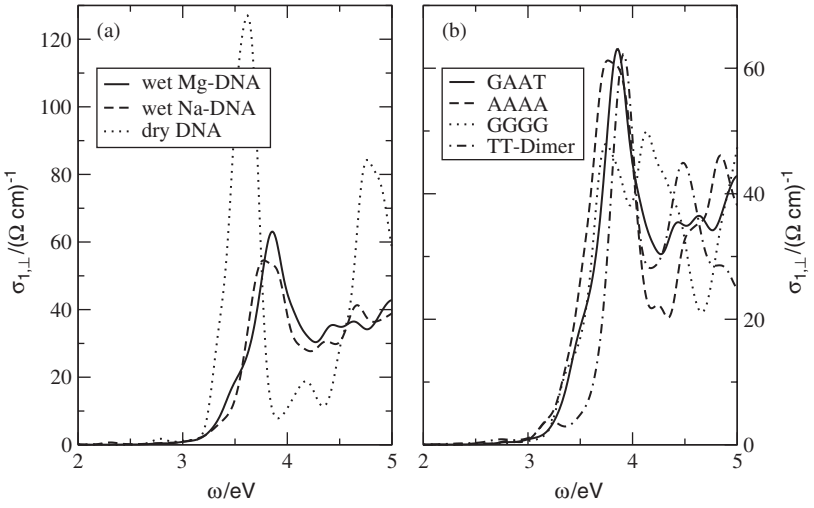


Fig. 20.7. Robustness of the π–π* transition regarding environment and DNA sequence. Panel (a) shows the effects of the surrounding water molecules and counter-ions on the optical conductivity of a 5′-GAAT-3′ sequence. The results for different B-DNA sequences in a wet environment containing Mg counter-ions are compared in panel (b). The spectra for the wet 5′-GAAT-3′ (5′-AAAA-3′, 5′-GGGG-3′, and TT-dimer) sequences have been averaged over the results of nine (five) structure snapshots. The theoretical line spectra are broadened with Gaussian functions of width 0.1 eV.

B-DNA in Fig. 20.7 (a). Although the spectra of wet and dry B-DNA considerably differ in intensities, the peak positions agree quite well. Furthermore, as seen from Fig. 20.7 (b), peak position and shape of the π–π^* transition depend only weakly on sequence, whereas remarkable differences are observed in the range 4–5 eV.

In Fig. 20.8 we compare our results with recent optical experiments [31] and find nice agreement for the range of the π–π^* transition above 10 000 cm^{-1}. However, we find pronounced polarization dependence of the conductivity (see Fig. 20.5) whereas the experimental spectra are almost isotropic. The latter can be explained by substantial variations in the axis of the macroscopically orientated DNA duplex [31].

The spectra below 500 cm^{-1} are also affected by vibrational modes of the double helix structure and the surrounding water molecules, which are not included in our DFT calculations. In particular, the optical conductivity below 100 cm^{-1} has been interpreted in Ref. [32] as a pure water dipole relaxation. However, our DFT calculations clearly show an electronic conductivity even at very low frequencies (see Fig. 20.8). As already discussed above, the low-frequency conductivity results from a doping of the DNA by water molecules near the bases. Therefore, the theoretically observed conductivity at low frequencies, which is approximately an order of magnitude smaller than the experimental values, would be further increased if more water could enter the DNA structure on additional damage sites like ends, breaks, or nicks. Our results may also be relevant to 'bubbles' of DNA, which are locally denaturated by thermal fluctuations or active proteins. For such bubbles, water can presumably enter the base region during the time scale of the bubble opening.

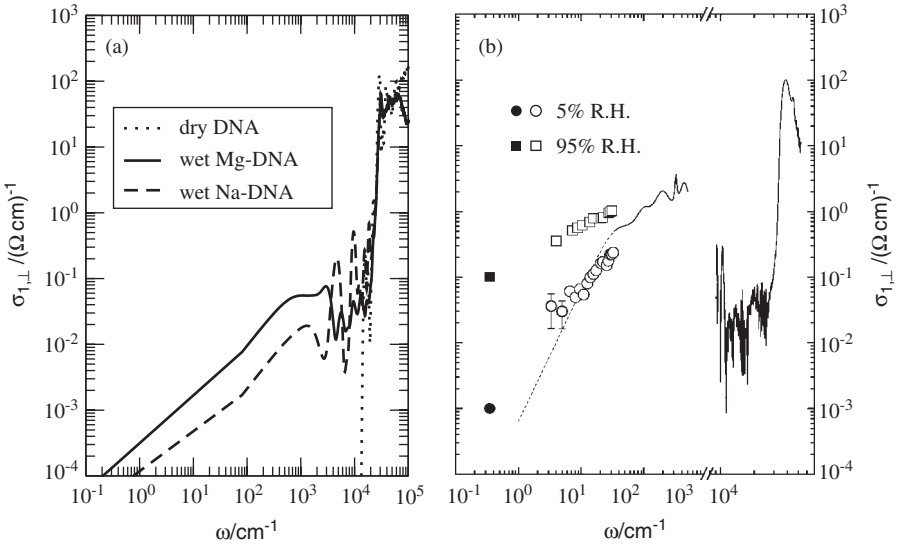


Fig. 20.8. Panel (a) is a log–log plot of the optical conductivity from Fig. 20.7, panel (a). Panel (b) shows experimental data for DNA in a 5 per cent and 95 per cent relative humidity environment taken from Ref. [31].

20.5 CONCLUSIONS

In this chapter we have outlined two approaches to the study of DNA as a dynamical, environmentally dependent electronic molecule.

The first is based on semi-empirical Slater–Koster–Hückel theory to model the electronic structure of the π-stack as a function of some helical parameters. The main advantage lies in the extremely short run times (seconds) which, in principle, allows the method to be applied to very large-scale electronic structure calculations of long DNA molecules and DNA bundles. The second approach uses DFT and thus allows us to realistically calculate the effects of the DNA environment. We calculated the contactless optical conductivity and proposed an easy mechanism for doping based on water molecules interacting with the DNA ends, breaks, and nicks. These detailed calculations and their accuracy come at a price. The calculation of the optical conductivity for a given structure snapshot from classical MD took about five days on AMD's Opteron CPU and 2–3 GB RAM. Obviously, the results open many questions which will stretch current computational limits. For example, since DNA bubbles typically contain ten or more bases, carrying out a study of these interesting biological conformations would far outstrip current accessible computer capacity.

While we have developed these approaches with an interest in DNA, they can of course be extended to the study of other fluctuating molecules with strong environmental modulation. We hope to further develop the MD/DFT interface in particular, and extend the methods to other systems such as charge transfer proteins or conducting polymers.

20.6 ACKNOWLEDGEMENTS

DLC and RRPS acknowledge the support of the Center for Biophotonics Science and Technology (NSF grant No. PHY 0120999). DLC and AH acknowledge the support of the U.S. Department of Energy, Office of Basic Energy Science, Office of Basic Energy Sciences, Division of Material Research. AH is grateful for the support of the Deutsche Forschungsgemeinschaft (grant No. HU 993/1-1).

20.7 APPENDIX A. SLATER–KOSTER INTER-ATOMIC MATRIX ELEMENTS

In the following we derive a general formula for the calculation of interaction matrix elements between two atomic p_z-orbitals belonging to adjacent base pairs. In general, the orbitals point in different directions, since the local z-axis is different for each base, mainly due to a propeller twist angle, in the case of A-DNA also because of a large inclination angle. The local z-axis is obtained by fitting a plane to each single base whose coordinates were obtained from DNA X-ray diffraction data. The normal of each base is used as the direction for all p_z-orbitals within that base. Given their directions, we first expand the orbitals p_z, p'_z in a global coordinate system with the z-axis pointing in the direction of the helix

$$\begin{aligned} p_z &= \alpha\tilde{p}_x + \beta\tilde{p}_y + \gamma\tilde{p}_z \\ p'_z &= \alpha'\tilde{p}_x + \beta'\tilde{p}_y + \gamma'\tilde{p}_z \end{aligned} \tag{10}$$

where the expansion coefficients are the direction cosines. If (l, m, n) is the unit vector pointing from one orbital to the other, the Slater–Koster relations [21]

$$E_{xx} = l^2 V_{pp\sigma} + (1 - l^2) V_{pp\pi} \tag{11}$$

$$E_{yy} = m^2 V_{pp\sigma} + (1 - m^2) V_{pp\pi}$$

$$E_{zz} = n^2 V_{pp\sigma} + (1 - n^2) V_{pp\pi}$$

$$E_{xy} = lm\,(V_{pp\sigma} - V_{pp\pi})$$

$$E_{xz} = ln\,(V_{pp\sigma} - V_{pp\pi})$$

$$E_{yz} = mn\,(V_{pp\sigma} - V_{pp\pi})$$

can be used to obtain the following formula for the interaction matrix element

$$\begin{aligned} V = {} & \alpha\alpha' E_{xx} + \beta\beta' E_{yy} + \gamma\gamma' E_{zz} \\ & + (\alpha\beta' + \beta\alpha') E_{xy} + (\alpha\gamma' + \gamma\alpha') E_{xz} \\ & + (\beta\gamma' + \gamma\beta') E_{yz} \end{aligned} \tag{12}$$

The Slater–Koster matrix elements $V_{pp\sigma}$ and $V_{pp\pi}$ are given by Eq. (1) and contain the distance dependence.

20.8 APPENDIX B. PARAMETER FITTING

The Slater–Koster parameters $\eta_{pp\sigma}$, $\eta_{pp\pi}$ and cut-off R_c of Eq. (1) were determined from fitting to DFT data as follows. Starting from the DZP basis set of SIESTA, we first identify the $2p_z$-orbitals. These are the first-ζ orbitals of p_z symmetry. (The second-ζ orbitals correspond to excited atomic states and have rather small contributions to the wavefunctions and matrix elements (~ 20 per cent). We consider the reduced SIESTA Hamiltonian matrix in this basis, which has two diagonal blocks representing the on-site and intra-base-pair couplings and one off-diagonal block representing inter-base-pair couplings. The Slater–Koster matrix elements calculated as outlined in Appendix A are directly fitted to the off-diagonal block matrix elements. The fitting was done with simulated annealing [33]. Finally, since the DFT code uses a global z-axis along the helix, which is different from the local z-axis due to an inclination and a propeller twist angle associated with each base, we turn to parallel reference base pairs for the fitting process. Starting from canonical B-form DNA, we used two $G \cdot C$ base pairs (GG dimer) and set their separation to exactly 3.34 Å. The resulting fitted parameters are $\eta_{pp\sigma} = 5.27$, $\eta_{pp\pi} = -2.26$, and $R_c = 0.87$ Å.

20.9 REFERENCES

1 R.G. Endres, D.L. Cox and R.R.P. Singh, cond-mat/0201404.
2 A. Hübsch, R.G. Endres, D.L. Cox and R.R.P. Singh, Phys. Rev. Lett., 94 (2005) 178102.
3 S.O. Kelly and J.K. Barton, Science, 283 (1999) 375.
4 S.R. Rajski, B.A. Jackson and J.K. Barton, Mutat Res., 447 (2000) 49.
5 N.C. Seeman, Nature, 421 (2003) 427.
6 F.C. Simmel and W.U. Dittmer, Small, 1 (2005) 284.
7 R.G. Endres, D.L. Cox and R.R.P. Singh, Rev. Mod. Phys., 76 (2004) 195.
8 D. Porath, A. Bezryadin, S. De Vries and C. Decker, Nature, 403 (2000) 635.
9 P. Carpena, P. Bernaola-Galván, P. Ch. Ivanov and H.E. Stanley, Nature, 418 (2002) 955.
10 A. Omerzu, M. Licer, T. Mertelj, V.V. Kabanov and D. Mihailovic, Phys. Rev. Lett., 93 (2004) 218101.
11 D.D. Eley and D.I. Splivey, Trans. Faraday Soc., 58 (1962) 411.
12 P. Tran, B. Alavi and G. Grüner, Phys. Rev. Lett., 85 (2000) 1564.
13 K.-H. Yoo, D.H. Ha, J.-O. Lee, J.W. Park, J. Kim, J.J. Kim, H.-Y. Lee, T. Kawai and H.Y. Choi, Phys. Rev. Lett., 87 (2001) 1981.
14 Z.G. Yu and X. Song, Phys. Rev. Lett., 86 (2000) 6018.
15 D. Sánchez-Portal, P. Ordejón, E. Artacho and J.M. Soler, Int. J. Quantum Chem., 65 (1997) 453; E. Artacho, D. Sánchez-Portal, P. Ordejón and A. Garcia, J.M. Soler, Phys. Status Solidi (b), 215 (1999) 809; P. Ordejón, E. Artacho and J.M. Soler, Phys. Rev. B, 53 (1996) R10441.
16 N. Troullier and J.L. Martins, Phys. Rev. B, 43 (1991) 1993.
17 L. Kleinman and D.M. Bylander, Phys. Rev. Lett., 48 (1982) 1425.
18 O.F. Sankey and D.J. Niklewski, Phys. Rev. B, 40 (1989) 3979.
19 S. Huzinaga, J. Andzelm, M. Klobukowski, E. Radzio-Andzelm, Y. Sakai and H. Tatewaki (Eds.), Gaussian basis sets for molecular calculations, Elsevier Science Pub. Co., New York, 1984.
20 J.P. Perdew, K. Burke and M. Ernzerhof, Phys. Rev. Lett., 77 (1996) 3865.
21 J.C. Slater and G.F. Koster, Phys. Rev., 94 (1954) 1498.
22 W.A. Harrison, Electronic Structure and the Properties of Solids, Dover Publications, Inc., New York, 481 (1989) 48.
23 A.A. Stuchebrukhov, J. Chem. Phys., 118 (2003) 7898.
24 T. Cramer, S. Krapf and T. Koslowski, J. Phys. Chem. B, 108 (2004) 11812.
25 M.D. Newton, J. Phys. Chem., 90 (1986) 3734.
26 H. Sugiyama and I. Saito, J. Am. Chem. Soc., 118 (1996) 7063.
27 M.L. Zhang, M.S. Miao, V.E. van Doren, J.J. Ladik and J.W.J. Mintmire, J. Chem. Phys., 111 (1999) 8696.
28 M.A. Young, G. Ravishanker and D.L. Beveridge, Biophys. J., 73 (1997) 2313.
29 D.A. Case, T.E. Cheatham, III, T. Darden, H. Gohlke, R. Luo, K.M. Merz, Jr., A. Onufrier, C. Simmerling, B. Wanz and R. Woods, J. Computat. Chem., 26 (2005) 1668–1688; D.A. Pearlman, D.A. Case, J.W. Caldwell, W.R. Ross, T.E. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel and P. Kollman, Comp. Phys. Commun., 91 (1995), http://amber.scripps.edu.
30 F.L. Gervasio, P. Carloni and M. Parrinello, Phys. Rev. Lett., 89 (2002) 108102.
31 E. Helgren, A. Omerzu, G. Grüner, D. Mihailovic, R. Podgornik and H. Grimm, cond-mat/0111299.
32 M. Briman, N.P. Armitage, E. Helgren and G. Grüner, Nano Lett., 4 (2004) 733.
33 http://www.ulib.org/webRoot/Books/Numerical Recipes/bookc.html.

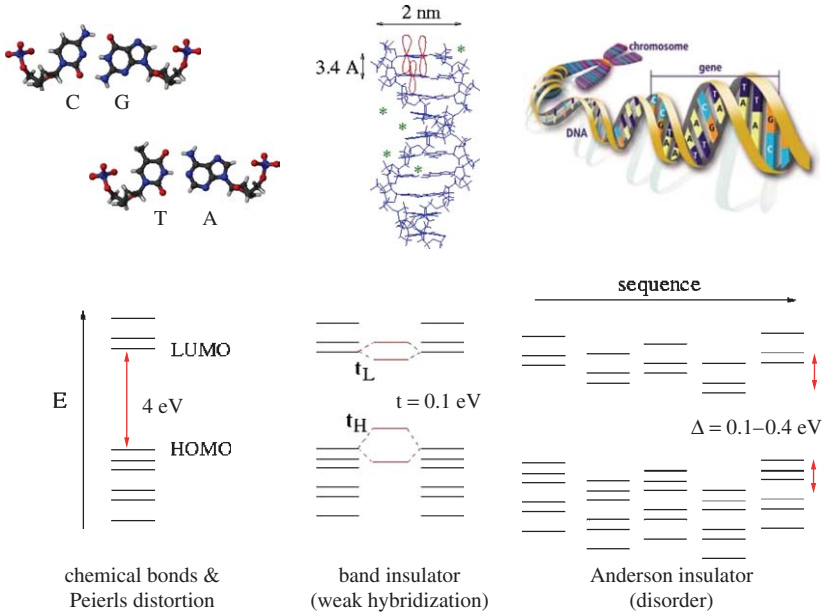


Plate 28. The different length scales of DNA with their characteristic electronic features. At the single base scale, the occupied π and unoccupied π^* orbitals are separated by the HOMO–LUMO gap (~ 4 eV), which is partially caused by the single and double atom–atom bonds reminiscent of the Peierls transition. At the scale of a few base pairs, the base orbitals weakly couple ($t \sim 0.1$ eV) leading to a band insulator picture [8]. In particular, the coupling of the HOMOs (t_H) and of the LUMOs (t_L) is important for hole and electron transport, respectively. At longer scales (10–100 base pairs), the energetic disorder ($\Delta > 0.1$ eV) originating from DNA sequence, backbone, water and counter-ions, and structural distortions cause Anderson localization in the quasi-one-dimensional system [9,10] (illustration appears on page 394 of this volume).

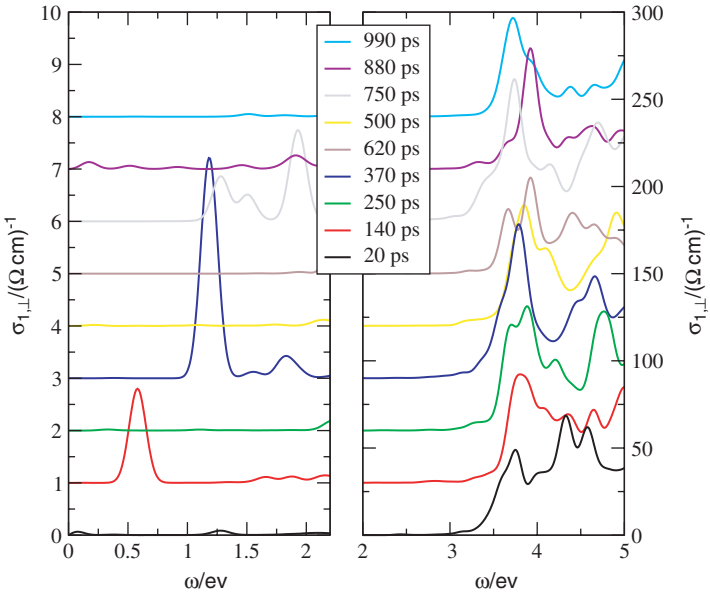


Plate 29. Optical conductivity for several structure snapshots of a 5′-GAAT-3′ sequence with Na counter-ions spread over the MD simulation time of 1 ns (illustration appears on page 402 of this volume).

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 21

Electronic transport and localization in short and long DNA

H. Wang [1], R. Marsh [1], J. P. Lewis [1] and R. A. Römer [2]

[1]*Department of Physics and Astronomy, Brigham Young University, Provo, UT 84602-4658, U.S.A.*
[2]*Department of Physics and Centre for Scientific Computing, University of Warwick, Coventry CV4 7AL, U.K.*

Abstract

The question of whether DNA conducts electric charges is intriguing to physicists and biologists alike. The suggestion that electron transfer/transport in DNA might be biologically important has triggered a series of experimental and theoretical investigations. Here, we review recent theoretical progress by concentrating on quantum chemical, molecular dynamics-based approaches to short DNA strands and physics-motivated tight-binding transport studies of long or even complete DNA sequences. In both cases, we observe small, but significant differences between specific DNA sequences such as periodic repetitions and aperiodic sequences of AT bases, λ - DNA, centromeric DNA, and promoter sequences as well as random-ATGC DNA.

21.1 INTRODUCTION

Charge transfer in DNA is currently the subject of intense theoretical and experimental investigations [1–4]. DNA, which is the blueprint of life, is being considered as a molecular wire in a new generation of electronic devices and computers. However, its electronic properties are elusive and remain controversial. Despite the current debate, the subject is far from new. Soon after Watson and Crick discovered the double-helix structure of DNA [5], Eley and Spivey were the first to suggest that DNA could serve as an electronic conductor [6] . The notion of a molecular wire is thought to apply to the DNA double helix because of its π - electron (the π - way) system of bases stacked upon each other. More recently, Barton and colleagues [7] measured the fluorescence of an excited molecule and found that it no longer emitted light when attached to DNA. Their results suggested that this 'fluorescence quenching' was due to the charge on the excited donor molecule leaking along the length of the DNA to a nearby acceptor molecule.

Other extensive experimental and theoretical work over the past decade has led to substantial clarification of charge-transfer mechanisms in DNA [7–20]. The dominant mechanisms appear to be both short-range quantum mechanical tunnelling and long-range thermally activated hopping. Guanine has the highest occupied molecular orbital (HOMO) level of the four bases, and can act as a trap for holes. Experiments on repeats of this base are used to investigate long-range hopping, and models have been developed to clarify the long-range hopping data in G-repeats [14]. Charge transport in DNA is also made more complex because of the influences of the local environment, such as counterions, thermal vibrations, contact resistance, and sequence variability, which are difficult to control [21–25]. The charge-transfer mechanisms in DNA and/or whether DNA is a good conductor or not remains somewhat unsettled. Indeed, theory is of great help in understanding these phenomena, but given the computational cost of full-scale calculations on realistic DNA systems, theoretical efforts to date have mostly been limited to small- and medium-size model systems [26–28], to dry DNA molecules [23,29,30], or to larger systems using model Hamiltonians [31–42] and semi-empirical studies [22,43–53].

In this review, we shall first focus on the use of quantum chemical methods which can treat smaller, but atomistically correct segments of DNA in Section 21.2. After an introduction to the construction of the DNA molecules and the density-functional-based methods in Sections 21.2.1 – 21.2.3, we then present results, many of which are new, in Sections 21.2.4 to 21.2.5. In Section 21.3, we use the lessons learned from the atomistic approach and now study an effective and necessarily rather coarse-grained Hamiltonian model of DNA to reveal the interplay of sequence fidelity and transport. Again, models, methods and DNA sequences are introduced in Sections 21.3.1 – 21.3.3. Sections 21.3.4 and 21.3.5 include the obtained results. We conclude and summarise in Section 21.4.

21.2 QUANTUM CHEMICAL METHODS FOR *SHORT* DNA STRANDS

Within a density-functional-based local orbital tight-binding-like formalism, more complex problems can be investigated with a modest decrease in accuracy. This is particularly useful where a quantum mechanical description is important to the investigated system's fundamental chemistry, yet where a smaller model system would inadequately describe the proper physical environment. With the increase in computational power, great effort has been made by the electronic-structure community to optimize the performance of quantum mechanical methods. Calculating larger systems without making stringent approximations has only been possible within the past few years. In this chapter, we theoretically investigate the electronic states of model DNA structures as the molecule undergoes classical thermal motion at room temperature by means of marrying classical molecular dynamics simulations with an electronic structure density-functional method. We investigate the dynamics of the DNA structure and its impact on the electronic structure. A similar approach was recently used to postulate the charge migration mechanism in DNA, with injected charges being gated in a concerted manner by thermal motions of hydrated counter-ions [17]. Here we study a longer oligonucleotide duplex than in previous studies, and demonstrate with the complete system that its electronic states dynamically localize. The mechanism is an Anderson *off-diagonal* dynamic disorder model similar to the static disorder that leads to localized band-tail states

in amorphous semiconductors [54–57]. The concept of static Anderson localization in DNA has previously been considered by Ladik [58,59]. We show that localization in DNA reaches far deeper in energy than just band-tail states. We demonstrate for the first time this effect in a hydrated poly(dA)–poly(dT) 10 base-pair fragment; this represents one complete turn of the B-DNA double helix.

21.2.1 Generating the poly(dA)–poly(dT) DNA structures

In this chapter, we consider thermal fluctuations of a poly(dA)–poly(dT) DNA 10-mer duplex fragment at room temperature from classical MD simulations; therefore, aperiodic structures of DNA are generated throughout the simulation. With our local-orbital density-functional method, we compare the electronic states of an idealized model periodic canonical B-DNA poly(dA)–poly(dT) DNA structure with those of thermally distorted aperiodic poly(dA)–poly(dT) DNA structures generated from the MD simulation.

Canonical B-DNA 10-base-pair models of poly(dA)–poly(dT) were built into an Arnott B-DNA [60] model using the nucgen DNA builder contained within AMBER 5.0 [61].Classical molecular dynamics trajectories of the B-DNA models, including explicit water and sodium counter-ions, were generated using CHARMM (version c26n1) [62]. Both models were solvated with enough pre-equilibrated TIPSP [62] water to add 12.0 Å to the maximal distance extent of the DNA. Net-neutralizing Na^+ ions [63] were placed off the phosphate oxygen bisector and then minimized (with larger, 5.0 , van der Waals radii) in vacuo prior to solvating the system. Equilibration involved the application of harmonic positional restraints (25.0 kcal/mol^{-2}) and 250 steps of ABNR minimization, followed by 25 ps of MD where the temperature was ramped up from 50 to 300 K in 1 ps intervals. The initial equilibration was performed with the Cornell *et al*. force field [64]. Subsequent equilibration with the force field of Langley [65] involved 250 steps of ABNR minimization followed by 5 ps of MD with position restraints.

All production simulations were performed without any restraints and the force field of Langley. Production simulation was performed for 10 ns with CHARMM (version c26n1) [62] in a consistent manner. This involved constant temperature (300 K, mass = 1 000) [66] and pressure (1 atm, piston mass = 500 amu, relaxation time = 20 ps^{-1}) [67], 2 fs time steps with the application of SHAKE [68] on hydrogen atoms, accurate use of the particle mesh Ewald method [69] (1.0 grid size with sixth-order B-spline interpolation and a Ewald coefficient of 0.34) in rhombic dodecahedral unit cells ($x = y = z$, $\alpha = 60°$, $\beta = 90°$, $\gamma = 60°$), a heuristically updated atom-based pairlist built to 12.0 Å and cut-off at 10.0 Å with a smooth shift of the van der Waals energies. These methods have proven reliable for representing DNA duplex structure [70,71] and the BMS force field very accurately models B-DNA crystal structures [65,72].

After an initial equilibration of an explicitly solvated 10-mer B–DNA poly(dA)–poly(dT) with explicit Na^+ ions, production molecular dynamics simulations (applying an accurate particle mesh Ewald treatment of the electrostatics) were performed for 10 ns. As shown in Fig. 21.1, a plot of the all-atom root-mean-squared deviation over the entire run is rather stable, and although thermal fluctuations are clearly evident, no large-scale distortions of the structure were observed (beyond sugar repuckering, and expected base and backbone fluctuations).

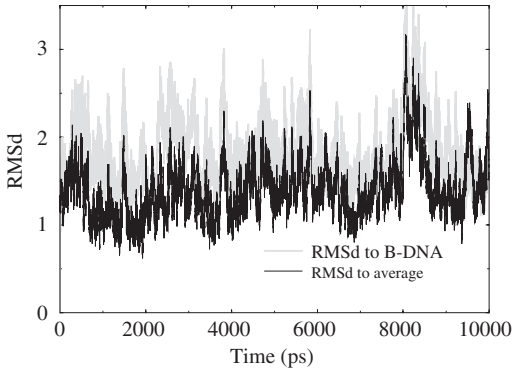


Fig. 21.1. Shown in black and grey are the all-atom best-fit root-mean-squared deviations (in Å) as a function of time compared with canonical B-DNA (grey) and the straight coordinate average structure from the 1.5–2.5 ns portion of the trajectory (at 0.5 ps intervals).

21.2.2 Electronic structure calculations of molecular dynamics snapshots

A stable portion of the trajectory from 1.5–2.5 ns, at 0.5 ps intervals, was analysed further using the FIREBALL DFT methodology [73]. FIREBALL is a first principles tight-binding molecular dynamics (TBMD) simulation technique based on a self-consistent version of the Harris–Foulkes [74,75] functional [76]. In this method, confined atomic-like orbitals are used as a basis set for the determination of the occupied eigenvalues and eigenvectors of the one-electron Hamiltonian. The 'fireball' orbitals, introduced by Sankey and Niklewski [77], are obtained by solving the atomic problem with the boundary condition that the atomic orbitals vanish outside and at a predetermined radius r_c where wavefunctions are set to be zero. This boundary condition is equivalent to an 'atom in the box'and has the effect of raising the electronic energy levels due to confinement. An important advantage of the Sankey and Niklewski basis set is that the Hamiltonian and the overlap matrix elements of the system are quite sparse for large systems, reducing overall computation time. A summary of the method is given in Ref. [73] and references therein. All poly(dA)–poly(dT) DNA atoms, including phosphate groups and backbone atoms, are included in the single-point calculations which contained 10 base pairs (644 atoms). Although the MD simulations are performed with full hydration and counter-ions, we include only 350 water molecules in our electronic structure calculations; this number of molecules represents approximately two solvation layers surrounding the molecule. Adding all water and cation atoms to represent the environment surrounding more correctly the DNA molecule will be the subject of future work.

21.2.3 Quantifying the degree of localization

The phenomena of Anderson localization [54,78] refers to the localization of mobile quantum mechanical entities, such as spin or electrons, due to impurities, spin diffusion or randomness. Anderson localization applied to DNA may come from two distinct

mechanisms, *diagonal* or *off-diagonal* disorder. *Diagonal* disorder-induced localization occurs from variations of the sequence along the base stack, and *off-diagonal* disorder occurs by variations either from bonding between bases along the stack or from hydrogen bonding variations across the double helix. The qualitative physics of localization is described by an Anderson model [54].

$$H = \sum_i \varepsilon_i c_i^\dagger c_i + \sum_{i,j} t_{i,j}\, c_i^\dagger c_j + t_{j,i}\, c_j^\dagger c_i \tag{1}$$

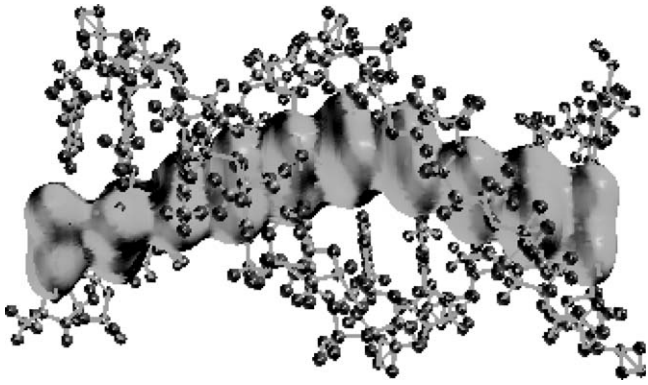
where each molecular orbital (MO) i of a base has energy ε_i and interacts with its nearest neighbour base MO j ($i \neq j$) with a Hamiltonian hopping interaction of t_{ij}. The Anderson model of *diagonal* disorder randomly varies the on-site Hamiltonian matrix elements (diagonal) ε_i [78] and describes the A–T–G–C random sequencing of DNA [31].

Here we focus on B-DNA structures of poly(dA)–poly(dT) in which there exists only one base pair combination A–T; each strand has only a single type of base in its stack. In this system, only *off-diagonal* disorder [79] may occur. The bonds within a single base are strong, but thermal fluctuations coupled with weak π-bonding occurs along the stack and the weak hydrogen bonds across the strands of the DNA double helix allow individual bases significant freedom of movement, including transient base-pair opening and DNA breathing events over millisecond time scales [80] and large fluctuations in the structure [81]. Stochastic fluctuations of the weak bonding modulates the electronic coupling, t_{jj}, between adjacent bases. If the dynamic fluctuations of t_{ij} are large enough, localized electronic states are produced as in an amorphous solid.

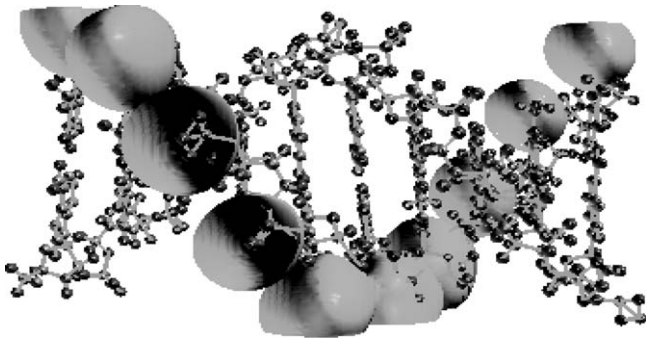
We quantify the spatial extent of an electronic state by defining the number of accessible atoms, W, from the electronic state quantum *entropy*. From a particular state ν, the wavefunction $\psi(\nu)$ has a Mulliken population $p_i(\nu)$ on atom i, which loosely is considered the probability that an electron is in state ν and resides on a particular atom i. The populations are normalized, $\sum p_i(\nu) = 1$. From probability theory, we define a quantum entropy for state ν as, $S(\nu) = -\sum_i p_i(\nu)\ \ln\ p_i(\nu)$. For example, a state ν with equal probabilities over N_0 atoms ($N_0 \leqslant N_{\text{Total}}$), gives an entropy of $\ln N_0$. From Boltzmann's equation, we can determine the number of accessible atoms $W(\nu)$ for electronic state ν as $S(\nu) = \ln W(\nu)$, or $W(\nu) = e^{S(\nu)}$. Our example state with equal probabilities spread over N_0 atoms gives the expected result, $W(\nu) = N_0$. For the complex electronic states of DNA, the number of accessible atoms $W(\nu)$ gives a quantitative, and easily calculable, measure for how many atoms a particular electronic state $\psi(\nu)$ reaches.

21.2.4 Electronic states of a periodic poly(dA)–poly(dT) DNA

To demonstrate that localization is not due to limitations of using localized orbitals, a 10-base pair periodic structure of poly(dA)–poly(dT) was created based on the Arnott B-DNA [60] fibre model. Each base pair is rotated by 36° and translated by 3.38 Å; therefore, 10 base pairs complete one full pitch of the double helix and periodicity is enforced in the program. The population densities for the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) are plotted in



HOMO



LUMO

Fig. 21.2. Population densities for the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) are shown for periodic poly(dA)–poly(dT) DNA (10 base pairs). Both molecular orbitals exhibit very extended and periodic (Bloch-like) states throughout the molecule (see Color Plate 30).

Fig. 21.2. As seen from this figure, both the HOMO and LUMO states exhibit very extended and periodic (Bloch-like) states throughout the molecule. No localization is evident.

21.2.5 Electronic states of sampled poly(dA)–poly(dT) DNA configurations

We now consider results for a single configuration from the MD simulation (labelled step 3 001, the first coordinate set 0.5 ps after a 1.5 ns production simulation). Figure 21.3 shows the number of accessible atoms, $W(\nu)$, for each electronic state at this time step for the dehydrated structure. The $W(\nu)$ for the hydrated DNA structure is shown in Fig. 21.4. For both structures it is important to note that near the HOMO and LUMO, the number of accessible atoms is quite small (<30), demonstrating a large degree of localization for the wavefunctions. This localization extends over several eV and is deeper than just the band-tail states. States further away from the HOMO and LUMO become considerably

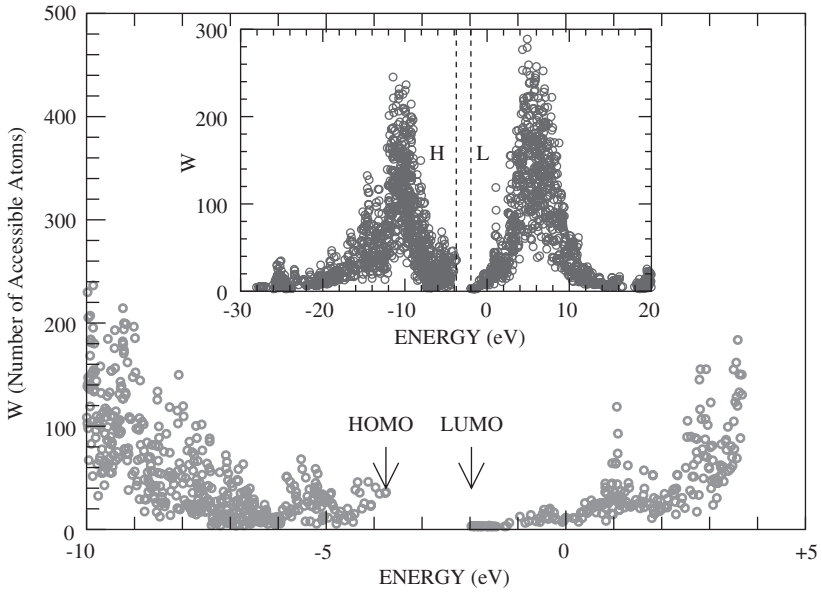


Fig. 21.3. Number of accessible atoms, $W(\nu)$, for each electronic state near the HOMO and LUMO levels. Inset shows number of accessible atoms for all levels. The system contains 10 base pairs of DNA (644 atoms).

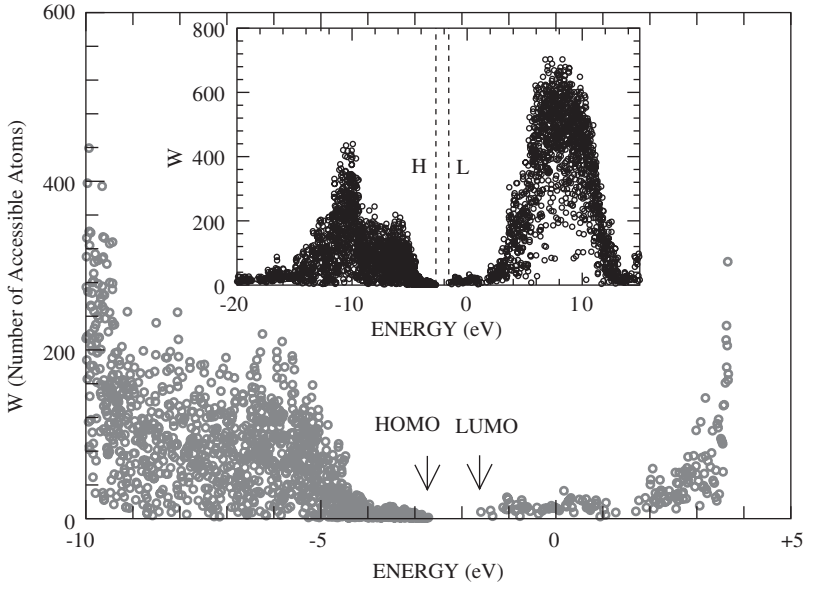


Fig. 21.4. Number of accessible atoms, $W(\nu)$, for each electronic state near the HOMO and LUMO levels. Inset shows number of accessible atoms for all levels. The system contains 10 base pairs of DNA (644 atoms) and 350 water molecules.

delocalized and the number of accessible atoms is much larger. The number of accessible atoms is also small for the lowest energy levels; these deep states consist mainly of 2s levels of oxygen and nitrogen atoms. For the hydrated DNA molecule, the localized states near the HOMO are mainly due to the surrounding water molecules. Just below these water-related localized electronic states are the localized electronic states residing on the DNA bases. This may account for the smaller band gap of the hydrated structure compared with the electronic structure of the dehydrated DNA. Overall, the electronic structures for both the hydrated and dehydrated DNA molecules show remarkable similarities. These results imply that the aquatic environment does not significantly alter DNA's electronic structure. Therefore, we focus our studies on the electronic structures of dehydrated DNA molecules.

The degree of localization for two example band states (1 074 and 614 – larger number implies higher eigenvalue) in the dehydrated DNA structures can be seen in Fig. 21.5 where population density plots of a localized and delocalized state are shown. As more configurations are analysed, we see consistently that the number of accessible atoms for the energy levels near the HOMO primarily consist of around 20 atoms. However, as a function of time, different sets of atoms are involved. To determine where the localization occurs, we compute a residence of each state according to the specific DNA component – adenine base, thymine base, ribose backbone, or phosphate group and determine where the high probability regions are located. Further investigation indicates the residence localization for the highly localized states near the HOMO are contained approximately on single bases in the DNA molecule; adenine for states very near the HOMO and thymine for states slightly lower in energy. This regional population information for the HOMO on adenine is plotted in Fig. 21.6. The more extended states (from approximately 8eV to approximately 18eV below the HOMO) are found to reside throughout the various DNA components.

As the simulation proceeds in time, the residence of the HOMO level moves from base to base along the poly(dA)–poly(dT) system and large jumps in sequence are possible over this 0.5 ps resolution time scale. This fluctuating residency of the HOMO is visualized in Fig. 21.7, which shows population density plots for a series of snapshots at different times (t = 3 001, 3 004, 3 007 and 3 010). The separation between these snapshots is 1.5 ps. Figure 21.8 shows the location of the HOMO for all 100 snapshots where the electronic structure was calculated in this work. The population is localized on different adenine bases as time progresses and appears to chaotically oscillate between one end of the DNA molecule to the other. The HOMO level's localization on one adenine base is traded for localization on another adenine base through the dynamical simulation. Physically, this trading ought to reflect concerted fluctuations assignable to off-diagonal dynamical disorder in a regular homo-oligonucleotide duplex. Based on these results, it is conceivable that electron (hole) transfer will occur as two or more localized MO levels are dynamically trading places. Moreover, this swapping may be gated by thermal fluctuations of hydrated counter-ions, in accordance with the ion-gating transport mechanism proposed in Ref. [17]

Finally, it is of considerable interest to compare our above results to the known literature data on this theme. Specifically, our findings are in parallel with the most recently established dependence of electronic coupling between DNA bases in the stack on DNA conformational states: a diminuation of the coupling between the DNA purine bases due

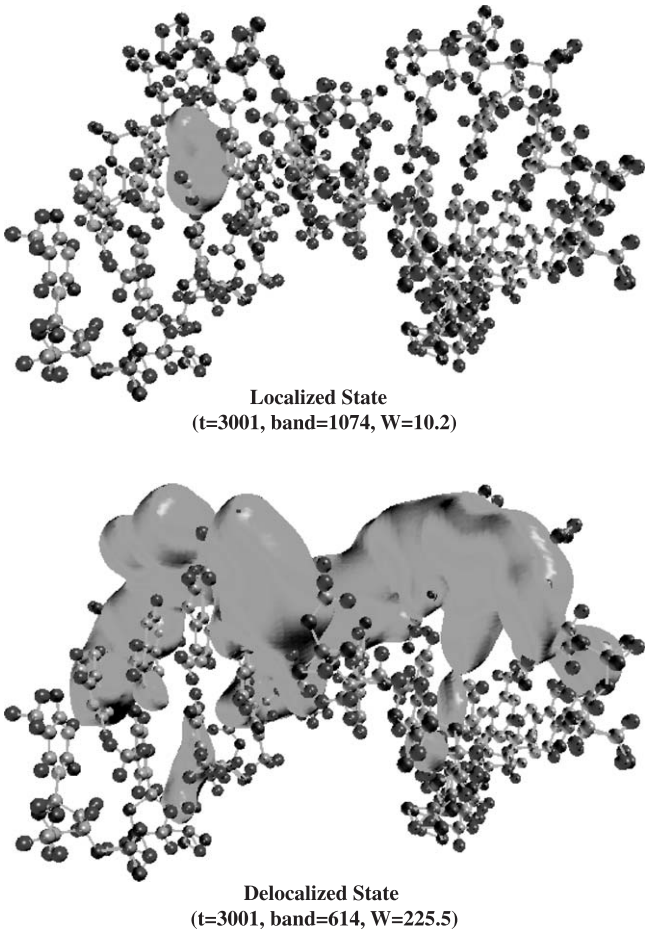


Fig. 21.5. Example of a localized and a delocalized state for two different states in poly(dA)–poly(dT) at time step 3 001. For reference, the HOMO is band 1 094 (see Color Plate 31).

to the pertinent conformational changes would 'arrest' the HOMO at one particular base. Whereas, conformationally induced increases in the above coupling ought to promote the 'HOMO trading' we revealed here. Our results are also in accordance with the analogous approach put forth most recently in Ref. [46] and in Ref. [88]. To be capable of formulating reasonable suggestions for experimentalists, we would need more detailed calculations not only on poly(dG)–poly(dC), but also on DNA with mixed-base sequences.

21.3 EFFECTIVE TIGHT-BINDING HAMILTONIANS FOR LONG DNA STRANDS AND COMPLETE SEQUENCES

In this section, we focus on whether DNA, when treated as a quantum wire in the fully coherent low-temperature regime, is conducting or not. To this end, we study and

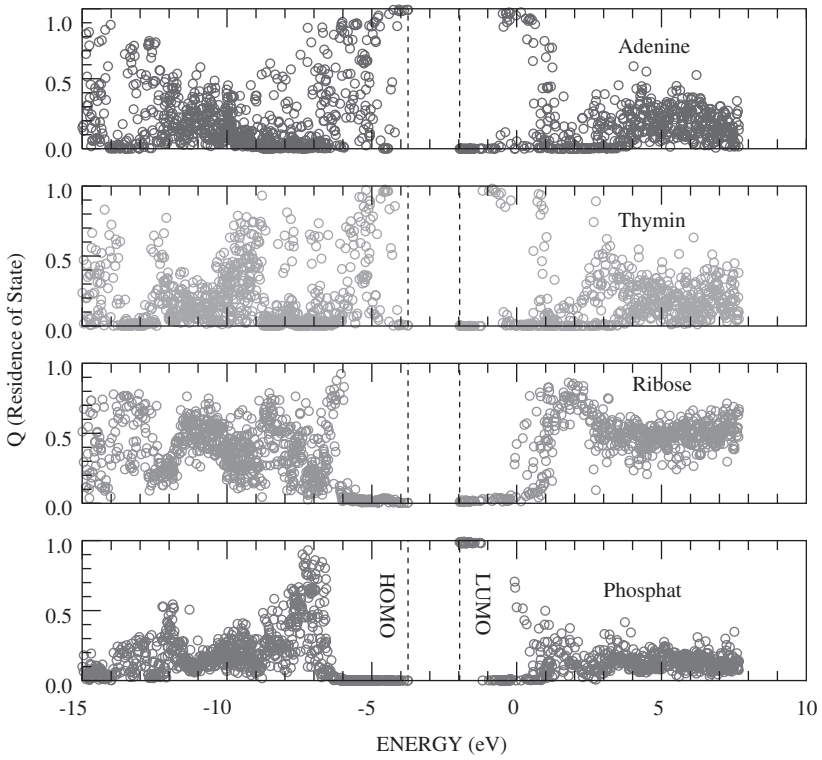


Fig. 21.6. Residence of state gives the location of the wavefunction for each energy state. States very near the HOMO level are located primarily on the adenine bases. For any given state, the sum of the four residences adds to unity.

generalize a tight-binding model of DNA which has been shown to reproduce experimental [25] as well as *ab initio* results [89]. A main feature of the model is the presence of sites which represent the sugar–phosphate backbone of DNA but along which no electron transport is permissible. We emphasize that the model is constructed to take into account the HOMO–LUMO gap observed in the DFT-based studies in Section 2 as well as the observed absence of transport along the backbone. We measure the effectiveness of the electronic transport by the *localization length* ξ, which roughly speaking parametrizes whether an electron is confined to a certain region ξ of the DNA (insulating behaviour) or can proceed across the full length L ($\leqslant \xi$) of the DNA molecule (metallic behaviour).

21.3.1 The ladder model

A convenient tight-binding model for DNA can be constructed as follows: it has two central conduction channels in which individual sites represent an individual base; these are interconnected and further linked to upper and lower sites, representing the backbone,

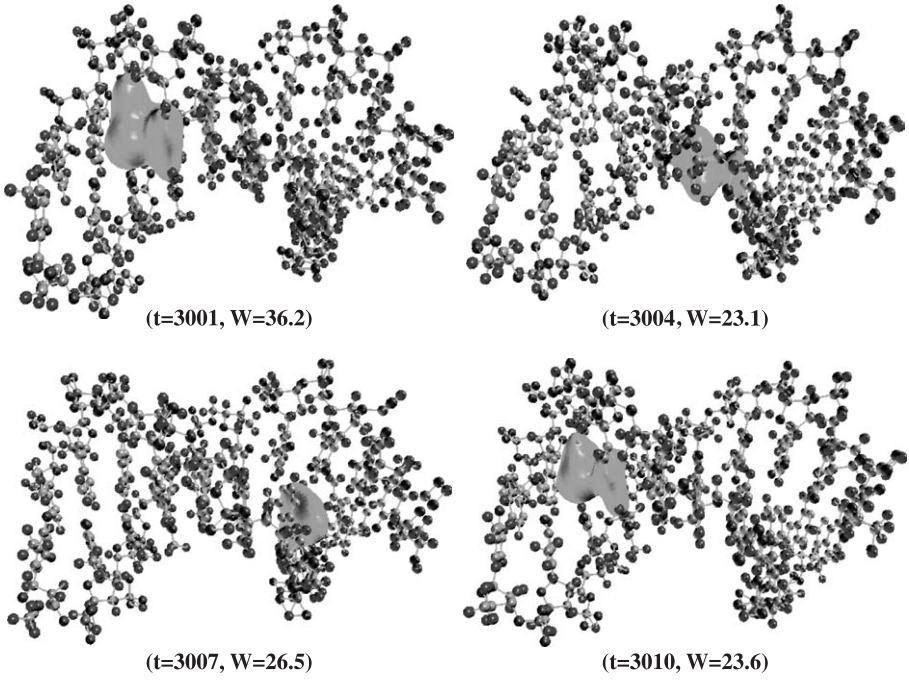


Fig. 21.7. Population density plots for the localised HOMO state as a function of time. The time between snapshots is 1.5 ps (see Color Plate 32).

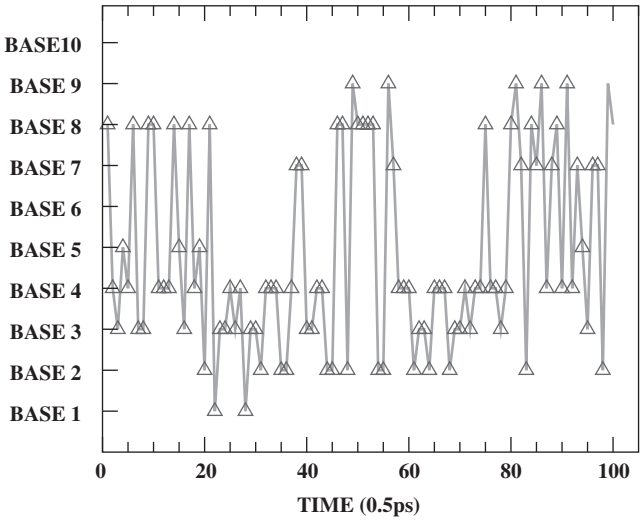


Fig. 21.8. Location of the HOMO as function of time. The 10 bases are the 10 adenine bases on one strand of the DNA. The HOMO is located only on adenine bases.

but are *not* interconnected along the backbone. Every link between sites implies the presence of a hopping amplitude. The Hamiltonian H_{L} for this ladder-like model is given by

$$\begin{aligned} H_L = & \sum_{i=1}^{L} \sum_{\tau=1,2} (t_{i,\tau} |i,\tau\rangle\langle i+1,\tau| + \varepsilon_{i,\tau} |i,\tau\rangle\langle i,\tau|) \\ & + \sum_{q=\uparrow,\downarrow} (t_i^q |i,\tau\rangle\langle i,q(\tau)| + \varepsilon_i^q |i,q\rangle\langle i,q|) \\ & + \sum_{i=1}^{L} t_{1,2} |i,1\rangle\langle i,2| \end{aligned} \tag{2}$$

where $t_{i,\tau}$ is the hopping amplitude between sites along each branch $\tau = 1$, 2 and $\varepsilon_{i,\tau}$ is the corresponding on-site potential energy. t_i^q and ε_i^q give hopping amplitudes and on-site energies at the backbone sites. Also, $q(\tau) = \uparrow,\downarrow$ for $\tau = 1$, 2, respectively. The parameter t_{12} represents the hopping between the two central branches, i.e., perpendicular to the direction of conduction. Quantum chemical calculations with semi-empirical wavefunction bases using the SPARTAN package [90] results suggest that this value, dominated by the wavefunction overlap across the hydrogen bonds, is weak and so we choose $t_{12} = 1/10$[1]. As we restrict our attention here to pure DNA, we also set $\varepsilon_{i,\tau}$ for all i and τ. Note that in this way, the energy gap has been made to be symmetric about $E = 0$. Hence when comparing with the results in Section 21.2, a constant shift according to the neglected ionization potentials has to be added.

The model (2) clearly represents a dramatic simplification of DNA. Nevertheless, in Ref. [25] it has been shown that an even simpler model – in which base pairs are combined into a single site – when applied to an artificial sequence of repeated GC base pairs, poly(dG)–poly(dC) DNA, reproduces experimental data current–voltage measurements when $t_i = 0.37$ eV and $t_i^q = 0.74$ eV are being used. This motivates the above parametrization of $t_i^q = 2t_i$ and $t_{i,\tau} \equiv 1$ for hopping between like (GC/GC, AT/AT) pairs. Assuming that the wavefunction overlap between consecutive bases along the DNA strand is weaker between unlike and non-matching bases (AT/GC, TA/GC, etc.) we thus choose 1/2. Furthermore, since the energetic differences in the adiabatic electron affinities of the bases are small [91], we choose $\varepsilon_i = 0$ for all i. Owing to the non-connectedness of the backbone sites along the DNA strands, the model (2) can be further simplified to yield a model in which the backbone sites are incorporated into the electronic structure of the DNA. The effective ladder model reads as

$$\begin{aligned} \tilde{H}_L = & \sum_{i=1}^{L} t_{1,2} |i,1\rangle\langle i,2| + \sum_{\tau=1,2}^{L} t_{i,\tau} |i,\tau\rangle\langle i+1,\tau| \\ & + \left[\varepsilon_{i,\tau} - \frac{\left(t_i^{q(\tau)}\right)^2}{\varepsilon_i^{q(\tau)} - E} \right] |i,\tau\rangle\langle i,\tau| + h.c. \end{aligned} \tag{3}$$

[1] Simulations with larger $t_{12} \approx 1/2$ give qualitatively similar results.

Thus the backbone has been incorporated into an *energy-dependent* on-site potential on the main DNA sites. This effect is at the heart of the enhancement of localization lengths due to increasing binary backbone disorder reported previously [42].

21.3.2 The numerical approach to localization in a Hamiltonian tight-binding model

There are several approaches suitable for studying the transport properties of the model (2), and these can be found in the literature on transport in solid-state devices, or, perhaps more appropriately, quantum wires. Since the variation in the sequence of base pairs precludes a general solution, we will use two methods well known from the theory of disordered systems [78].

The first method is the iterative transfer-matrix method (TMM) [92–96], which allows us in principle to determine the localization length ξ of electronic states in systems with cross-sections $M = 1$ [25] and 2 (ladder) and length $L >> M$, where typically a few million sites are needed for L to achieve reasonable accuracy for ξ. However, in the present situation we are interested in finding ξ also for viral DNA strands of typically only a few ten thousand base-pair long sequences. Thus in order to restore the required precision, we have modified the conventional TMM and now perform the TMM on a system of fixed length L_0. This modification has been previously used [97–99] and may be summarized as follows: after the usual forward calculation with a global transfer matrix T_{L_0}, we add a backward calculation with transfer matrix $T^{\mathrm{b}}_{L_0}$. This forward–backward-multiplication procedure is repeated K times. The effective total number of TMM multiplications is $L = 2KL_0$ and the global transfer-matrix is $\tau_L = \left(T^{\mathrm{b}}_{L_0}\, T_{L_0}\right)^K$. It can be diagonalized as for the standard TMM with $K \to \infty$ to give ${\tau_L}^{\dagger}\tau_L \to \exp[+/-\mathrm{diag}(4KL_0/\xi_\tau)]$with $\tau = 1$ or $\tau = 1{,}2$ for fishbone and ladder model, respectively. The largest ξ_τ for all τ then corresponds to the localisation lengths of the electron on the DNA strand and will be measured in units of the DNA base-pair spacing (0.34 nm).

The second method that we will use is the recursive Green function approach pioneered by MacKinnon [100,101]. It can be used to calculate the DC and AC conductivity tensors and the density of states (DOS) of a d-dimensional disordered system and has been adopted to calculate all kinetic linear-transport coefficients such as thermoelectric power, thermal conductivity, Peltier coefficient and Lorentz number [102].

The main advantage of both methods is that they work reliably (i) for short DNA strands ranging from 13 (DFT studies [103] base pairs up to 30 base-pairs length, which are being used in nanoscopic transport measurements [89] as well as (ii) for somewhat longer DNA sequences as modelled in the electron transfer results, and (iii) even for complete DNA sequences which contain, e.g., for human chromosomes, up to 245 million base pairs [104].

21.3.3 Long DNA sequences: λ-DNA, centromers and (super-)promoters

We shall use two naturally occurring long DNA sequences ('strings'). (i) λ-DNA [105] is DNA from the bacteriophage virus. It has a sequence of 48 502 base pairs and is

biologically very well characterized. Its ratio α of like to unlike base-pairs is $\alpha_\lambda = 0.949$. (ii) Centromeric DNA for chromosome 2 of yeast has 813 138 base pairs [106] and $\alpha_{\text{centro}} = 0.9555$. This DNA is also rich in AT bases and has a high rate of repetitions, which should be favourable for electronic transport.

Another class of naturally existing DNA strands is provided by so-called promoter sequences. We use a collection of 4 986 of these which have been assembled from the TRANSFAC database and cover a range of organisms such as mouse, human, fly and various viruses. Promoter sequences are biologically very interesting because they represent those places along a DNA string where polymerase enzymes bind and start the copying process that eventually leads to synthesis of proteins. On average, these promoters consist of approximately 17 base pairs, much too short for a valid localization length analysis by TMM. Therefore, we concatenate them into an 86 827 base-pair long *super-promoter* with $\alpha_{\text{super-p}} = 0.921$. In order to obtain representative results, 100 such super-promoters have been constructed, representing different random arrangements of the promoters, and the results presented later will be averages. As usual, averages of ξ are computed by averaging the normally distributed $1/\xi$ values.

Occasionally, we show results for 'scrambled' DNA. This is DNA with the same number of A, T, C, G bases, but with their order randomized. Clearly, such sequences contain the same set of electronic potentials and hopping variations, but would perform quite differently in a biological context. A comparison of their transport properties with those from the original sequence thus allows us to measure how important the exact fidelity of a sequence is. On average, we find for these sequences $\alpha_{\lambda/s} = 0.899$, $\alpha_{\text{centro/S}} = 0.9951$, and $\alpha_{\text{super-}p/S} = 0.901$.

A convenient choice of artificial DNA strand is a simple, 100 000 base-pair long *random* sequence of the four bases, random-ATGC DNA, which we construct with equal probability for all four bases ($\alpha_{\text{random}} = 0.901$). We shall also 'promote' these random DNA strings by inserting all 4 086 promoter sequences at random positions in the random-ATGC DNA ($\alpha_{\text{random}} = 0.910$).

21.3.4 Results for localization lengths

We have computed the energy dependence of the localization lengths for all sequences of Section 21.3.3. In addition, λ-DNA, centromeric DNA and the super-promoter DNA were also scrambled 100 times and the localization length of each resulting sequence measured and the appropriate average constructed. Also, we constructed 100 promoted random-ATGC DNA sequences.

As shown previously [42], the energy dependence of ξ reflects the backbone-induced two-band structure. The obtained $\xi(E)$ values for the lower band are shown in Fig. 21.9. In the absence of any on-site disorder, we find two prominent peaks separated by $t_{1,2}$ and $\xi(E) = \xi(-E)$. We also see that λ-DNA has roughly the same $\xi(E)$ dependence as random-ATGC-DNA. The super-promoter has larger ξ values compared with random-ATCG- and λ-DNA. Most surprisingly, centromeric DNA – the longest investigated DNA sequence – has a much larger localization length than all other DNA sequences. The order of like-to-unlike pair-ratios is $\alpha_{\text{centro}} = 0.955 > \alpha_\lambda = 0.949 > \alpha_{\text{super}-P} = 0.921 > \alpha_{\text{random}} = 0.901$ and one might expect that transport is favoured in sequences with large

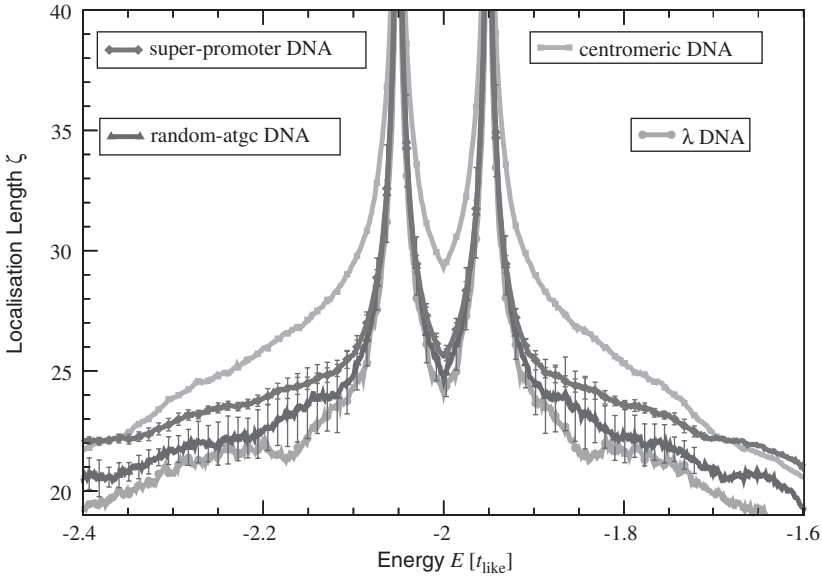


Fig. 21.9. Localization lengths ξ versus Fermi energy E for various clean DNA strands. Only every 10th symbol is shown. Error bars reflect the standard deviation after sampling the different sequences for random-ATGC and promoted DNA. The energy is measured in unit of hopping energy between like base pairs, i.e., $t_{\text{like}} = t_i = 0.37$ eV.

α. From Fig. 21.9, it is clear that this is not the case, λ -DNA has the smallest localization lengths, but the second largest α.

In Fig. 21.10, we add results for scrambled and promoted DNA. We find that promoting a given DNA sequence leads to small increases in localization length ξ for random DNA, whereas scrambling can lead to increase (centrometric and λ -DNA) as well as decrease (super-promoter). These results suggest that the promoters have a tendency towards larger localization lengths and thus enhanced transport.

21.3.5 Promoter sequences and *E. coli* binding sites

Let us now turn our attention to the transport properties of individual promoters rather than the artificially constructed super-promoters. Since their average length is 17 base pairs and thus comparable to the localization lengths measured in the longer sequences, we can no longer use the TMM, but need to employ the RGFM mentioned in Section 21.3.2. While this method is capable of computing all thermo-electric transport coefficients, we shall restrict ourselves here to presenting results for the conductance.

In Fig. 21.11, we show results for averaged conductance in the upper band; both arithmetic and typical conductance have been calculated. We first note that the double-peak structure of Fig. 21.9 has vanished and only a single peak remains. This is because our results have been computed with perfectly conducting leads attached to both ends of the DNA strands. This is close to the experimental situation, but the purely off-diagonal disorder in the DNA model is now masked by the ordered leads.

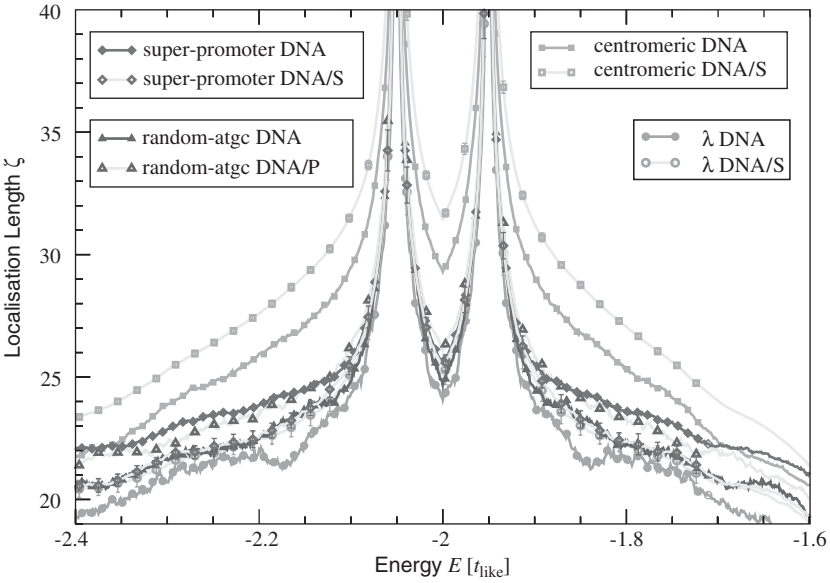


Fig. 21.10. Localization lengths ξ versus Fermi energy E for various clean DNA (solid symbols as in Fig. 21.9, error bars not shown for clarity), scrambled DNA (DNA/S, (open ◊, ⊏, ○) and promoted DNA (DNA/R, open Δ) strands. Only every 10th (20th) symbol is shown for clean (scrambled/promoted) DNA. Error bars reflect the standard deviation after sampling the different sequences for random-ATGC, scrambled and promoted DNA.

Next, we observe that the promoters and their scrambled copies have larger conductances than random and λ-promoters. λ-promoters have been constructed by cutting sequences with the same lengths as the true promoters out of λ-DNA at randomly selected positions along the DNA. Since α for random and λ-DNA is different, this allows us to check whether it is the order of base pairs or the value of which dominates the value of G. Since $\alpha_{\text{promoter}} = 0.928 < \alpha_\lambda = 0.955$, but $G_{\text{promoter}} > G_\lambda$, it appears that as before the transport properties are not simply large if α is large. This suggests that it is indeed the fidelity of the sequence which is also important.

Typical and arithmetic averages share similar characteristics when comparing different sequences as shown in Fig. 21.11. However, the typical values are systematically smaller than their arithmetic counterparts. We therefore expect the distributions to be highly non-Gaussian and in Fig. 21.12 we see that this is indeed the case.

We first note that both the original promoter as well as the scrambled version (/*S*) appear to have a slightly larger weight at $G > 0.05$ whereas both random and λ -DNA are peaked at $G \approx 0.025$. In addition, we find that there is a peak in the conductance distribution *P(G)* at $G \approx 0.26$. This peak is most pronounced for the original promoter and their scrambled cousins, but much smaller for the artificial random and λ-promoter.

In Figs 21.11 and 21.12, we have also included results for computationally inferred 802 *E. coli* binding sites [107]. Sequence-specific DNA-binding proteins perform a variety of roles in the cell, including transcriptional regulation. Our results show that the total conductance of these sequences is smaller than for promoters.

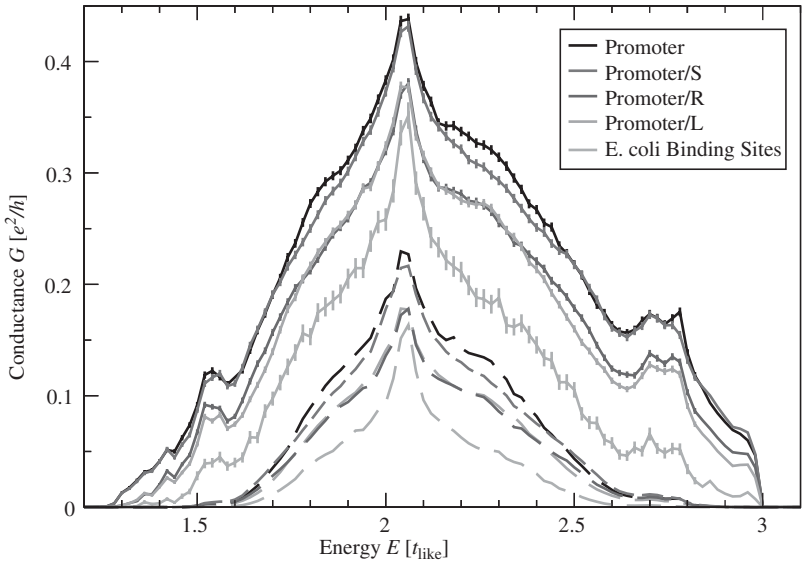


Fig. 21.11. Energy dependence of the conductance G for promoters, scrambled promoters (/S), random promoters (/R) and λ-promoters (/L). Solid lines denote the arithmetic, dashed lines the typical average of G. The error bars denote standard deviation obtained from the 4 896 different promoters considered in each category (original and /S, /R, /L). These are not repeated for the typical averages for clarity.

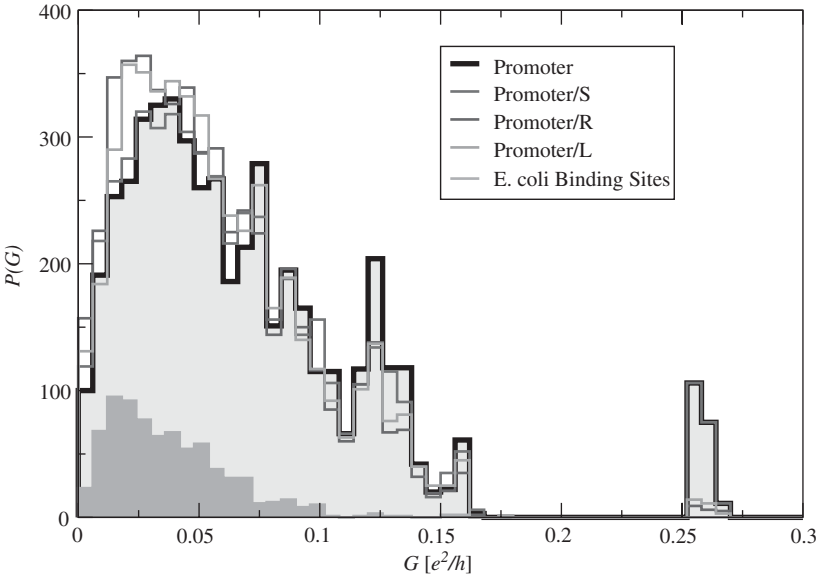


Fig. 21.12. Distribution function $P(G)$ for the conductances averaged over the energy range $[-5, 5]$. Only promoter and *E. coli* results have been shaded.

However, their average length is approximately 25 so that the average *conductivity* is in fact larger when compared with promoters. This might be important in a biological context where one could envisage proteins to identify their binding site differences on local conductivities.

21.4 SUMMARY

The results presented in this chapter are preliminary results but indicate a marked difference in the nature of the electronic HOMO–LUMO states for the periodic and aperiodic structures of duplex DNA. These results indicate that the HOMO–LUMO states for the periodic structure are quite extended, as would be expected for Bloch-like states while the HOMO–LUMO states for the aperiodic structure demonstrate more localization. The concept of static localization in short DNA has previously been considered by Ladik [58,59], and our results show that such a localization in our structure for aperiodic poly(dA)–poly(dT) DNA reaches far deeper in energy than just the band-tail states. The localization phenomenon observed in the DNA double helix is the so-called Anderson localization, which attributes to the *off-diagonal* disorder. This disorder results from dynamical variations in DNA intramolecular interactions and coupling of DNA with its environment. Turning our attention to longer DNA sequences, we next used this insight by modelling DNA as an off-diagonally disordered Anderson chain. However, in addition and contradistinction to previous studies using Anderson-type models, we included the sugar–phosphate backbone explicitly and by doing so retained the essential semiconducting structure as observed in some experiments. Our results for the localization lengths suggest extended states even in non-periodic DNA up to 20 base-pair distances. This is roughly consistent with the previous results. Next, we studied how transport properties differ between sequences and found that promoter sequences seem to have a tendency towards larger localization length, i.e., enhanced transport. This might point towards the importance of an electronic mechanism in the initial stages of DNA polymerase. Our results warrant further investigation, as the role of the dynamical localization and the sequence dependence may very well suggest an important mechanism of charge transport along the DNA molecule.

21.5 ACKNOWLEDGEMENTS

RAR gratefully acknowledges discussions with A. Rodriguez and M.S. Turner. JPL acknowledges discussions with D. Drabold, O. F. Sankey and T. Cheatham who prepared MD simulations. While the chapter has been prepared jointly, calculations in Section 21.2 were done by J. P. Lewis, H. Wang and R. Marsh, whereas Section 21.3 is based on results of R. A. Römer. The work was partially supported by the Royal Society.

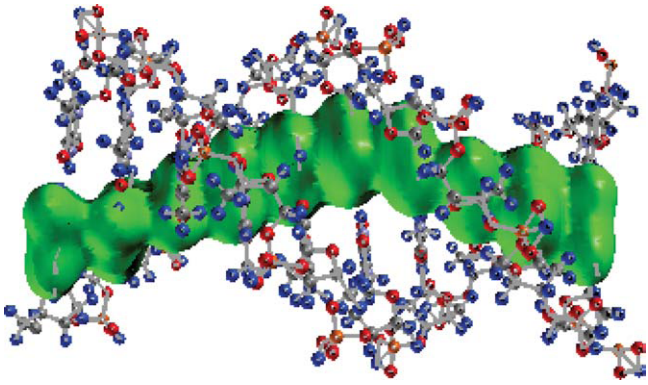
21.6 REFERENCES

1 G.B. Schuster (Ed.), Long-range charge transfer in DNA I, Topics in current chemistry, Vol. 236, Springer-Verlag, Heidelberg, 2004.
2 G.B. Schuster (Ed.), Long-range charge transfer in DNA II, topics in current chemistry, Vol. 237, Springer-Verlag, Heidelberg, 2004.

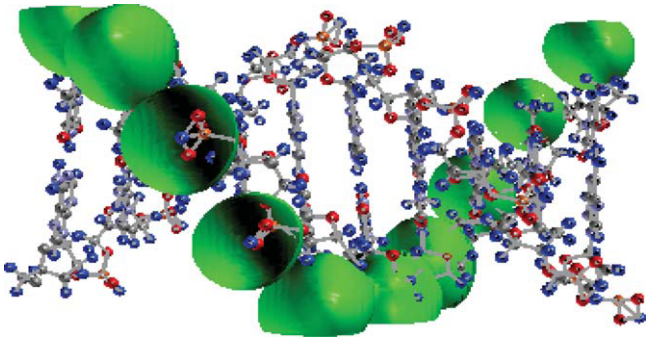
3 RG. Endres, D.L. Cox and R.R.P. Singh, Rev. Mod. Phys., 76 (2004) 195.
4 D. Porath, G. Cuniberti and R. Di Felice, Topics in Current Chemistry, 237 (2004) 183.
5 J.D. Watson and F.H.C. Crick, Nature, 171 (1953) 737.
6 D.D. Eley and D.I. Spivey, Trans. Faraday Soc., 58 (1962) 411.
7 S.O. Kelley and J.K. Barton, Science, 283 (1999) 375.
8 B. Giese, S. Wessely, M. Spormann, U. Lindemann, E. Meggers and M.E. Michel-Beyerle, Angew. Chem. Int. Ed., 38 (1999) 996.
9 P.T. Henderson, D. Jones, G. Hampkian, Y. Kan and G.B. Schuster, Proc. Natl. Acad. Sci. USA, 96 (1999) 8353.
10 C. Wan, T. Fiebig, S.O. Kelley, C.R. Treadway, J.K. Barton and A.H. Zewail, Proc. Natl. Acad. Sci. USA, 96 (1999) 6014.
11 E. Meggers, M. E. Michel-Beyerle and B. Giese, J. Am. Chem. Soc., 120 (1998) 12950.
12 E. Meggers, D. Kusch, M. Spichty, U. Wille and B. Giese, Angew. Chem. Int. Ed., 37 (1998) 460.
13 K. Fukui and K. Tanaka, Angew. Chem. Int. Ed., 37 (1998) 158.
14 F.D. Lewis, T. Wu, Y. Zhang, R.L. Letsinger, S.R. Greenfield and M.R. Wasielewski, Science, 277 (1997) 673.
15 A.M. Brun and A. Harriman, J. Am. Chem. Soc., 116 10383 (1994); *ibid*, 114 (1992) 3656.
16 Y.A. Berlin, A.L. Burin and M.A. Ratner, J. Am. Chem. Soc., 123 (2001) 260 and references therein.
17 R.N. Barnett, C.L. Cleveland, A. Joy, A.U. Landman and G.B. Schuster, Science, 294 (2001) 567.
18 J. Jortner, M. Bixon, T. Langenbacher, and M. Michel-Beyerle, Proc. Natl. Acad. Sci. USA, 95 (1998) 12759.
19 D.N. Beratan, S. Priyadarshy and S. Risser, Chem. Biol., 4 (1997) 3.
20 S. Priyadarshy, S.M. Risser and D.N. Beratan, J. Phys. Chem., 100 (1996) 17678.
21 Ch. Adessi, S. Walch and M.P. Anantram, Phys. Rev. B, 67 (2003) 081405.
22 F.C. Grozema, L.D.A. Siebbeles, Yu. A. Berlin and M.A. Ratner, Chem. Phys. Chem., 6 (2002) 536.
23 P. de Pablo, F. Moreno-Herrero, J. Colchero, J. GomezHerrero, P. Herrero, A.M. Baro, P. Ordejon, J.M. Soler, and E. Artacho, Phys. Rev. Lett., 85 (2000) 4992.
24 P. Maragakis, R.L. Barnett, E. Kaxiras, M. Elstner and T. Frauenheim, Phys. Rev., B66 (2002) 241104.
25 G. Cuniberti, L. Craco, D. Porath and C. Dekker, Phys. Rev. B, 65 (2002) 241314.
26 J. Reynisson and S. Steenken, Phys. Chem. Chem. Phys., 4 (2002) 527.
27 H. Sugiyama and I. Saito, J. Am. Chem. Soc., 118 (1996) 7063.
28 Y. Yoshioka *et al*., J. Am. Chem. Soc., 121 (1999) 8712.
29 H. Wang, J.P. Lewis and O.F. Sankey, Band-gap tunneling states in DNA, Phys. Rev. Lett. 93, (2004) 016401.
30 J.P. Lewis, T.E. Cheatham, E. . Starikov, H. Wang and O.F. Sankey, J. Phys. Chem., B, 107 (2003) 2581.
31 S. Roche, Phys. Rev. Lett., 91 (2003) 108101.
32 S. Roche, D. Bicout, E. Macia′ and E. Kats, Phys. Rev. Lett., 91 (2003) 228101 .
33 K. Iguchi, Int. J. Mod. Phys. B, 17 2565 (2003).
34 S. Roche and E. Macia′, Mod. Phys. Lett. B, 18 (2004) 847.
35 H. Yamada, E.B. Starikov, D. Hennig and J.F.R. Archilla, Eur. Phys. J. E, 17 (2005) 149.
36 H. Yamada, Phys. Lett. A, 332 (2004) 65.
37 H. Yamada, Int. J. Mod. Phys. B, 18 (2004) 1697.
38 K. Iguchi, Int. J. Mod. Phys. B, 18 (2004) 1845.
39 C.-T. Shih, Phys. Stat. Sol. (b), 243 (2006) 378.
40 C.-T. Shih, Phys. Rev. E, (2005) under review.
41 D.K. Klotsa, R.A.R. mer and M.S. Turner, In Proceedings 27th International Conference on the Physics of Semiconductors(Q5 129), Flagstaff, Arizona (2004) 328.
42 D.K. Klotsa, R.A. Römer and M.S. Turner, Biophys. J., 89 (2005) 2187.
43 R. Bruinsma, G. Gruer, M.R. D'Orsogna and J. Rudnik, Phys. Rev. Lett., 85 (2000) 4393.
44 G. Brunaud, F. Castet, A. Fritsch and L. Ducasse, Phys. Chem. Chem. Phys., 5 (2003) 2104.
45 I.V. Kurnikov, G.S. M. Tong, M. Madrid and D. Beratan, J. Phys. Chem. B, 106 (2002) 7.
46 A.A. Voityuk, K. Siriwong and N. Rosch, Phys. Chem. Chem. Phys., 3 (2001) 5421.
47 E.B. Starikov, Phil. Mag., 85 3435 (2005).
48 D. Henning, E.B. Starikov, J.F.R. Archilla and F. Palmero, J. Bio. Phys., 30 (2004) 227.
49 E.B. Starikov, J. Photochem. Photobio. C, 3 (2002) 147.

50 E.B. Starikov, Phys. Chem. Chem. Phys., 4 (2002) 4523.
51 E.B. Starikov, Mod. Phys. Lett. B, 18 (2004) 825.
52 T. Tanabe, K. Noda, M. Saito, E.B. Starikov and M. Tateno, Phys. Rev. Lett., 93 (2004) 043201.
53 J. Cuevas, E.B. Starikov, J.F.R. Archilla and D. Henning, Mod. Phys. Lett. B, 18 (2004) 1319.
54 P.W. Anderson, Phys. Rev., 109 (1958) 1492.
55 J. Dong and D.A. Drabold, Phys. Rev. Lett., 80 (1998) 1928.
56 P. Thomas and H. Overhof, Properties and application of amorphous materials, Kluwer Academic Publisher, Dordrecht (2001) 261.
57 W. Gotze, Philos. Mag., B, 43 (1981) 219.
58 J. Ladik, M. Seel, P. Otto and A.K. Bakhshi, Chem. Phys., 108 (1986) 203.
59 Y.-J. Ye, R.S. Chen, J. Shun and J. Ladik, Solid State Commun., 119 (2001) 175.
60 S. Arnott, and D.W. Hukins, Biochem. Biophys. Res. Comm., 47 (1972) 1504.
61 D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham, S. Debolt, D. Ferguson, G. Seibel, and P. Kollman, Comput. Phys. Comm., 91 (1995) 1.
62 B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M.J. Karplus, Computat. Chem., 4 (1983) 187.
63 J. Aqvist, J. Phys. Chem., 94 (1990) 8021.
64 W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmayer, T. Fox, J.W. Caldwell and P.A. Kollman, J. Am. Chem. Soc., 117 (1995) 5179.
65 D.R.J. Langley, Biomol. Struct. Dyn., 16 (1998) 487.
66 W.G. Hoover, Phys. Rev., A, 31 (1985) 1695.
67 S.E. Feller, Y. Zhang, W. Pastor, and B.R. Brooks, J. Chem. Phys., 103 (1995) 4613.
68 J.P. Ryckaert, G. Ciccotti and H.J.C. Berendsen, Comput. Phys., 23 (1977) 327.
69 U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee and L.G. Pedersen, J. Chem. Phys., 103 (1995) 8577.
70 D.L. Beveridge and K.J. McConnell, Curr. Opin. Struct. Biol., 10 (2000) 182.
71 T.E. Cheatham and P.A. Kollman, Annu. Rev. Phys. Chem., 51 (2000) 435.
72 T.E. Cheatham and M.A. Young, Biopolymers, 56 (2001) 232.
73 J.P. Lewis, K.R. Glaesemann, G.A. Voth, J. Fritsch, A.A. Demkov, J. Ortega and O.F. Sankey, Phys. Rev. B, 64 (2001) 195103.
74 J. Harris, Phys. Rev. B, 31 (1985) 1770.
75 W. Foulkes and R. Haydock, Phys. Rev. B, 39 (1989)12520.
76 A.A. Demkov, J. Ortega, O.F. Sankey and M.P. Grumbach, Phys. Rev. B, 52 (1995) 1618.
77 O.F. Sankey and D.J. Niklewski, Phys. Rev. B, 40 (1989) 3979.
78 R.A.R mer and M. Schreiber, Numerical investigations of scaling at the Anderson transition, in T. Brandes and S. Kettemann (Eds.), The Anderson Transition and its Ramifications – Localisation, Quantum Interference, and Interactions, Springer, Berlin, 2003.
79 P. Biswas, P. Cain, R.A. Römer and M. Schreiber, Phys. Stat. Sol. (b), 218, (2000) 205, ArXiv: cond-mat/ 0001315.
80 U. Dornberger, M. Leijon and H. Fritzche, J. Biol. Chem., 274 (1999) 6957.
81 W.K. Olson, A.A. Gorin, X.J. Lu and L.M. Hock, Proc. Natl. Acad. Sci. USA, 95 (1998) 11163.
82 H. Hellmann, Einfuhrung in die quantumchemie, Franz Duetsche, Leipzig 1937.
83 R.P. Feynman, Phys. Rev., 56 (1939) 340.
84 R.A.R mer and P. Ziesche, J. Phys. A: Math. Gen., 34 (2001) 1485.
85 M. Preuss, W.G. Schmidt, K. Seino, J. Furthmuller and F. Bechstedt, J. Comput. Chem., 25 (2003) 112.
86 S.D. Wetmore, R.J. Boyd and L.A. Eriksson, Chem. Phys. Lett., 322 (2000) 129.
87 V.M. Orlov, A.N. Smirnow and Y. Varshavsky, Tetrahedron, 48 (1976) 4377.
88 A. Troisiand and G. Orlandi, J. Phys. Chem. B, 106 (2002) 2093.
89 O.R. Davies and J.E. Inglesfield, Phys. Rev. B, 69 (2004) 195110.
90 SPARTAN version 5.0, user's guide, Wavefunction Inc. Irvine, CA 92612.
91 S.S. Wesolowski, M.L. Leininger, P.N. Pentchev and H.F. Schaefer III, J. Am. Chem. Soc., 123 (2001) 4023.
92 J.-L. Pichard and G. Sarma, J. Phys. C, 14 (1981) L127.
93 J.-L. Pichard and G. Sarma, J. Phys. C, 14 (1981) L617.

94 A. MacKinnon and B. Kramer, Z. Phys. B, 53 (1983) 1.
95 B. Kramer and A. MacKinnon, Rep. Prog. Phys., 56 (1993) 1469.
96 A. MacKinnon, J. Phys. Condens. Matter, 6 (1994) 2511.
97 K. Frahm, A.M ller-Groeling, J.L. Pichard and D. Weinmann, Europhys. Lett., 31 (1995) 169.
98 R.A. Römer and M. Schreiber, Phys. Rev. Lett., 78 (1997) 4890.
99 M.L. Ndawana, R.A. Römer, and M. Schreiber, Europhys. Lett., 68 (2004) 678.
100 A. MacKinnon, J. Phys.: Condens. Matter, 13 (1980) L1031.
101 A. MacKinnon, Z. Phys. B, 59 (1985) 385.
102 R.A. Römer, C. Villagonzalo and A. MacKinnon, J. Phys. Soc. Japan, 72 (suppl. A) 167 (2002).
103 P.J. Pablo, F. Moreno-Herrero, J. Colchero, J. Gomez Herrero, P. Hererro, P. Baro, A.M. an Ordejon, J. M. Soler and E. Artacho, Phys. Rev. Lett., 85 (2000) 4992.
104 B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts and J. Watson, Molecular biology of the cell, Garland, New York, 1994.
105 Bacteriophage lambda, complete genome gi|9626243|ref|NC_001416.1|9626243, Genbank Accession number NC_001416, http://www.ncbi.nlm.nih.gov/entrez/.
106 CEN2, Chromosome II centromere, http://www.yeastgenome.org/.
107 K. Robison, A.M. McGuire and G.M. Church, J. Molec. Biol., 284 (1998) 241.

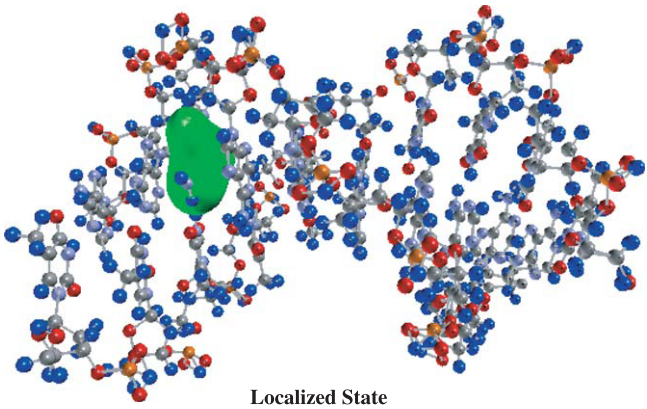


HOMO

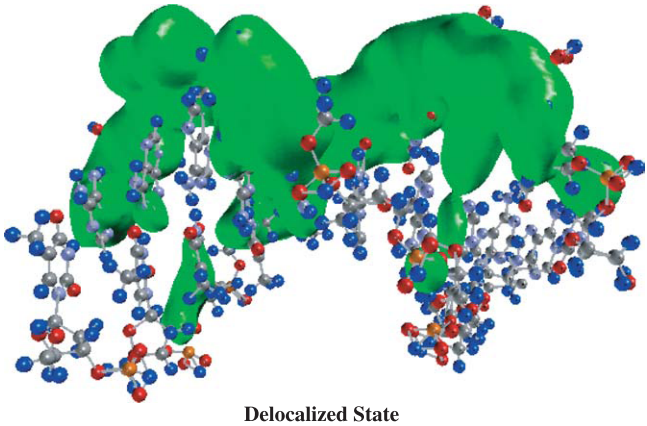


LUMO

Plate 30. Population densities for the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) are shown for periodic poly(dA)–poly(dT) DNA (10 base pairs). Both molecular orbitals exhibit very extended and periodic (Bloch-like) states throughout the molecule (illustration appears on page 412 of this volume).



Localized State
(t=3001, band=1074, W=10.2)



Delocalized State
(t=3001, band=614, W=225.5)

Plate 31. Example of a localized and a delocalized state for two different states in poly(dA)–poly(dT) at time step 3 001. For reference, the HOMO is band 1 094 (illustration appears on page 415 of this volume).

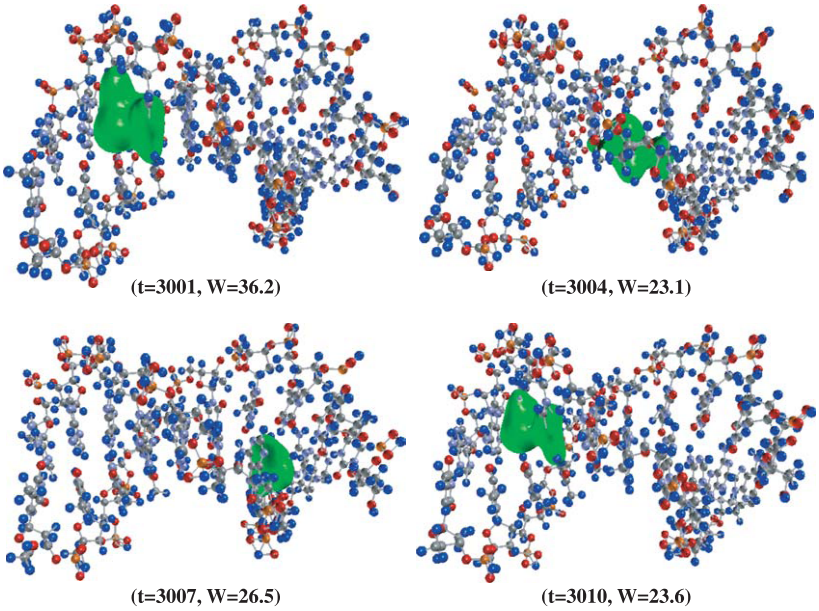


Plate 32. Population density plots for the localised HOMO state as a function of time. The time between snapshots is 1.5 ps (illustration appears on page 417 of this volume).

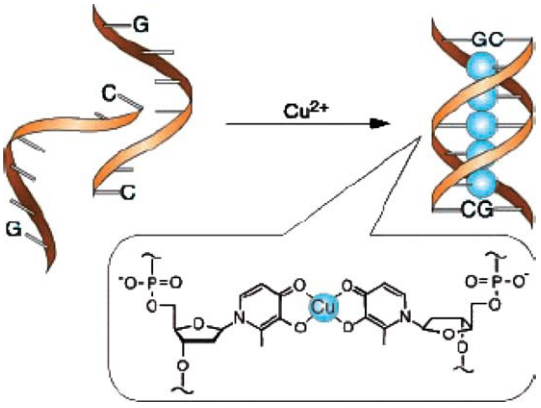


Plate 33. The figure shows the self-assembly process of artificial DNA strands in which some of the native nucleobases (except the terminal ones) can be substituted by hydroxypyridone bases: if Cu(II) ions are present in solution during synthesis, they can be incorporated along the axis and form molecules with a magnetic moment. The inset shows the artificial H–Cu(II)–H pair: the Cu(II) ion in each plane is tightly coordinated by four surrounding electronegative O atoms (illustration appears on page 488 of this volume). (From [17], with permission; copyright ©2003 Science.)

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 22

Polaronic charge transport mechanism in DNA

D. Hennig[1] and J. F. R. Archilla[2]

[1] *Freie Universität Berlin, Fachbereich Physik, Institut für Theoretische Physik, Arnimallee 14, 14195 Berlin, Germany*
[2] *Group of Nonlinear Physics, Departamento de Física Aplicada I, ETSI Informática, University of Sevilla, Avda Reina Mercedes, s/n. 41012, Sevilla, Spain*

Abstract

For the detailed understanding of the conduction mechanism in DNA we use models based on the concept of polaron and breather solutions. We describe how charge transport relies on the coupling of the charge carrying unit to the vibrational modes of DNA allowing for the formation of polaron-like localized states. The mobility of these localized states is discussed particularly in the presence of parametrical and structural disorder inherent to biomolecules. It is demonstrated that long-range coherent charge transport along the DNA structure is supported by mobile polaron and breather solutions, suggesting that DNA seems suitable for the design of functional nanostructures as ingredients in molecular nanoelectronic devices.

22.1 INTRODUCTION

Electronic transport through DNA is crucial for its biological functions, such as the repair mechanism after radiation damage and biosynthesis [1]. As proposed in [2] electron transport through DNA proceeds along a one-dimensional pathway constituted by the overlap between π–orbitals in neighbouring base pairs along a strand. Recent measurements have indeed suggested that DNA forms an effective one-dimensional molecular wire [3–13], which offers promising applications in molecular electronics based on biomaterials [11,12,14]. Opposed to that, it has been reported that DNA is insulating [15] and for short oligomers built up from base pairs of the same type semiconductivity was observed [16]. This qualitative discrepancy, shared also by theoretical findings, instigated the still ongoing debate whether DNA is conductive or not. In fact, a theoretical investigation based on first-principle calculations suggests the absence of DC-conductivity in DNA [17]. On the other

hand, there stand the studies using various assumptions for modelling DNA structure which focus on different aspects, such as the influence of aperiodicity, temperature-driven fluctuations and aggregation effects on the energetic control of charge migration which have all come to the conclusion that DNA is a conductor [18–23].

The detailed understanding of the conduction mechanism of DNA still remains an unsolved task. Different attempts to model the charge transport of DNA have been based on transport via coherent tunnelling [2], classical diffusion under the conditions of temperature-driven fluctuations [20], incoherent phonon-assisted hopping [24,25], variable range hopping between localized states [22], and charge carriages mediated by polarons [26–28] and solitons [29]. The impact of correlated disorder on the electronic conduction in DNA was studied recently in [30].

Strongly backed by recent experimental results [31], theoretical considerations have been directed towards the idea that the constructive interplay between the charge, carrying system and the vibrational degrees-of-freedom of DNA conspire to form mobile polarons which can establish coherent ET [20,21]. Inspired by these observations, we invoke for the theoretical description of charge transport in DNA the concept of polarons and breather solutions. The structure of the bent double helix of λ–DNA is modelled by a sterical network of oscillators in the frame of the base-pair picture [32–34] taking into account deformations of the hydrogen bonds within a base pair and twist motions between adjacent base pairs. The electron motion is described by a tight-binding system. It is assumed that the electron motion is predominantly influenced by vibrational modes of the double helix. The non-linear interaction between the electron and the vibrational modes causes the formation of polarons or electron–vibron breathers supporting charge localization and eventually transport.

We review the role played by static parametrical disorder and complex structural effects due to irregular deviations from the ordered helical shape of the double helix (reflected in random arrangements of the positions of the bases) with regard to the charge transport in DNA. Results concerning the charge transport in synthetic DNA polymers, consisting of a single type of base pair are then represented. Finally, we demonstrate that with the application of an external electric field, ET in DNA polymers can be controlled in the sense that the charge can be steered to any desired position on the chain.

22.2 MODEL HAMILTONIAN FOR POLARON-LIKE CHARGE TRANSFER ALONG DNA

DNA chains are quite flexible and exhibit structural fluctuations influencing the transport properties. In our DNA model we cover the basic features of the DNA double-helix structure (abstracting from the detailed chemical structure) and consider the latter as a bent double-stranded system for which in the frame of a base-pair picture the bases are treated as single non-deformable entities. The helicoidal structure of DNA is then conveniently described in a cylindrical reference system where each base pair possesses two degrees-of-freedom, namely, a radial variable measuring the transversal displacements of the base pair (that is, deformations of the H-bond), and the angle with a reference axis in a plane perpendicular to the helix backbone which defines the twist of the helix [34]. In Fig. 22.1 we represent a sketch of the structure of the DNA model.

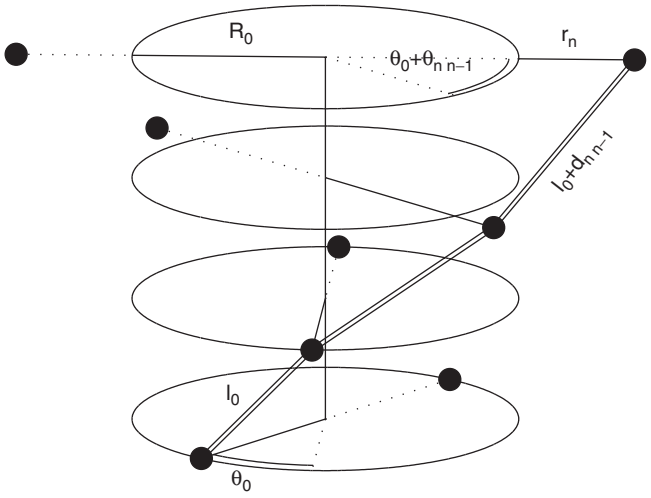


Fig. 22.1. Sketch of the helicoidal structure of the DNA model, the bases being represented by bullets. Geometrical parameters R_0, θ_0, l_0 and the radial and angular variables r_n and θ_{nn-1}, respectively, are indicated.

The charge migration process is influenced by the dominant vibrational modes of DNA stemming from transverse vibrations of the bases relative to each other, that is the stretchings/compressions of the base pair distance within a base pair plane and the torsional variations of the helicoidal twist [20,35]. The influence of other vibrational degrees-of-freedom (e.g., longitudinal acoustic phonons along the strands are significantly restrained by the stretch-proof sugar–phosphate chains in the backbone) can be discarded and the motion is restricted to the base-pair planes [19]. Furthermore, the angular twist and radial vibrational motions evolve independently on two different time scales and can be regarded as decoupled degrees-of-freedom in the harmonic treatment of the normal mode vibrations of DNA [36].

The Hamiltonian for the electron migration along a strand in DNA consists of three parts

$$H = H_{\text{el}} + H_{\text{rad}} + H_{\text{twist}} \tag{1}$$

with H_{el} representing the part corresponding to the particle charge transport over the base pairs, H_{rad} describes the dynamics of the H-bond vibrations, and H_{twist} models the dynamics of the relative twist angle between two consecutive base pairs. The electronic part is given by a tight-binding system

$$H_{\text{el}} = \sum_n E_n |c_n|^2 - V_{nn-1}(c_n^* c_{n-1} + c_n c_{n-1}^*) \tag{2}$$

The index n denotes the site of the nth base on one of the two strands and c_n determines the probability that the electron occupies this site. E_n is the local electronic energy and $V_{nn\text{-}1}$ stands for the transfer matrix element which is responsible for the transport of the electron along the stacked base pairs.

The vibronic part H_{rad} takes into account radial displacements of the base units from their equilibrium positions within the base-pair plane connected with deformations of the hydrogen bonds linking two bases. The corresponding part of the Hamiltonian treating the bond vibrations classically and harmonically is given by

$$H_{\text{rad}} = \frac{1}{2}\sum_n M_n[\dot{r}_n^2 + \Omega_r^2 r_n^2] \tag{3}$$

and the radial coordinate r_n measures the deviation from the equilibrium length of a hydrogen bond along the line bridging two bases of a base pair, Ω_r is the frequency and M_n is the reduced mass of a base pair.

The twist motion part reads as

$$H_{\text{twist}} = \frac{1}{2}\sum_n J_n[\dot{\theta}_{nn-1}^2 + \Omega_\theta^2 \theta_{nn-1}^2] \tag{4}$$

where θ_{nn-1}^2 is the relative angle between two adjacent base pairs quantifying displacements from its equilibrium twist angle θ_0, and J_n is the reduced moment of inertia.

We emphasize that in the context of charge transport in DNA the non-linear dynamics of large-amplitude structural transitions is not of interest because charge transport processes are thought to be mainly facilitated by soft fluctuational vibrational modes allowing the harmonic approximation of their dynamics.

The interaction between the electronic variable and the structure variables r_n and θ_{nn-1}^2 stems from the parameter dependence of the electronic parameters E_n and V_{nn-1}. The diagonal term expressing the most efficient coupling is of the form

$$E_n = E_n^0 + kr_n \tag{5}$$

and takes into account the modulation of the on-site electronic energy E_n^0 by the radial vibrations of the base pairs. In turn, the actual charge occupation has its impact on the local radial distortion of the helix. As for as the transfer matrix elements V_{nn-1} are concerned, we assume that they depend on the three-dimensional distance between two consecutive bases along a strand in the following fashion

$$V_{nn-1} = V_0(1-\alpha d_{nn-1}) \tag{6}$$

The quantity α regulates how strongly V_{nn-1} is influenced by the distance and the latter is determined by

$$\begin{aligned} d_{nn-1} = \{&a^2 + (R_0 + r_n)^2 + (R_0 + r_{n-1})^2 \\ &- 2(R_0 + r_n)(R_0 + r_{n-1})\cos(\theta_0 + \theta_{nn-1})\}^{1\backslash 2} - l_0 \end{aligned} \tag{7}$$

with

$$l_0 = \sqrt{\mathrm{a}^2 + 4R_0^2\sin^2(\theta_0/2)} \tag{8}$$

Expanding the expression (7) up to first order around the equilibrium positions gives

$$d_{nn-1} \approx \frac{R_0}{l_0}\left[(1-\cos\theta_0)(r_n + r_{n-1}) + \sin\theta_0 R_0 \theta_{nn-1}\right] \tag{9}$$

Realistic parameters for DNA molecules are given by [34,35]: $a = 3.4$ Å, $R_0 \approx 10$ Å, $\theta_0 = 36°$, $J = 4.982 \times 10^{-45}$ kgm^2, $\Omega_\theta = [0.526-0.744] \times 10^{12}$ s^{-1}, $\Omega_r = 6.252 \times 10^{12}$ s^{-1}, $V_0 = 0.1$ eV and $M = 4.982 \times 10^{-25}$ kg

We scale the time according to $t \to \Omega_r t$ and introduce the dimensionless quantities:

$$\begin{gathered}\widetilde{r}_n = \sqrt{\frac{M\Omega^2}{V_o}}\, r_n,\ \widetilde{k}_n = \frac{k_n}{\sqrt{M\Omega_r^2 V_0}},\ \widetilde{E}_n^0 = \frac{E_n^0}{V_0} \\ \widetilde{\Omega} = \frac{\Omega_\theta}{\Omega_r},\ \widetilde{V} = \frac{V_0}{J\Omega_r^2},\ \widetilde{\alpha} = \sqrt{\frac{V_0}{M\Omega_r^2}}\,\alpha,\ \widetilde{R}_0 = \sqrt{\frac{M\Omega_r^2}{V_0}}R_0\end{gathered} \tag{10}$$

Subsequently we omit the tildes.

With the use of the expression (9) the equations of motion derived from the Hamiltonian (2)–(4) read as

$$\begin{aligned} i\tau\dot{c}_n = {} & (E_n^0 + kr_n)c_n \\ & -(1-\alpha d_{n+1,n})c_{n+1} - (1-\alpha d_{nn-1})c_{n-1}\end{aligned} \tag{11}$$

$$\begin{aligned} \ddot{r}_n = {} & -r_n - k|c_n|^2 - \alpha\frac{R_0}{l_0}(1-\cos\theta_0) \\ & \times\left\{[c_{n+1}^* c_n + c_{n+1}c_n^*] + [c_n^* c_{n-1} + c_n c_{n-1}^*]\right\}\end{aligned} \tag{12}$$

$$\ddot{\theta}_{nn-1} = -\Omega^2\theta_{nn-1} - \alpha V\frac{R_0^2}{l_0}\sin\theta_0[c_n^* c_{n-1} + c_n c_{n-1}^*] \tag{13}$$

and the quantity $\tau = \hbar\Omega_r/V_0$ appearing in Eq. (11) determines the time-scale separation between the fast electron motion and the slow bond vibrations. We remark that in the limit case of $\alpha = 0$ and constant $E_n^0 = E_0$ the set of coupled equations represents the Holstein system widely used in studies of polaron dynamics in one-dimensional lattices [37]. Furthermore, in the linear limit case emerging for $\alpha = k = 0$, the Anderson model is obtained for random E_n^0 [38,39]. We remark that the underlying model is suitable for the study of charge transport in DNA in a vacuum when, in particular, the damping influence of the biological environment on the structural dynamics can be discarded. In a subsequent section we take the impact of viscous damping on the ET into account (see below in Section 22.6).

The values of the scaled parameters are given by $\tau = 0.0411$, $\Omega^2 = [0.709 - 1.417]\times 10^{-2}$, $V = 0.0823$, $R_0 = 34.862$ and $l_0 = 24.590$. In our model study, we treat the electron-mode coupling strengths k and α as adjustable parameters and adopt the values of these free parameters such that not too strong deformations of the helix result, which

is essential for our used harmonic treatment of the dynamics of the structural coordinates. Afterwards we study synthetic DNA molecules consisting of a single type of base pairs for which we are able to compute reliable values of the coupling strengths with quantum chemical methods (see further below).

22.3 LOCALIZED POLARON-LIKE STATES

Caused by the non-linear interplay between the electronic and the vibrational degrees-of-freedom of the helix, the formation of polaronic electron-vibration compounds is possible [20,21,24,26,40–43]. Being interested in the simulation of a non-linear charge transport mechanism in DNA, we embark as the first step on the construction of localized stationary solutions of the coupled system (11)–(13). Exploiting the fact that the adiabaticity parameter τ is small, i.e., the fast charge transport along the base stacks and the slow bond vibrations and the even slower rotational twist motions evolve on distinct time scales, the inertia in Eqs (12) and (13) is negligible. Therefore, one can solve in the adiabatic limit the resulting static equations, which results in the instantaneous displacements

$$\begin{aligned} r_n = & -k|c_n|^2 - \alpha \frac{R_0}{l_0}(1-\cos\theta_0) \\ & \times \left\{[c^*_{n+1}c_n + c_{n+1}c^*_n] + [c^*_n c_{n-1} + c_n c^*_{n-1}]\right\} \end{aligned} \tag{14}$$

$$\theta_{nn-1} = -\frac{\alpha V}{\Omega^2}\frac{R_0^2}{l_0}\sin\theta_0\,[c^*_n c_{n-1} + c_n c^*_{n-1}] \tag{15}$$

Upon insertion of (14) and (15) into Eq. (11) one obtains a non-linear discrete Schrödinger equation for the electronic amplitude

$$\begin{aligned} i\tau\dot{c}_n = & \left[E_n^0 - k^2|c_n|^2 - k\alpha\frac{R_0}{l_0}(1-\cos\theta_0)\right. \\ & \left.\times\left\{[c^*_{n+1}c_n + c_{n+1}c^*_n] + [c^*_n c_{n-1} + c_n c^*_{n-1}]\right\}\right]c_n \\ & -(1+b_{n+1n})c_{n+1} - (1+b_{nn-1})c_{n-1} \end{aligned} \tag{16}$$

with

$$\begin{aligned} b_{nn-1} = & \alpha\left(\frac{R_0}{l_0}\right)^2\left\{(1-\cos\theta_0)\left[k\frac{l_0}{r_0}[|c_n|^2 + |c_{n-1}|^2]\right.\right. \\ & + \alpha(1-\cos\theta_0)([c^*_{n+1}c_n + c_{n+1}c^*_n] \\ & \left.\left. + 2[c^*_n c_{n-1} + c_n c^*_{n-1}] + [c^*_{n-1}c_{n-2} + c_{n-1}c^*_{n-2}])\right] \right. \\ & \left. + \frac{\alpha V}{\Omega^2}R_0^2\sin^2\theta_0[c^*_n c_{n-1} + c_n c_n c^*_{n-1}]\right\} \end{aligned} \tag{17}$$

In order to search for stationary localized solutions (polaron states), we substitute $c_n = \Phi_n \exp[-iEt/\tau]$ in Eq. (16) and obtain the difference system

$$\begin{aligned} E\Phi_n = &\Big[E_n^0 - k^2\,|\Phi_n|^2 - k\alpha\frac{R_0}{l_0}(1-\cos\theta_0) \\ &\times \{[\Phi_{n+1}^*\Phi_n + \Phi_{n+1}\Phi_n^*] + [\Phi_n^*\Phi_{n-1}^* + \Phi_n\Phi_{n-1}^*]\}\Big]\Phi_n \\ &- (1+B_{n+1n})\Phi_{n+1} - (1+B_{nn-1})\Phi_{n-1} \end{aligned} \tag{18}$$

where B_{nn-1} is given by

$$\begin{aligned} B_{nn-1} = \alpha\left(\frac{R_0}{l_0}\right)^2\Bigg\{&(1-\cos\theta_0)\Bigg[k\frac{l_0}{R_0}[|\Phi_n|^2 + |\Phi_{n-1}|^2] \\ &+ \alpha(1-\cos\theta_0)([\Phi_{n+1}^*\Phi_n + \Phi_{n+1}\Phi_n^*] \\ &+ 2\,[\Phi_n^*\Phi_{n-1} + \Phi_n\Phi_{n-1}^*] + [\Phi_{n-1}^*\Phi_{n-2} + \Phi_{n-1}\Phi_{n-2}^*])\Bigg] \\ &+ \frac{\alpha V}{\Omega^2}R_0^2\sin^2\theta_0[\Phi_n^*\Phi_{n-1}^* + \Phi_n\Phi_{n-1}^*]\Bigg\} \end{aligned} \tag{19}$$

The ground state of the system (18) is computed with the numerical map method outlined in [44,45].

In Fig. 22.2 we depict typical patterns of polaron states. First we discuss the ordered (periodic) case arising, e.g., for synthetically produced DNA molecules consisting of a single type of base pairs (e.g., poly(G)–poly(C) DNA polymers) surrounded by vacuum, which provides realistic ideal conditions for the consideration of DNA conductivity in experimental and theoretical studies [16]. The electronic contribution is shown in Fig. 22.2(a). In the ordered case the electronic wave function is localized at the central lattice site and the amplitudes decay monotonically and exponentially with growing distance from this central site. We remark that the spatial extension of the polaron decreases with increased coupling strength κ and there is a smooth transition from large to small polarons. Analogously, enlarging the off-diagonal coupling strength α leads also to an enhanced degree of localization. The size of the polarons plays a crucial role for their mobility in the sense that for large up to medium polarons coherent motion can be activated, whereas small polarons are immobile due to their pinning to the discrete lattice [46,47]. The associated patterns of the static radial and angular displacements of the bases are shown in Fig. 22.2(b) and 22.2(c), respectively. Like the electronic wave function, the radial and angular displacements are exponentially localized at the central lattice site (base pair). Notice that due to the overall minus sign of the radial and angular excitation patterns, the H-bridges get compressed while the helix experiences a local unwinding around the occupation peak of the localized electron.

For a more realistic study we take static diagonal disorder in the on-site electronic energy E_n^0 into account. The random values caused, for example, by the inhomogeneous broadening of the sites of distinct base pairs with different energies were simulated by random potentials $[-\Delta E, \Delta E]$ with mean value $\overline{E} = 0$ and different mean standard deviations ΔE. (Note that any

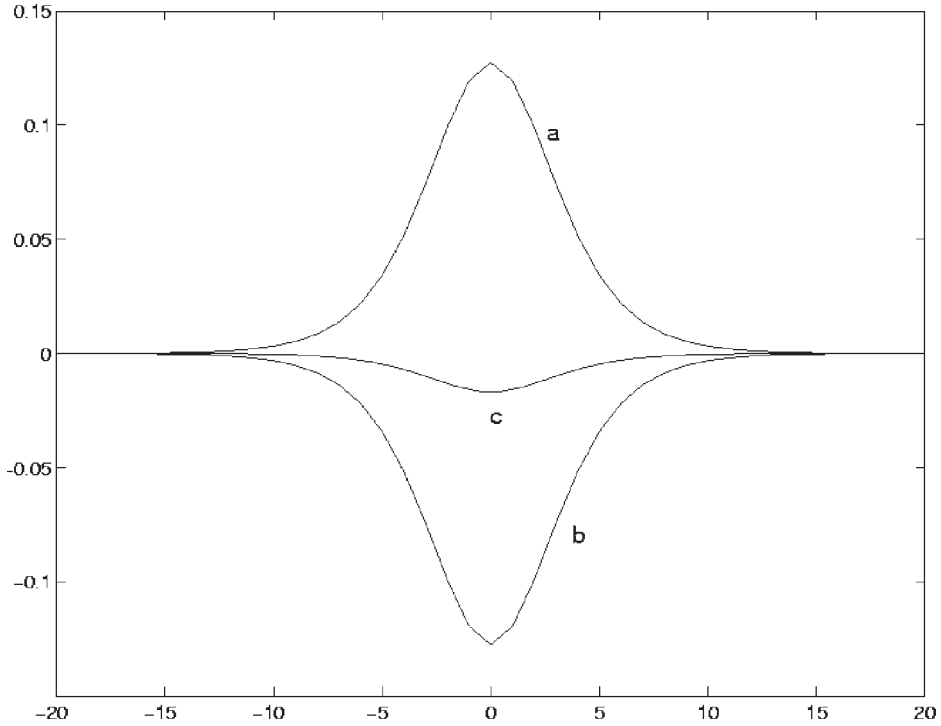


Fig. 22.2. Profiles of the stationary polaron states for the ordered (periodic) case. (a) The electronic amplitudes $|\Phi_n|^2$. (b) The static radial displacements r_n. (c) The static angular twists θ_n.

$E_0 c_n$ term on the r.h.s. of Eq. (11) with constant E_0 can be eliminated by a gauge transformation $c_n \rightarrow \exp(-iE_0 t/\tau) c_n$.)

Generally, the localized excitation patterns do not change qualitatively when diagonal disorder is taken into account. Compared with the ordered case, one merely observes a shift of the excitation peak away from the central lattice site depending on the respective realization of the random potential. However, since in the presence of diagonal disorder the translational invariance of the lattice system is broken, the localized amplitude pattern is no longer a reflection symmetric with respect to a distinguished lattice site as opposed to the corresponding symmetric localized state of the ordered system. Moreover, we find that the combined effect of the two localization mechanisms, that is, non-linear polaron and linear Anderson-mode formation, respectively, leads to enhancement of the degree of localization compared with the case when only one of the two mechanisms acts.

Furthermore, real DNA molecules exhibit random structural imperfections of their double helix caused, e.g., by the random base-pair sequence and the solvent conditions that may leave the involved helix structures in irregularly distorted shapes. Accordingly, the structural disorder is incorporated in our model by randomly distributed equilibrium positions of the bases so that the resulting irregular helical DNA matrix deviates from the perfectly regular helix structure.

To be precise, we consider randomly distributed torsional coordinates θ_{nn-1} and radii r_n. Since the positions (actually, the distances) govern the value of the corresponding transfer matrix elements $V_{nn-1}=V_0(1-\alpha d_{nn-1})$ non-diagonal disorder is induced in the latter at the same time. Random transfer integrals along the helix were used to represent random base-pair sequences for which the charge transport along DNA was discussed recently in a model relying on localized electron states [19].

22.4 CHARGE TRANSPORT ESTABLISHED BY MOBILE POLARONS

We study now the charge transport supported by mobile polarons propagating along DNA. To activate polaron motion an established method is provided by proper stimulation of localized internal normal modes of certain shape and frequency, which are also called pinning modes [48]. In this way kinetic energy is injected into the system in the form of an initial kick of the momentum variables of the vibrational degrees-of-freedom. Results regarding activation of polaron motion in the ordered case are presented in Fig. 22.3. We integrated the set of coupled non-linear equations (11)–(13) with a fourth-order Runge–Kutta method and the accuracy of the computation was checked through monitoring the conservation of the total energy as well as the norm $\Sigma_n|c_n(t)|^2=1$. The applied kicking strength is $\lambda_r= 0.02$ corresponding to an amount of radial kinetic energy of the order of 200 meV. Interestingly, we found that there exists no angular pinning mode component ξ_θ (see further below) so that alone the excitation of the radial component ξ_r produces long-lived stable polaron and breather motions. Hence, our findings suggest that higher-frequency radial fluctuational modes of DNA instigate the propagation of localized structures along DNA rather than low-frequency twist fluctuational modes.

Figure 22.3(a) demonstrates that the electronic component of the polaron moves with uniform velocity along the lattice maintaining its localized shape throughout the journey. In the same manner the radial polaron part travels as a localized wave along the strand so that the electron propagation is accompanied by a local and temporal contraction of the radii. However, the angular static polaron component splits into two parts when the radial pinning mode is initially imposed. At the starting position there remains a standing breather of comparatively large amplitudes which reverse periodically their sign in the course of time. Thus the helix experiences alternately a local winding and unwinding around its central base pair. In addition a second breather of relatively small and exclusively negative amplitude propagates in unison with the electronic and radial polaron components. The result is that the region of the helix which is traversed by the electron gets locally untwisted. As far as the quantitative assessment of the charge transport is concerned we note that the breather travels across 70 sites (base pairs) in a 1000 (dimensionless) time units, meaning that physically the charge conduction across the base pairs of DNA over a distance of 238 Å takes $\approx$0.16 ns leading to a higher transport rate than a typical one found for protein conductivity [19]. This underlines the status ascribed to DNA molecules as promising candidates establishing fast and efficient charge transport in (albeit only theorized yet) molecular electronic engineering, utilizing them as molecular wires.

In the presence of *diagonal static disorder*, that is, when the on-site electronic energies E_n^0 are represented by random numbers distributed in the interval $|E_n^0|<\Delta E$ with width

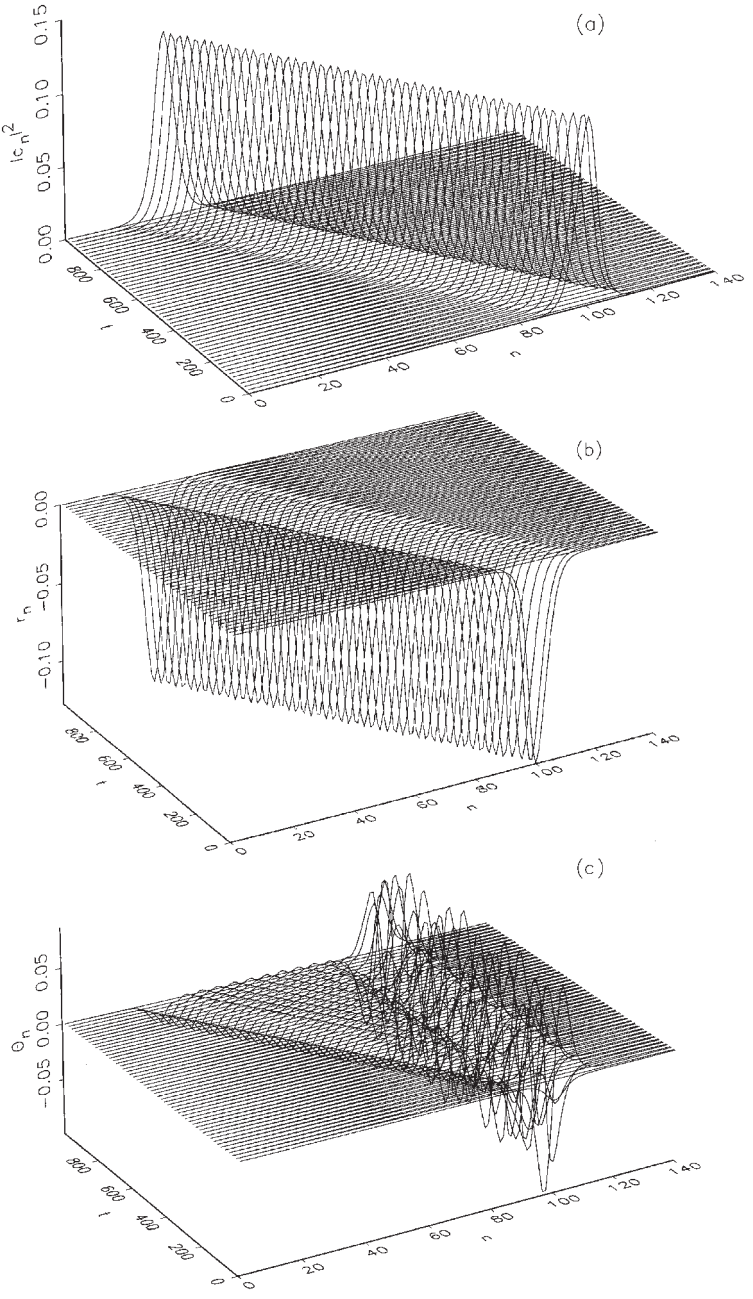


Fig. 22.3. Breather motion along the DNA in the ordered case. Motion is initiated by an initial kick of the velocities $\{\dot{r}_n\}$ directed along the corresponding pinning mode component. The kick amplitude is $\lambda_r = 0.25$. (a) Uniformly moving electronic breather. (b) The associated moving breather in the radial displacements. (c) The angular breather contribution consisting of a small amplitude breather pinned at the starting site and a relatively large-amplitude moving breather component propagating in alliance with the electron and radial breathers.

ΔE, we find that mobile breathers exist up to a critical degree of disorder ΔE_{crit} and ET persists. (Let us stress that it was indeed experimentally found that the random sequence $\lambda-$DNA is an electrical conductor [8].) Beyond this critical value, continuous propagation of the breathers along the lattice is prevented. More precisely, while for a relatively small amount of disorder, $\Delta E = 0.05$, the electronic breather (together with its radial and torsional counterparts) is able to travel uniformly along the lattice for a 10 times increased $\Delta E = 0.5$, the breather motion is restricted to a neighbourhood of the initially excited position around which it wanders in an oscillatory manner. The temporal evolution of the electron breather centre is shown in Fig. 22.4.

Concerning the impact of *structural disorder* we considered random distributions of the base-pair spacings $r_n - r_0 \in [-\Delta,\Delta]$ and the twist angles $\theta_n - \theta_0 \in [-\Delta,\Delta]$ with different mean standard deviations Δ.

For representative results we averaged over several realizations of structural disorder. The results for the temporal evolution of the electron breather centre are depicted in Fig. 22.5. We conclude that structural disorder does not significantly affect the mobility of the breathers. For growing degree of disorder merely the propagation velocity gets reduced. Even in the case of fairly strong disorder $\Delta = 10$ percent (dashed line) the breather is still able to propagate steadily.

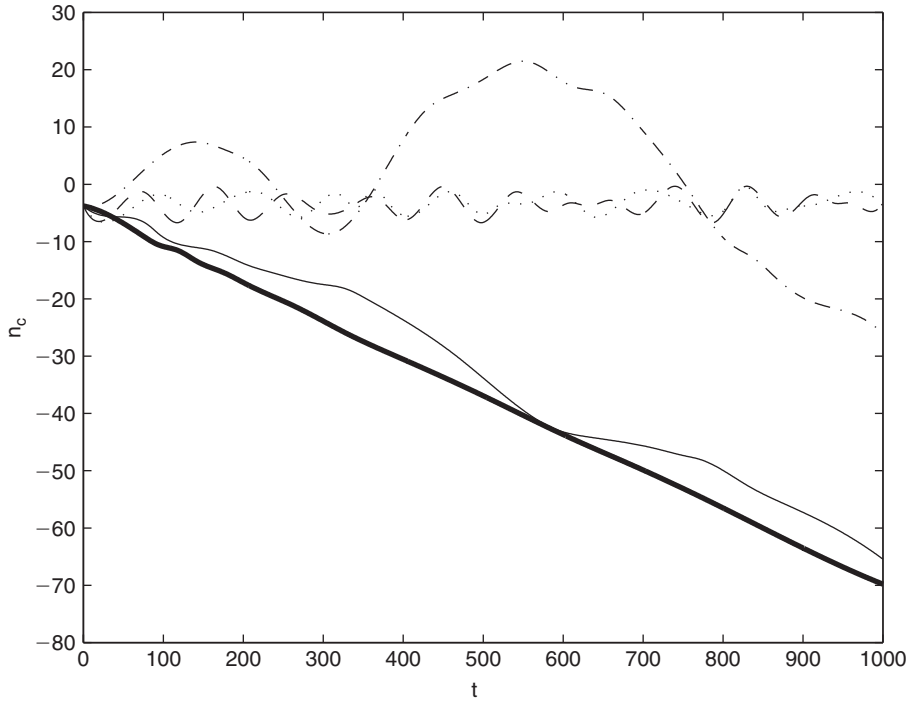


Fig. 22.4. Diagonal disorder. The position of the centre of the electron breather as a function of time and for different amounts of disorder. Full line (ordered case), short dashed line ($\Delta E = 0.025$), dashed–dotted line ($\Delta E = 0.050$), dotted line ($\Delta E = 0.100$) and long dashed line ($\Delta E = 0.500$).

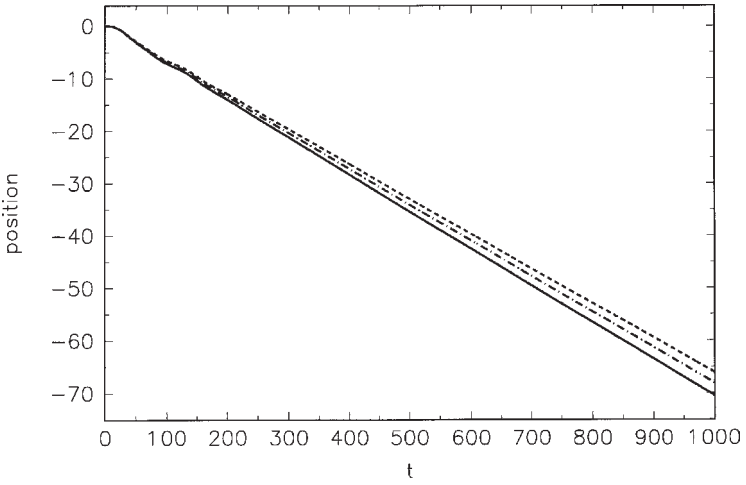


Fig. 22.5. Structural disorder and random double-helix structure. The temporal behaviour of the position of the electron breather centre for different degrees of randomness in the radii r_n and angles θ_n. Assignment of the line types to the increasing mean standard deviations ranging up to 10 percent of the respective mean value: full line (1 percent), dashed–dotted line (5 percent) and dashed line (10 percent).

We also studied the polaron mobility under the impact of combined static diagonal and structural disorder and found that the strong pinning effect generated by static diagonal disorder prevails over the more harmless and merely velocity-reducing influence of structural irregularities. Particularly, the last result differs from the opposing finding reported in [49] suggesting that hopping conductivity in proteins and DNA seems to be influenced first of all by the molecular conformations and not by their sequences.

22.5 CHARGE TRANSPORT IN POLY(DA)–POLY(DT) AND POLY(DG)–POLY(DC) DNA POLYMERS

In this section we theoretically describe the charge transport in synthetic DNA polymers built up from single types of base pairs on the basis of the model introduced above.

Concerning the interaction between the electronic and the vibrational degrees-of-freedom the quantum chemical computation of the geometry dependence of the electronic parameters E_n and $V_{nn\text{-}1}$ reveals that their most significant modulation originates from radial distortions of the helix which are related to hydrogen and covalent bond deformations, respectively. However, the influence of small-angle deformations on the values of the electronic parameters can be discarded. Hence the corresponding twist part of the Hamiltonian (4) can be omitted in the forthcoming study.

Furthermore, unlike in the previous studies, we use for our computations credible values for the electron-mode coupling strengths κ and α, as a result of quantum chemical computational procedure described in Ref. [50]. To this end, quantum chemical calculations have been performed on symmetrical homodimers consisting of two nucleoside

Watson–Crick base pairs (adenosine–thymidine (AT) and guanosine–cytidine (GC) base-pair steps (BPS)) stacked over each other, to mimic the conventional A-DNA and B-DNA conformations [51]. Taking into account the DNA backbone at least in the form of intact sugar moieties, instead of substituting it with protons or methyl groups, is necessary for the consistency of the calculations [52] and to correctly describe charge transfer through DNA duplexes [53]. E.B. Starikov [54] has used a semi-empirical all-valence-electron PM3 Hamiltonian [55] within the MOPAC7 version of CI (configuration interaction) approximation. We have chosen the PM3-CI method, but not an *ab initio* one, since:

(a) The molecular fragments involved are very large.
(b) Similar *ab initio* calculations even for smaller segments experience difficulties with SCF convergence [56].
(c) The work [52] used similar semi-empirical Hamiltonian (AM1) for analogous molecular fragments.
(d) PM3-CI approximation is good at describing excited states of nucleoside base-pair steps [54].
(e) CI approximation is indispensable when charged states of nucleic acid bases are considered using semi-empirical quantum chemistry (cf., e.g., [57] and references therein).

Here the effects of small, but non-negligible (up to 0.1 Å), radial stretching–compressing of Watson–Crick hydrogen bonds (WC H-bonds) on the MO energies of the BPS under study have been estimated. Specifically, in the above sense, the equilibrium lengths of the WC H-bonds are perturbed in only one of the base pairs in all the BPS involved, and the resulting changes in the energies of the highest occupied molecular orbital and of the occupied molecular orbital next to the highest one (HOMO and HOMO-1, respectively), as compared with the equilibrium values of these energies, are monitored. According to the Koopmans theorem, well known in quantum chemistry (cf., e.g., [56] and references therein), the HOMO energy is approximately equal to the molecular ionization potential and, in the tight-binding approximation, could be viewed as the site energy. Along with this, the difference between the HOMO and HOMO-1 energies is approximately twice the hopping integral. With these considerations, the κ and α parameters of our model can be estimated by calculating linear regression of the corresponding site energy and hopping integral changes, respectively, onto WC H-bond distance perturbations. A very tight linear correlation between the former and the latter ones is achieved. Interestingly, there is not any correlation between the tight-binding Hamiltonian parameter changes and BPS twist angle/helical pitch perturbations ('helical pitch' is the distance a in Eq. (8)). The perturbations in the two latter parameters are also relatively small (up to 10° and 0.1 Å, respectively). For the coupling parameters of the (dA)–poly(dT) DNA polymer, the values thus obtained are: $k = 0.0778917$ eV/Å and $\alpha = 0.053835$ Å^{-1}. The corresponding values for the poly(dG)–poly(dC) DNA polymer are determined as $k = -0.090325$ eV/Å and $\alpha = 0.383333$ Å^{-1}.

Like in the previous study, we construct stationary localized (polaronic) electron–vibration solutions with the help of the non-linear map approach. In Fig. 22.6 we depict the profiles of the (standing) polaron states for the poly(dA)–poly(dT) and the poly(dG)–poly(dC) DNA polymer, respectively. In both cases, the polarons are of fairly

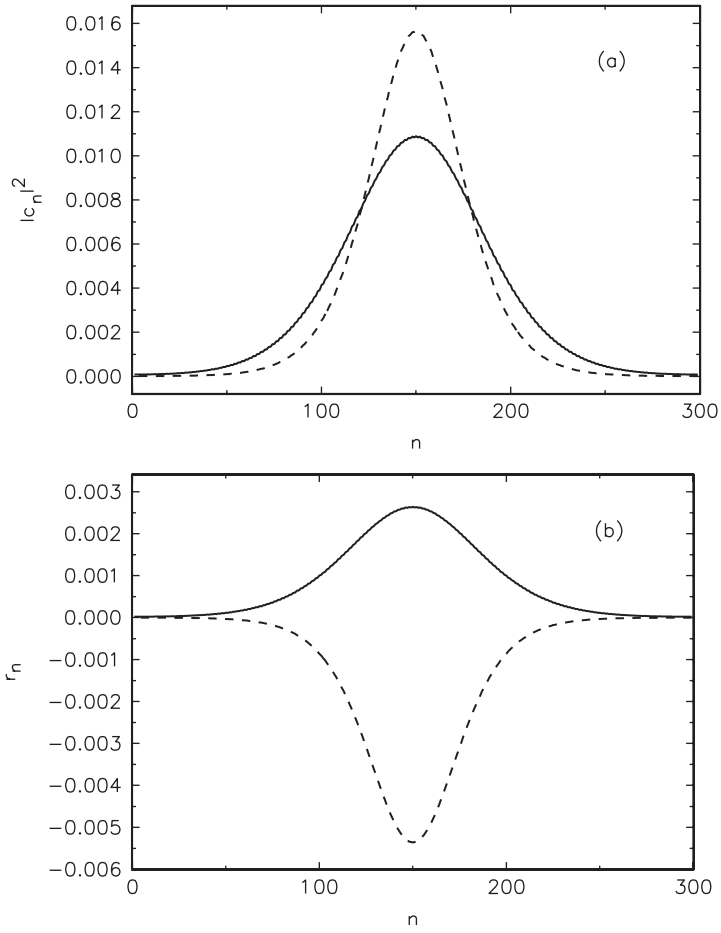


Fig. 22.6. The spatial pattern of the polaronic electron–vibration compound. (a) The electronic part. Full (dashed) line: poly(dA)–poly(dT) (poly(dG)–poly(dC)) DNA polymer. (b) Radial deformation pattern. Assignment of line types as in (a).

large extension (width). Regardless of the DNA polymer type, the electronic wavefunction is localized at the central lattice site and the envelope of the amplitudes decays monotonically and exponentially with growing distance from this central site (base pair). However, the electronic wavefunction of the poly(dG)–poly(dC) DNA polymer is stronger localized than that of its poly(dA)–poly(dT) counterpart. Accordingly, the attributed radial displacement patterns are exponentially localized at the central lattice site. Concerning the resulting static radial helix deformations, we find that there is a drastic difference between the poly(dA)–poly(dT) and the poly(dG)–poly(dC) DNA polymers. In the former case, the overall non-positive radial amplitudes imply that the H-bridges experience contractions. In contrast, in the latter case the H-bridges get stretched. Nevertheless, all these deformations are rather weak, i.e., of the order of 1.510^{-3}Å.

As for as the ET is concerned, polaron motion is activated with the use of the discrete gradient method [58]. With this method we obtain suitable initial perturbations of the momentum coordinates p_n^r which initiate coherent motion of the polaron compound.

We consider first the case of the poly(dG)–poly(dC) DNA polymer. Interestingly, only the electronic component sets off to move directionally along the lattice, whereas the vibrational amplitude pattern remains at the initial position. Hence, long-range ET is achievable. Contrary to that, we found that the poly(dA)–poly(dT) DNA polymer does not exhibit such a good charge mobility as its poly(dG)–poly(dC) counterpart. The different propagation scenarios are properly illustrated using the time evolution of the first momentum of the electronic occupation probability defined as $\overline{n}(t)=\Sigma_n n|c_n(t)|^2$. One can clearly observe that, in the (dG)–(dC) case, electron propagation proceeds uninhibitedly with uniform velocity (see Fig. 22.7). In comparison, the electron within the poly(dA)–poly(dT) travels with smaller velocity and, after having traversed eight lattice sites, its motion stops, so that the electron eventually becomes trapped.

In conclusion, we have found that conductivity in synthetically produced DNA molecules depends on the type of the single base pair of which the polymer is built. While a polaron-like mechanism (relying on the non-linear coupling between the electron amplitude and radial vibrations of the base pairs) is responsible for long-range and stable ET in (dG)–(dC) DNA polymers, the charge mobility is comparatively weak in the case of (dA)–(dT) DNA polymers.

Especially when it comes to designing synthetic molecular wires, these findings might be of interest. In fact, recent experiments suggest that ET through DNA molecules proceeds by polaron hopping [31]. Furthermore, our results comply with the findings of

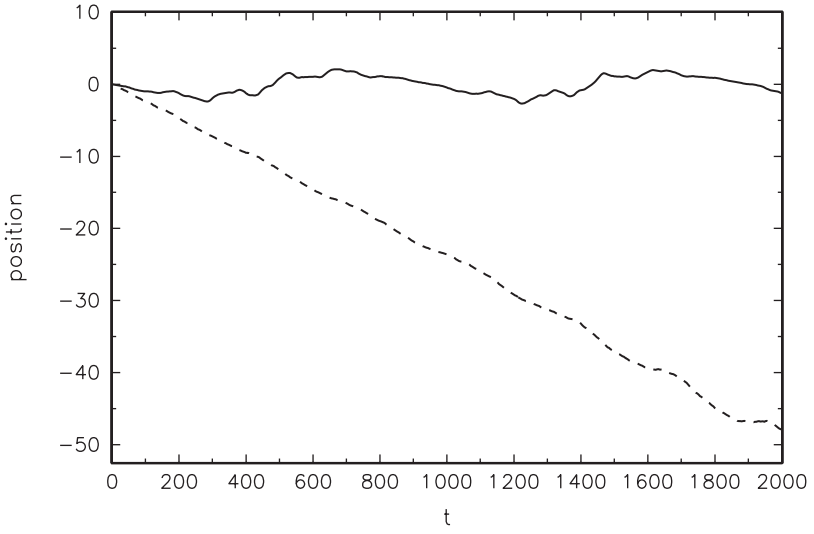


Fig. 22.7. Time evolution of the first momentum of the electronic occupation probability. Full (dashed) line: poly(dA)–poly(dT) (poly(dG)–poly(dC)) DNA polymer.

these experiments, which show also that poly(dG)–poly(dC) DNA polymers forms a better conductor than their poly(dA)–poly(dT) counterparts.

22.6 CONTROL OF ET IN DNA

In this section we investigate the combined effect of energetic disorder, viscous damping and an external field on the electron transfer in DNA where we focus our interest on the instigation of ET by the applied electric field. To this end, the Hamiltonian given in Eqs (2)–(4) is augmented by a further part representing the electron-field interaction

$$H_{\text{field}} = -eE(t)\sum_n na|c_n|^2 \tag{20}$$

with $E(t) = E_0\sin(\omega t)$ is the time-dependent periodic electric field directed along the strands and e is the electron charge.

The modified equations of motion then read as

$$\begin{aligned} i\tau\dot{c}_n = {} & (E_n^0 + kr_n)c_n - i\gamma c_n - \beta n d_{nn-1}\sin(\omega t)c_n \\ & - (1-\alpha d_{n+1,n})c_{n+1} - (1-\alpha d_{nn-1})c_{n-1} \end{aligned} \tag{21}$$

$$\ddot{r}_n = -r_n - k|c_n|^2 - \frac{R_0}{l_0}(1-\cos\theta_0) \tag{22}$$

$$\times \left\{\alpha[c_{n+1}^* c_n + c_{n+1}c_n^*] + [c_n^* c_{n-1} + c_n c_{n-1}^*]) - \beta n\sin(\omega t)|c_n|^2\right\}$$

$$\ddot{\theta}_{nn-1} = -\Omega^2\theta_{nn-1} - \frac{R_0^2}{l_0}\sin^2\theta_0 V\{\alpha\,[c_n^* c_{n-1} + c_n c_{n-1}^*] - \beta n\sin(\omega t)|c_n|^2\} \tag{23}$$

For a more realistic model of ET in DNA, we have incorporated the effect of the viscosity of the solvent associated with the friction of the electron motion (electronic energy dissipation), which is described by the additional damping term $-i\gamma c_n$ on the r.h.s. of Eq. (21). The damping constant lies in the range $\gamma = [0.001,0.01]$ corresponding to 'life times' on time scales ranging from 10ps to 100ps relevant for biological ET.

We demonstrate that the coupling of the electron–vibron system, that is, the localized standing electron–vibron compound, to the periodically varying external field, initiates directed motion. Furthermore, despite the electronic energy dissipation (related to the γ–term) and energetic disorder, long-lived stable ET can be accomplished. Concerning the activation of breather motion, we observe that the immediate initial energy injection through the external field supplies kinetic energy and the breather starts to move in a direction dictated by the initial phase of the external field. Through this driving due to the external periodic field, there is a permanent energetic input, which is also necesssary to compensate for the incessant energetic losses due to friction. We underline that, to render the stationary states mobile, no special pattern of the imposed external field is requested, which rather can be of a general periodic type.

To illustrate the typical propagation features of the mobile polarons, we represent in Fig. 22.8 the spatio-temporal evolution of the electronic, the radial and torsional polaron components, respectively.

As Fig. 22.8(a) clearly shows, under the impact of the external periodic field the (standing) electron becomes mobile and propagates directionally along the lattice with a velocity oscillating around a constant mean value (see further below). In addition, the amplitude pattern breathes with the frequency of the applied periodic field, so that an electronic breather arises. At each instant of time corresponding to a period duration $T = 2\pi/\omega$ of the external field from the moving large-amplitude breather, a small-amplitude breather is radiated which performs oscillatory (but confined) motions around the lattice position at which it has emerged. The energy content of the main moving breather $E_{\mathrm{el}} = \sum_n [E_n|c_n|^2 - V_{nn-1}(c_n^* c_{n-1} - V_n c_n c_{n-1}^*)$ gets periodically modulated by the external field and performs oscillations around a constant mean value $\overline{E}_{\mathrm{el}}$ in the course of time. There remains some excess energy, which can neither be absorbed and carried by the mobile main breather nor is it needed to compensate the energy losses due to friction. This excess energy is deposited into the lattice in the form of additional electronic small-amplitude pinned breathers. Nonetheless, by a suitable choice of the amplitude β of the applied field, the energy balance between the frictional energy loss and the external energy input can be optimized in the sense that there remains (virtually) no excess energy and the 'radiational' creation of the additional small-amplitude breathers is suppressed.

Similarly, the radial breather component depicted in Fig. 22.8(b) consists of a moving small-amplitude breather accompanying the electron and additional 'radiationally' generated standing breathers, the amplitudes of which are modulated by the external periodic field. We emphasize that no significant structural deformations of the double helix develop, and the maximal amplitude of the stretchings of the hydrogen bonds from the equilibrium length is restricted to values less than 0.04 Å, justifying the harmonic treatment of the corresponding bond potential. In the case of the torsional breather component, the radiation effect is not as pronounced as for their electronic and radial counterparts. Nevertheless, there still appears a moving torsional breather which propagates in unison with its electronic and radial counterparts. Meanwhile, the amplitudes of this moving torsional breather are small compared with the amplitudes of the slowly varying and immobile breather resting at the starting site. Hence, ET in DNA mediated by breathers is connected mainly with radial displacements rather than angular deformations along the transfer path. (This result is in accordance with our findings presented in Section 22.4.)

For the sake of further illustration of the ET, in Fig. 22.9 we display the velocity of the electron breather, as a function of the field amplitude β and for different values of the driving frequency ω. Interestingly, the dependence of the breather velocity on the field amplitude exhibits a resonance structure and the larger the driving frequency is, the broader is the resonance curve, whereas its peak position shifts towards larger field strengths. Therefore, for each frequency of the field, an optimal field amplitude exists which provides a maximal possible breather velocity. The value of the latter is the higher the lower the frequency of the applied field. In conclusion, with an appropriate choice of the amplitude β of the external field, the velocity of the ET can be tuned. We found that coherent ET can be stimulated when the frequency of the applied field lies in the range $0.1 < \omega < 3.25$.

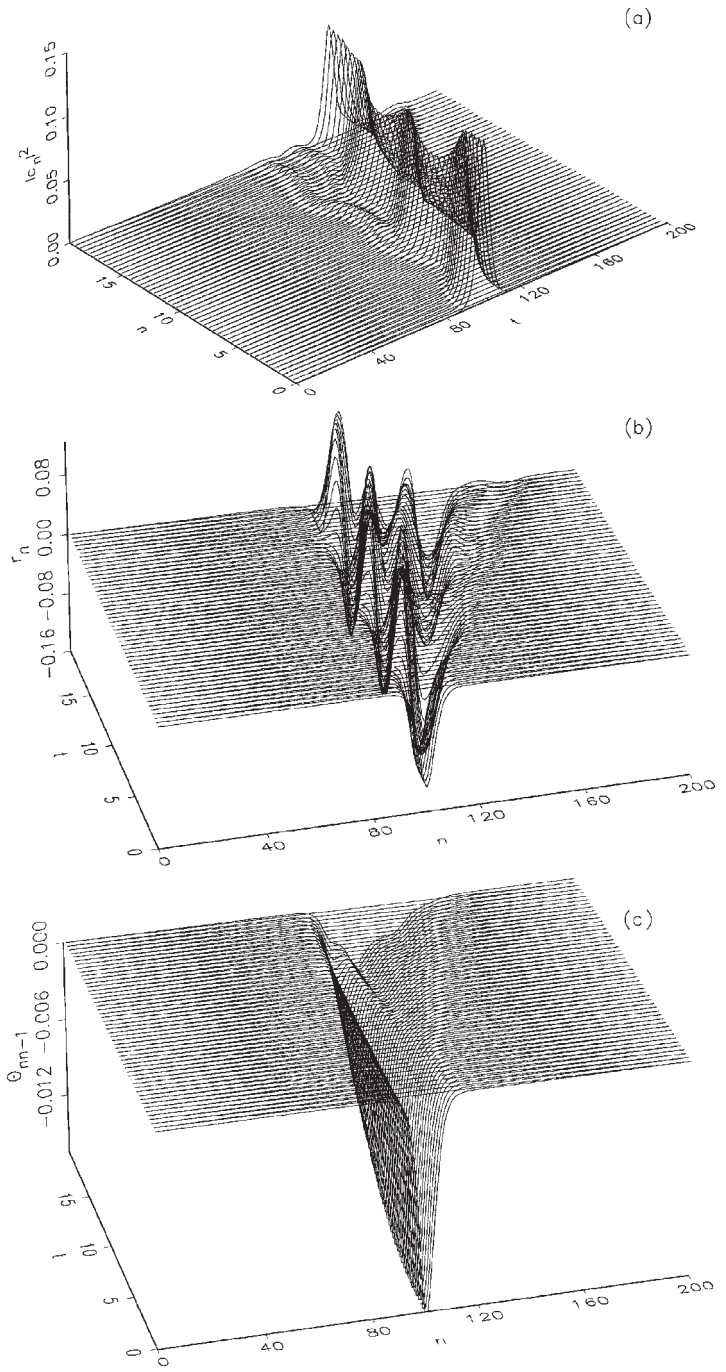


Fig. 22.8. Breather motion along the DNA in the ordered case. Parameters $\gamma = 0.001$, $\omega = 1$ and $\beta = 0.1$. (a) Electronic breathers. (b) Radial breathers. (c) Angular breathers.

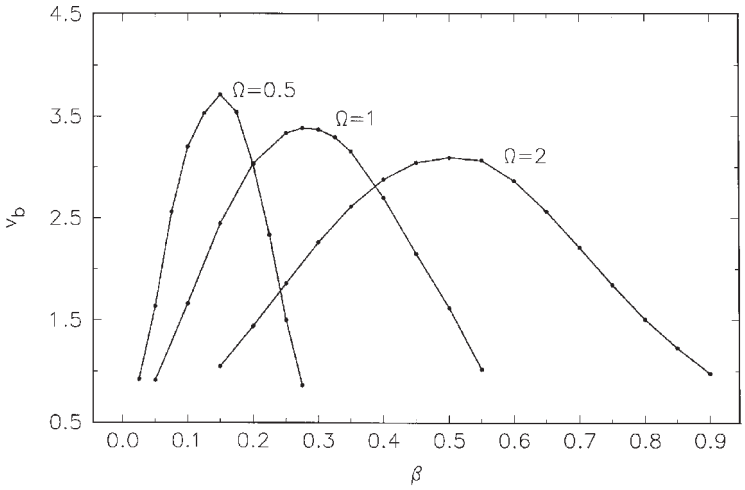


Fig. 22.9. The velocity of the moving electron breather as a function of the amplitude of the external periodic field β for different frequencies, as indicated on the graphs.

As for as the influence of disorder is concerned, we find that the stronger the degree of disorder is, the slower propagates the corresponding breather. Nevertheless, for an undercritical amount of disorder, ΔE , the localized states still maintain their non-linear breather character to such an extent that the initiation of their motion is possible up to a moderate strength of disorder $\Delta E < 0.25$, and thus conductivity persists. It has indeed been experimentally found that the random sequence of $\lambda-$DNA is an electrical conductor [8].

We demonstrate now that the application of a low-frequency external field offers the possibility of controlling the range of ET in DNA. Our results are summarized in Fig. 22.10 depicting the temporal behaviour of the first momentum of the electronic occupation probability for relatively strong disorder $\Delta E = 0.2$. The frequency of the imposed external field is $\omega = 0.1$ and its amplitude β is varied. For $\beta = 0.2$ we find that, in an initial phase, the centre of mass of the electron breather propagates steadily away from its starting position and moves over more than 30 sites corresponding to the directional long-range ET. Then the direction of the motion is reversed but the electron breather does not return completely to its starting position. Eventually, the dynamics of the electron breather can be viewed as oscillations around a fixed lattice site (base), being 22 sites apart from the starting position. Remarkably, when the amplitude β is enlarged, the maximum of these oscillations diminishes, while their frequency increases. Furthermore, the central position, around which the oscillations take place, is shifted downwards to lower lattice sites. Conclusively, with the choice of an appropriate amplitude of the low-frequency external field, the control of the range of the electron excursion along the DNA is possible.

The results of our numerical simulations suggest that in order to activate and control the ET in DNA, strong electrical fields with field strengths of the order of 10^6–10^7 V/cm and oscillations, which are fast compared with the characteristic times of electronic transfer processes, are needed. In Refs [59–62] the proposed control of longe-rang ET in molecular systems utilizing a strong periodic field is based on similar assumptions with

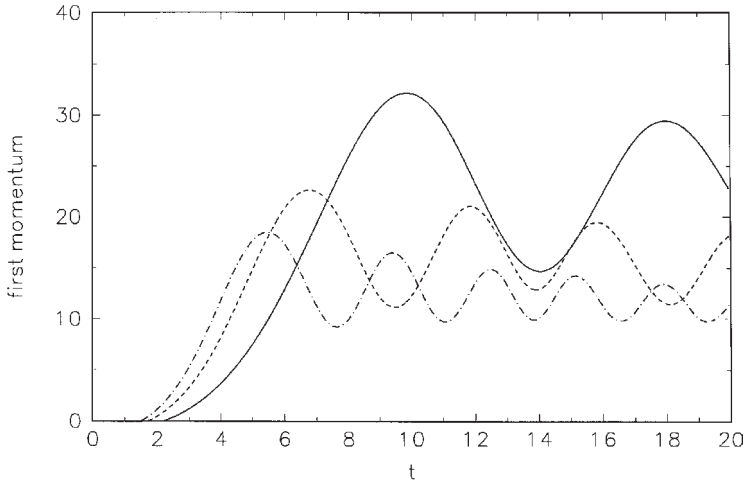


Fig. 22.10 The temporal behaviour of the position of the electron breather centre in the presence of a low-frequency external field. Parameters $\omega = 0.1$ and diagonal disorder $\Delta E = 0.1$. Full line ($\beta = 0.2$), dashed line ($\beta = 0.4$) and (dashed–dotted line ($\beta = 0.6$).

respect to the field strengths and the time-scale separation between the fast oscillations of the external field and the electron transfer rate.

22.7 SUMMARY

We have reviewed studies on the ET mechanism in DNA, with special reference to the interaction of the charge carrying unit with the vibrational modes of DNA, responsible for the formation of polarons and electron–vibron breathers. The steric structure of the double helix of twisted DNA has been described by a network system of coupled oscillators in the context of the base-pair picture considering angular and radial vibrational motions of the base pairs. The coupling between the electronic and vibrational degrees-of-freedom is established through modulations of the parameters of the electronic system (described by a tight-binding system), caused by structural changes of the DNA molecule.

In this chapter, we have focused special attention on the initiation of long-range, stable polaron and breather motion along the DNA structure, when static and/or structural disorder inherent in any real DNA molecule is taken into account (intrinsic static parametrical and/or structural disorder, reflecting structural imperfections and the impact of ambient solvent coordinates). As the first step, we have constructed exact stationary localized solutions (polarons) with the help of a non-linear map approach. Motion of the polaron is then activated through a kick mechanism utilizing the pinning mode. Alternatively, we have demonstrated that an applied electric field activates the long-range coherent motion of polarons and/or breathers travelling along the bases of the DNA strands.

We have shown that the directed motion prevails when the electron–vibration system is subjected to randomness. Even for a moderate amount of parametrical and structural

disorder the polaron compounds remain in localized shapes and support directed, long-range coherent electron transfer along the DNA double-helical strands. Regarding the vibrational degrees-of-freedom of the DNA, we have observed that the electron is accompanied by a vibrational breather in the radial variable causing local and temporal deformations of the traversed region of the double helix. Therefore, the efficient charge transport in DNA proceeds under the condition that the double helix undergoes structural changes, which is exemplary for the interplay of structure and function in flexible and adaptive biomolecules. Conclusively, DNA molecules might be useful for the design of functional nanowires as constituents of molecular nanoelectronic devices.

Finally, varying the amplitude of an applied external field, it turns out to be possible to tune the propagation velocity of the breather, which can be exploited to control ET in DNA. In addition, the electron breather can be steered to a desired lattice position through amplitude-tuning of the applied field, where the chosen initial phase of the latter governs the propagation direction. In this way even targeting the ET in DNA is achievable.

22.8 ACKNOWLEDGEMENTS

This work was sponsored by Ministerio de Educación y Ciencia, Spain, project FIS 2004–01183.

22.9 REFERENCES

1 M.D Purugganan, C.V. Kumar, N.J. Turro and J.K. Barton, Science, 247 (1988) 6548; D.N. Beratan, J.N. Onuchic, J.R. Winkler and H.B. Gray, Science, 258 (1997) 1740; C.J. Murphy, M.R. Arkin, Y. Jenkins, N.D. Ghatlia, S.H. Bossmann, N.J. Turro and J.K. Barton, Science, 262 (1993) 1025; M.R. Arkin, E.D.A. Stemp, R.E. Holmlin, J.K. Barton, A. Hormann, E.J.C. Olson and P.F. Barbara, *ibid,* 273 (1996) 475; P.J. Dandliker, R.E. Holmlin and J.K. Barton, *ibid,* 257 (1997) 1465.
2 D.D. Eley and D.I. Spivey, Trans. Faraday Soc., 58 (1962) 411.
3 D.B. Hall, R.E. Holmkin and J.K. Barton, Nature, 382 (1996) 731.
4 D.B. Hall and J.K. Barton, J. Am. Chem. Soc., 119 (1997) 5045.
5 M.R. Arkin, E.D.A. Stemp, S.C. Pulver and J.K. Barton, Chem. Biol., 4 (1997) 369.
6 Y. Okahata, T. Kobayashi, K. Tanaka and M. Shimomura, J. Am. Chem. Soc., 120 (1998) 6165.
7 E. Meggers, M.E. Michel-Beyerle and B. Giese, J. Am. Chem. Soc., 120 (1998) 12950.
8 H.-W. Fink and C. Schönenberger, Nature, 398 (1999) 407.
9 P. Tran, B. Alavi and G. Gruner, Phys. Rev. Lett., 85 (2000) 1564.
10 F.D. Lewis, T.F. Wu, Y.F. Zhang, R.L. Letsinger, S.R. Greenfield and M.R. Wasielewski, Science, 277 (1997) 673.
11 C.A. Mirkin, R.L. Letsinger R.C. Mucic and J.J. Storhoff, Nature, 382 (1996) 607; A.A. Voityuk, N. Rösch, M. Bixon and J. Jortner, J. Phys. Chem., 104 (2000) 9740; F.C. Grozema, Y.A. Berlin and L.D.A. Siebbels, J. Am. Chem. Soc., 122 (2000) 10903; Y.A. Berlin, A.L. Burin and M.A. Ratner, J. Am. Chem. Soc., 123 (2001) 260.
12 M. Ratner, Nature, 397 (1999) 480.
13 B. Giese, J. Amaudrut, A.K. Köhler, M. Spormann and S. Wessely, Nature, 412 (2001) 318 and references therein.
14 A.Yu. Kasumov, M. Kociak, S. Gueron, B. Reulet, V.T. Volkov, D.V. Klinov and H. Bouchiat, Science, 291 (2001) 280.
15 E. Braun, Y. Eichen, U. Sivan and G. Ben-Yoseph, Nature, 391 (1998) 775.

16 D. Porath, A. Bezryadin, S. de Vries and C. Dekker, Nature, 403 (2000) 635.
17 P.J. de Pablo, F. Moreno-Herrero, J. Colchero, J. Gómez Herrero, P. Herrero, A.M. Baró, P. Ordejón, J.M. Soler and E. Artacho, Phys. Rev. Lett., 85 (2000) 4992.
18 J. Jortner, M. Bixon, T. Langenbacher and M.E. Michel-Beyerle, Proc. Nat. Acad. Sci. USA, 95 (1998) 12759.
19 Y.-J. Ye, R.-S. Chen, A. Martinez, P. Otto and J. Ladik, Sol. Stat. Comm., 112 (1999) 139.
20 R. Bruinsma, G. Grüner, M.R. D'Orsogna and J. Rudnick, Phys. Rev. Lett., 85 (2000) 4393.
21 D.M. Basko and E.M. Conwell, Phys. Rev. E, 65 (2002) 061902.
22 Z.G. Yu and X. Song, Phys. Rev. Lett., 86 (2001) 6018.
23 M. Hjort and S. Stafström, Phys. Rev. Lett., 87 (2001) 228101-1.
24 D. Ly, Y.Z. Kan, B. Armitage and G.B. Schuster, J. Am. Chem. Soc., 118 (1996) 8747.
25 J. Jortner, Proc. Nat. Acad. Sci. USA, 95 (1998) 12759.
26 E. Conwell and S.V. Rakhmanova, Proc. Nat. Acad. Sci. USA, 97 (2000) 4556; S.V. Rakhmanova and E.M. Conwell, J. Phys. Chem. B, 105 (2001) 2056.
27 D. Ly, L. Sanii and G.B. Schuster, J. Am. Chem. Soc., 121 (1999) 9400.
28 S. Komineas, G. Kalosakas and A.R. Bishop, Phys. Rev. E, 65 (2002) 061905.
29 Z. Hermon, S. Caspi and E. Ben-Jacob, Europhys. Lett., 43 (1998) 482.
30 P. Carpena, P. Bernaola-Galvan, P.Ch. Ivanov and H.E. Stanley, Nature, 418 (2002) 955.
31 L. Cai, H. Tabata and T. Kawai, Appl. Phys. Lett., 77 (2000) 3105; K.-H. Yoo, D.H. Ha, J.-O. Lee, J.W. Park, J. Kim, J.J. Kim, H.-Y. Lee, T. Kawai and H.Y. Choi, Phys. Rev. Lett., 87 (2001) 198102; H.-Y. Lee, H. Tanaka, Y. Otsuka, K.-H. Yoo, J.-O Lee and T. Kawai, Appl. Phys. Lett., 80 (2002) 1670.
32 M. Peyrard and A.R. Bishop, Phys. Rev. Lett., 62 (1989) 2755.
33 L.V. Yakushevich, Quart. Rev. Biophys., 26 (1993) 201.
34 M. Barbi, S. Cocco and M. Peyrard, Phys. Lett. A, 253 (1999) 358.
35 L. Stryer, *Biochemistry*, Freeman, New York, 1995.
36 S. Cocco and R. Monasson, J. Chem. Phys., 112 (2000) 10017.
37 T.D. Holstein, Ann. Phys. NY, 8 (1959) 325, 343.
38 P.W. Anderson, Phys. Rev., 109 (1958) 1492.
39 N.F. Mott, Adv. Phys., 10 (1961) 107.
40 D. Hennig, J.F.R. Archilla and J. Agarwal, Physica D, 180 (2003) 256.
41 J.F.R. Archilla, D.Hennig and J.Agarwal, Charge transport in a nonlinear, three-dimensional DNA model with disorder in L. Vazquez, M.P. Zorzano, R.S. Mackay, (Eds.), Localization and energy transfer in nonlinear systems, World Scientific, Singapore, 2003, pp.153—160.
42 F. Palmero, J.F.R. Archilla, D. Hennig and F.R. Romero, N. J. Phys., 6 (2004) 13.1.
43 D. Hennig, Eur. Phys. J. B, 30 (2002) 211.
44 G. Kalosakas, S. Aubry and G.P. Tsironis, Phys. Rev. B, 58 (1998) 3094.
45 N.K. Voulgarakis and G.P. Tsironis, Phys. Rev. B, 63 (2001) 14302.
46 M. Peyrard and M.D. Kruskal, Physica D, 14 (1984) 88.
47 S. Flach and C.R. Willis, Phys. Rep., 295 (1998) 181.
48 D. Chen, S. Aubry and G.P. Tsironis, Phys. Rev. Lett., 77 (1996) 4776.
49 Y.-J. Ye and J. Ladik, Int. J. Quant. Chem., 52 (1994) 491.
50 D. Hennig, E.B. Starikov, J.F.R. Archilla and F. Palmero, J. Biol. Phys., 30 (2004) 227.
51 W. Saenger, *Principles of nucleic acid structure*, Springer Verlag, New York, 1984.
52 G. Brunaud, F. Castet, A. Fritsch, M. Kreissler and L. Ducasse, J. Phys. Chem. B, 105 (2001) 12665.
53 G. Cuniberti, L. Craco, D. Porath and C. Dekker, Phys. Rev. B, 65 (2002) 241314(R).
54 E.B. Starikov, Phys. Chem. Chem. Phys., 4 (2002) 4523.
55 J.J.P. Stewart, J. Comp. Chem., 10 (1989) 209.
56 A. Fortunelli and A. Painelli, Phys. Rev. B, 55 (1997) 16088.
57 E.S. Chen and E.C.M. Chen, Biochem. Biophys. Res. Comm., 289 (2001) 421.
58 M. Ibanes, J.M. Sancho and G.P. Tsironis, Phys. Rev. E, 65 (2002) 041902.
59 Y. Dakhnovskii, J. Chem. Phys., 100 (1994) 6492.
60 Y. Dakhnovskii, D.G. Evans, H.J. Kim and R.D Coalson, J. Chem. Phys., 103 (1995) 5459.
61 I.A. Goychuk, E.G. Petrov and V. May, Chem. Phys. Lett., 253 (1996) 428.
62 I.A. Goychuk, E.G. Petrov and V. May, J. Chem. Phys., 106 (1997) 4522.

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 23

Atomistic models of biological charge transfer

Thorsten Koslowski, Tobias Cramer and Nadine Utz

Institut für Physikalische Chemie, Universität Freiburg, Albertstrasse 23a, D-79104 Freiburg im Breisgau, Germany

Abstract

In this contribution, we highlight the potential of atomically resolved models to address the problem of ground-state charge transfer reactions in biological systems like (i) DNA or (ii) proteins that incorporate transition metal complexes as model cofactors. We describe an extended version of the Su–Schrieffer–Heeger Hamiltonian as an atomistic model and the procedures that enable its efficient numerical solution. As applications, we present the computation of charge transfer rates within DNA oligomers, complex DNA arrangements like bulges or three-way junctions, and in model proteins. The results are compared with experimental findings and phenomenological concepts of biological charge transfer.

23.1 INTRODUCTION

Charge transfer reactions within and between proteins and their cofactors play an important role in fundamental biological processes like photosynthesis and respiration. In addition, DNA charge transfer has become of considerable interest in the past decade [1,2]. This interest has been motivated by the attempt to understand the direction of mutations upon oxidative stress or UV irradiation via the migration of charges to preferential sites of trapping. Sensor applications and the prospect of having a nano-sized biopolymer wire at hand that can be easily replicated are tantalizing possibilities and have motivated many experimental and theoretical studies [3,4].

Charge transfer reactions can be described by Marcus' seminal theory [5], which considers the dynamics of an excess charge in a two-site donor–acceptor system coupled to a bath. The charge transfer process is interpreted with the help of a free energy surface, usually displayed and analysed along a generalized–yet frequently unspecified–reaction coordinate. On this energy surface and its excited-state counterpart, three energies which are characteristic for charge transfer reactions can be identified, they are illustrated with the help of Fig. 23.1. We have the total driving force of the reaction, ΔG, the electronic tunnel

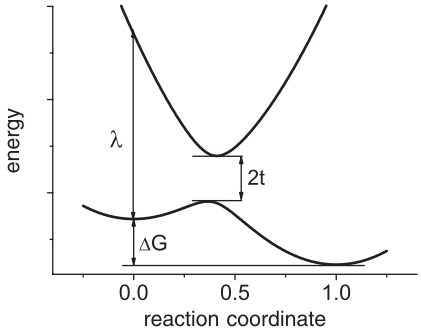


Fig. 23.1. Potential energy curve typical for charge transfer reactions. The characteristic energy parameters relevant to Marcus' theory are shown. For details, see text.

splitting t as a measure of the strength of the donor–acceptor coupling and the so-called reorganization energy λ. The latter is equal to the vertical optical transition of the system. It specifies a trapping potential in which an excess charge carrier may localize, provided λ is sufficiently strong. Both intramolecular electron–phonon coupling and the polarization of the dielectric environment of the molecule contribute to the reorganization energy, we have $\lambda = \lambda_{\text{in}} + \lambda_{\text{out}}$. Rather than specifying ΔG and λ, we may use the activation barriers of the forward and backward reactions. We wish to draw the readers' attention to the close relation of these concepts to small polaron theory and to the transformation that mediates between Marcus' treatment of polarization terms and the picture of Einstein phonons that exhibit a linear on-site electron–phonon coupling [6]. In the regime of a small tunnel splitting t, the classical charge transfer rate is given by

$$k_{\text{CT}} = \frac{t^2}{\hbar}\sqrt{\frac{\pi}{\lambda k_{\text{B}} T}}\exp(-E_A/k_{\text{B}}T) \tag{1}$$

In the strong-coupling case, pure Arrhenius behaviour with a prefactor of the order of 10^{12} to 10^{13} s^{-1} is assumed; at low temperatures, quantum corrections become important [7]. The scheme described above is not only relevant for direct donor–acceptor transfer, but also for configurations that involve a covalent linker or bridge. Here, the perturbation technique of electronic superexchange [8] or Löwdin's partitioning approach [9] can be applied. Numerous suggestions have been made to compute the above-listed energy parameters relevant to charge transfer within large and complex biological systems; in the following, we will only list the most important that focus on estimating the tunnel splitting t. In the widely used *pathways* approach of Beratan, Onuchic and Gray, t is summed from individual contributions through bonds, through hydrogen bonds and through space at a resolution of amino acids or comparable entities [10]. Dutton, Moser and co-workers, on the other hand, highlight the importance of through-bond charge transfer and provide an empirical rule, with t exponentially decreasing with an increasing donor– acceptor distance [11]. On a quantum mechanical level, path integral techniques [12] and Green's function methods [13] have made large systems treatable without explicit diagonalizations. High-level quantum chemical

calculations are limited to comparatively small systems, yet they provide an important frame of reference for parametrization.

Is there a serious need for an additional computational scheme addressing charge transfer? We believe that we are able to give a positive answer due to the properties of our approach that will be presented in the following section, and the list of applications given in the third part of this work. We will return to the question just formulated in the conclusions section.

23.2 METHODS

Whereas many biochemical processes and reactions can be adequately modelled on the basis of classical force fields applied within molecular dynamics or Monte Carlo simulations, the physical nature of the substrate of charge transfer reactions – an excess positive or negative charge with a centre of mass variable in space – strongly calls for a quantum mechanical description. Here, we apply a simple tight-binding Hamiltonian to describe the chemical bond within biological systems, which is extended by terms that induce charge localization. These can be considered as microscopic representations of the inner and outer sphere contributions to the reorganization energy, as encountered in the phenomenological theories of charge transfer referenced above.

Whenever a separation of σ and π orbitals is feasible – as for DNA nucleobases arranged within a polyglycophosphate scaffold – variants of the Su–Schrieffer– Heeger (SSH) model can be applied [14]. Its extension to systems with more than one valence orbital per atom is straightforward, but we will present the simpler single-orbital version here for the sake of convenience. The potential energy part of the SSH Hamiltonian incorporates a classical harmonic interaction between two atoms i and j and a distance-dependent tight-binding matrix element that gives rise to linear electron–phonon coupling:

$$\hat{V}_{\mathrm{SSH}} = \sum_{i<j} \frac{k_{ij}}{2} x_{ij}^2 - \sum_{ij} (t_{ij}^0 - \alpha_{ij} x_{ij})\, a_i^\dagger a_j \tag{2}$$

The deviation of the distance between a pair of atoms from that of a single bond is denoted by x_{ij}. The k_{ij} are the corresponding force constants, the t_{ij}^0 the tight-binding matrix elements and the α_{ij} the electron–phonon coupling constants. The $a_i^\dagger / a_j$ are the creation and annihilation operators acting on a basis of atomic orbitals, which is usually assumed to be orthogonal. We note that $a_i^\dagger a_i = n_i$, which is the number operator, and we define $n_{ij} = a_i^\dagger a_j$. The expectation value of the latter is equal to the bond order. The nuclear coordinates and momenta are treated as classical quantities.

Recently, we have extended the standard SSH model by the following elements in order to address the electronic structure of complex molecules of biochemical relevance [15]. To account for the heteroatomic character of biological systems, element-specific diagonal parameters t_{ii} have been introduced. Long-range coupling has been considered via additional distance-dependent tight-binding matrix elements $t_{ij}(r)$. For the computations relevant to DNA charge transfer, the parameters beyond those of the standard SSH model have been obtained by a careful fit to *ab initio* quantum chemical calculations and

to the experimental data, full details of this procedure and the parameters are given in [15]. Outer sphere degrees-of-freedom have been incorporated into the Hamiltonian by a non-retarded reaction field, which is based upon a straightforward extension of Marcus' treatment of two cavities that are dispersed in a dielectric medium [16]. The outer sphere reorganization energy can be written as:

$$\lambda_{\text{out}} = \frac{e^2}{4\pi\varepsilon_0}\left(\frac{1}{\varepsilon_\infty}-\frac{1}{\varepsilon_\text{s}}\right)\left(\sum_i \frac{\Delta z_i^2}{\sigma_i} - \sum_{i<j}\frac{\Delta z_i \Delta z_j}{r_{ij}}\right) \tag{3}$$

Here, ε_∞ and ε_s denote the high-frequency and the static limits of the dielectric response function, the σ_i are particle diameters and the r_{ij} the interparticle distances. Whenever fast and slow time scales can be separated, the charge differences Δz_i can be replaced by the corresponding number operators $n_i - \bar{n}_{i,0}$, and the model turns into the interaction part of the so-called attractive Hubbard Hamiltonian [17]. As in the tight-binding part of the SSH model, we consider the long-range Coulomb interactions as being absorbed into effective parameters.

To enable an efficient numerical mean-field treatment, the SSH Hamiltonian can be transformed by the introduction of a displaced phonon coordinate

$$q_{ij} = x_{ij} - \frac{\alpha_{ij}}{k_{ij}}\left(a_i^\dagger a_j + a_j^\dagger a_i\right) \tag{4}$$

to decouple nuclear and electronic degrees of freedom [18]. The resulting mean-field restricted Hartree–Fock expression that combines the polaron-transformed equation (2) and the reaction field (3) reads [15]:

$$\hat{H} = -\sum_{ij}\left(t_{ij}(r) - 4U_{ij}\bar{n}_{ij}\right) a_i^\dagger a_j - 2U_\text{d}\sum_i n_i(\bar{n}_i - \bar{n}_{i,0}) \tag{5}$$

We apply the usual corrections for counting the electron–electron interaction twice. The overbar denotes the expectation values taken within a self-consistent field (SCF) iterative calculation. The Hamiltonian (5) only depends on the tight-binding parameters t_{ij}, the diagonal Hubbard parameter U_d – as both given in Ref. [15] – and its off-diagonal counterpart U_{ij}, which equals $\alpha^2/2k = 0.32$ eV for π-bonded atoms in the original SSH parametrization [14] and is set to zero otherwise. For DNA charge transfer, these values provide an excellent fit for SSH model nucleobase π orbital energies and HOMO atomic orbital contributions to their *ab initio* density functional counterparts [15]. In our model, the U_d parameter denotes the local strengths of a reaction field that describes the interaction with all polarization degrees of freedom, and hence mainly represents outer-sphere contributions. The U_{ij} represent the locally transformed electron–phonon coupling and vibrational terms, i.e., the inner-sphere reorganization degrees of freedom. These contributions are treated on the same footing and are incorporated into the same electronic Hamiltonian. This fact is not only aesthetically pleasing, but now also enables an efficient solution of the emerging self-consistent field problem.

23.3 APPLICATIONS

23.3.1 DNA oligomers

As a first straightforward application directly related to photochemically induced DNA fragmentation, we have studied interguanine charge transfer through idealized DNA oligomers [15]. Here, in the presence of an excess positive charge, each guanine acts as a potential centre of attraction within a Hartree–Fock self-consistent field procedure. In contrast to the majority of LCAO-MO mean-field calculations as applied to small molecules, we find a multitude of local minima rather than a single basin of attraction of the SCF iteration. Each of these minima is characterized by a particular set of charge and bond orders.

Between two of these minima, we apply a linear interpolation scheme usually referred to as the linear synchronous transit (LST) approach, which to our knowledge was first applied by Clark and co-workers to inner sphere charge transfer between solvated organic molecules [19]. In doing so, we gradually turn the charges and bond orders corresponding to the first minimum into that of the second one, and the interpolation parameter serves as an obvious and convenient reaction coordinate. The thus obtained pictures resemble Fig. 23.1, from which the characteristic energies of charge transfer processes can be readily extracted. For double-strand oligomers with sequences of the type $A_3GA_nGA_3$, we have given estimates of the reaction coefficients as a function of the number n of adenine–thymine pairs that intervene two guanines. The corresponding half-lifes, $t_{1/2}$, are displayed in Fig. 23.2. They decrease exponentially with an increasing guanine–guanine distance with a characteristic decay constant of 1.38 $Å^{-1}$, which lies in the upper range of the experimental values [20–22]. This behaviour is typical for a tunnelling or – in a different terminology – superexchange mechanism. In contrast, it has been observed that for long AT sequences, k_{CT} is almost independent of n. This finding has been rationalized by the assumption of a second process with adenine nucleobases as stepping stones [23]. This

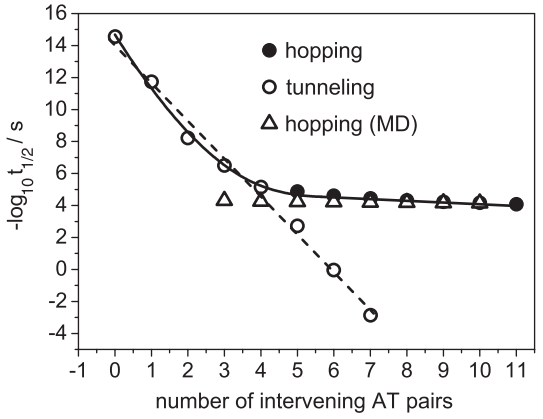


Fig. 23.2. Half-lifes computed for guanine–guanine charge transfer in the superexchange (or tunnelling) and hopping regimes.

mechanism is usually referred to as hopping or diffusive, as the rate equation for the populations p_i of charged species within a longer sequence of adenines,

$$\dot{p}_i = kp_{i-1} - 2kp_i + kp_{i+1} \tag{6}$$

resembles a spatially discretized one-dimensional diffusion equation. Here, k denotes the adenine–adenine CT coefficient. Using an A_6 bridge as an example, the first mechanism corresponds to tunnelling (or superexchange), $A_nG^+A_6GA_n \rightleftharpoons A_nGA_6G^+A_n$. If only the two central adenines were operative as stepping stones, the second sequence of processes could be formulated as $A_nG^+A_6GA_n \rightleftharpoons A_nGAAA^+AAAGA_n \rightleftharpoons A_nGAAAA^+AAGA_n \rightleftharpoons A_nGA_6G^+A_n$. In a system of rate equations with reaction coefficients extracted from all of the hopping processes, it is possible to compute the corresponding half-lifes [24]. We were rather surprised to find that the very model presented above also accounts for the second mechanism and permits the computation of the $t_{1/2}$ values [15], which are also presented in Fig. 23.2. As in the experiments, a crossover from a superexchange to a hopping regime can be observed at $n \simeq 3$–4 with $t_{1/2} \simeq 10^{-5}$s [25]. We note that the computation of the rates relevant to the second process is only possible if the potential energy surface relevant to adenine–adenine hopping exhibits at least one minimum. For idealized geometries, this is the case for $n > 4$. In a dynamically disordered system–as generated by the molecular dynamics simulations referenced below – smaller AT bridges can also carry an excess charge.

The question of the DNA dynamics upon the charge transfer kinetics has been addressed using snapshots from classical molecular dynamics simulations (see Fig. 23.3) as inputs of the electronic structure calculations [24]. In particular, the effective donor–acceptor couplings now exhibit a broad exponential distribution, yet the high efficiency of the computational scheme outlined above permits the proper calculation of averages via the consideration of several hundred configurations. Correlations in the effective tunnel splittings are short-lived (some 10 fs), thus permitting the calculation of thermodynamic averages, which are usually a factor of 10 smaller than the corresponding values computed for idealized structures.

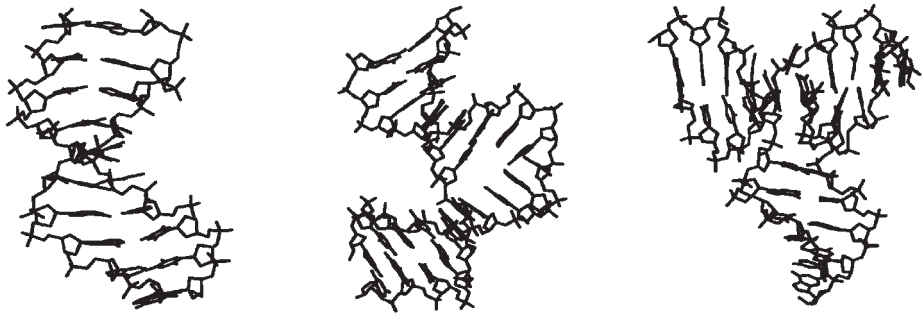


Fig. 23.3. Disordered and complex DNA structures: snapshot of an oligonucleotide molecular dynamics simulation, bulge and three-way junction (from left to right).

23.3.2 Complex arrangements of nucleobases

Considerable evidence exists that charge transfer is operative not only in carefully designed artificial DNA oligomers, but also in structures relevant to living organisms: Núñez *et al.* have been able to attach a rhodium intercalator to the 5′ end of a nucleosome core particle – a DNA–protein complex in which the majority of the eukaryotic genome is organized – and have induced DNA damage more than 80 Å away from the site of the initial oxidation process [26]. Further evidence suggests the possibility of *in vivo* charge transfer stems from long-range oxidative damage in whole nuclei [27]. As for simple DNA oligomers, the photochemically generated holes can become trapped at multiple guanine sites (G_n clusters with $n > 2$), where consecutive irreversible oxidation reactions occur and the DNA double strand can finally be split by chemical means, thus permitting a simple analysis of the relative length of migration. Using a sequence of 146 base pairs corresponding to an X-ray structure of the human satellite DNA bound to a histone – which is very close to the one used in the CT experiments – we were able to provide a microscopic rationalization of this behaviour [28].

We only note in passing that more complex arrangements of DNA, like three-way junctions, dendrimers and defect structures like bulges, have been the object of applying this scheme, and that detailed predictions about the reaction kinetics have been made in either case: bulges exhibit small sequences of unpaired bases located between two otherwise unperturbed double helices (cf. Fig. 23.3). Here, the charge transfer kinetics depends crucially on the value of a global geometric parameter, the so-called kink angle between the two double-strand parts of the oligomer. A maximum average k_{CT} of $10^6\ s^{-1}$ has been observed for a kink angle of 87°, indicating CT sufficiently fast to compete with DNA oxidation ($k_{ox} \simeq 10^{-4}\ s^{-1}$) [24]. In three-way junctions, the CT is highly anisotropic, reflecting the pronounced T-shape asymmetry of the junction (see Fig. 23.3) with predicted transfer rates of $10^{11}\ s^{-1}$ along the horizontal arm of the structure, which is slowed by a factor of 10^4 once CT to the vertical arm is considered [29].

23.3.3 Model proteins

In addition to computing charge transfer parameters for DNA, we have studied the kinetics of model proteins that are complexed by two $Ru[NH_3]_5^{2.5+}$ entities bound to two histidine residues [30]. For these calculations, only the chemical bond and the polarization contributions of a Hamiltonian related to Eq. (5) have been considered. The geometries of our models represent α-helices, isolated β-sheets and an antiparallel arrangement of β-sheets connected by a turn. The α-helices and the isolated β-sheets have the sequence C-His-Gly_n-His-N, the β-turn is based upon experimental geometries and has the sequence C-Gly_l-His-Gly_m-His-N with $l + m = 22$. To the E2 nitrogen of each histidine, the ruthenium complex was coordinated. For the aromatic histidine residues, we constructed a basis of $2p$ orbitals perpendicular to the C–N-plane. In this manner, only a single d-orbital couples to the π system, considerably reducing the degrees of freedom relevant to charge transfer, although a full consideration of all d orbitals is straightforward. The tight-binding approach presented here permits the computation of the energy parameters

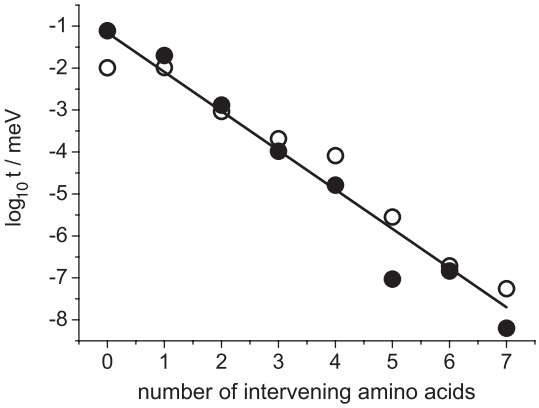


Fig. 23.4. Decadic logarithm of the effective through-bond tunnel splitting t as a function of the number of glycines intervening two ruthenium-complexed histidines for α-helices (○) and β-sheets (●).

relevant to charge transfer both in the absence and in the presence of long-range through-space interactions or hydrogen bonds, orientation effects are automatically taken into account. The tunnel splittings computed for charge transfer solely mediated by through-bond interactions show an exponential dependence on the number of bonds separating the donor and the acceptor, as evident from Fig. 23.4. We obtain an average reduction of t by a factor of 0.5 per bond, which supports the *pathways* scheme of Beratan, Onuchic and Gray [10], which uses a reduction factor of 0.4 to 0.6. For all but the smallest donor–acceptor separations, through-space CT is faster within α-helices than through-bond transfer alone. For β-sheets, through-space interactions do not speed up the electron transfer due the stretched character of the protein and the resulting large donor–acceptor separations. The dependence of t upon the distance is exponential, with an only slightly larger coefficient of 0.95 Å^{-1} than that given by Dutton and Moser for CT through a protein medium (0.7 Å^{-1}) [11]. For through-space charge transfer, orientational effects are important, they can amount to one order of magnitude in t and two orders in k_{CT}. Through-hydrogen bond charge transfer dominates for the β-turn for chromophores far apart in the sequence of amino acids.

23.4 CONCLUSIONS

In this contribution, we have introduced a simple, yet chemically specific model suited to the computation of ground-state charge transfer potential energy surfaces. It bridges high-level *ab initio* calculations usually restricted to small systems and models related to the solid-state sciences, which are ususally coarse-grained on a level of entire nucleobases or amino acids. The methodology suggested here is atomistic, with a strong appeal to structural biochemists and biologists; it shows an intimate relation to standard concepts of the chemical bond within the highly flexible linear combination of the atomic orbitals approach, solved on the level of Hartree–Fock mean field theory; it provides full variational control rather than operating in a perturbative weak-coupling regime; no *ad hoc* assumptions concerning an artificial donor–bridge–acceptor partitioning or the introduction of a transfer energy are required; and we have an immediate grasp of the

reaction coordinate. We hope to be able to address electronic structure and CT problems for even more complex systems like biomolecule–nanoobject contacts, charge migration from biopolymers to surfaces or within aggregates of conducting polymers and the molecules of life.

23.5 ACKNOWLEDGEMENTS

It is a pleasure to thank the present and former members of our group that have worked on biological charge transfer: L. Engel, G. Ganzenmüller, S. Krapf, A. Staehle, C. Schäfer, G. Rink, C. Wittekindt and F. Zehender. We wish to acknowledge helpful comments by J. Armbruster, A. Blumen, E. Conwell, Th. Friedrich, P. Gräber, V. Helms, R. Naaman, E. Neumann, S. Larsson, W. von Niessen, E. Starikov and M. Zwolak. This work has been supported by the Deutsche Forschungsgemeinschaft (DFG) via an SFB 428 project grant.

23.6 REFERENCES

1 B. Schuster (Ed.), Long-range charge transfer in DNA, Springer, Berlin, 2004.
2 R.G. Endres, D.L. Cox and R.R.P. Singh, Rev. Mod. Phys., 76 (2004) 195.
3 A.H. Flood, J.F. Stoddart, D.W. Steuerman and J.R. Heath, Science, 306 (2004) 2055.
4 M. Di Ventra and M. Zwolak, DNA electronics in H. Singh-Nalwa (Ed.), Encyclopedia of Nanoscience and Nanotechnology, American Scientific Publishers, New York, 2004, p. 475.
5 R.A. Marcus, J. Chem. Phys., 24 (1956) 966.
6 D.C. Licciardello, D.L. Stein and F.D.M. Haldane, Phil. Mag. B, 43 (1981) 189.
7 M. Bixon and J. Jortner, Faraday Discuss. Chem. Soc., 74 (1982) 7.
8 H.M. McConnell, J. Chem. Phys., 35 (1961) 508.
9 S. Larsson, J. Am. Chem. Soc., 103 (1981) 4034.
10 J.N. Onuchic, D.N. Beratan, J.R. Winkler and H.B. Gray, Ann. Rev. Biophys. Biomol. Struct., 21 (1992) 349.
11 C.C. Moser, C.C. Page, X. Chen and P.L. Dutton, J. Biol. Inorg. Chem., 2 (1997) 393.
12 P.G. Wolynes and A. Kuki, Science, 236 (1987) 1647.
13 A.A. Stuchebrukhov and R.A. Marcus, J. Phys. Chem., 99 (1995) 7581
14 W.P. Su, J.R. Schrieffer and A. Heeger, Phys. Rev. Lett., 41 (1979) 1698.
15 T. Cramer, S. Krapf and T. Koslowski, J. Phys. Chem. B, 108 (2004) 11812.
16 R.A. Marcus and N. Sutin, Biochim. Biophys. Acta, 811 (1985) 265.
17 R. Micnas, J. Ranninger and S. Robaszkiewicz, Rev. Mod. Phys., 62 (1990) 113.
18 M. Rateitzak and T. Koslowski, Chem. Phys. Lett., 377 (2003) 455.
19 G. Rauhut and T. Clark, J. Am. Chem. Soc., 115 (1993) 9127.
20 E. Meggers, M.E. Michel-Beyerle and B. Giese, J. Am. Chem. Soc., 120 (1998) 12950.
21 C. Wan, T. Fiebig, O. Schiemann, J.K. Barton and A.H. Zewail, Proc. Natl. Acad. Sci. USA, 97 (2000) 14052.
22 F.D. Lewis, R.L. Letsinger and M.R. Wasielewski, Acc. Chem. Res., 34 (2001) 159.
23 B. Giese, J. Amaudrut, A.-K. Köhler, M. Spormann and S. Wessely, Nature, 412 (2001) 318.
24 T. Cramer, T. Steinbrecher, A. Labahn and T. Koslowski, PCCP, 7 (2005) 4039.
25 B. Giese and M. Spichty, Chem. Phys. Chem., 1 (2000) 195.
26 M.E. Núñez, K.T. Noyes and J.K. Barton, Chem. Biol., 9 (2002) 403.
27 M.E. Núñez, G.P. Holmquist and J.K. Barton, Biochemistry, 40 (2001) 12465.
28 T. Cramer, S. Krapf and T. Koslowski, Phys. Chem. Chem. Phys., 6 (2004) 3160.
29 T. Cramer, A. Volta, A. Blumen and T. Koslowski, J. Phys. Chem. B, 108 (2004) 16586.
30 N. Utz, L. Engel, T. Friedrich and T. Koslowski, Zeitschrift für physikalische chemie, 219 (2005) 1391.

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 24

Nonlinear models in DNA conductivity

Victor D. Lakhno

Institute of Mathematical Problems of Biology, Russian Academy of Sciences, Pushchino, Moscow Region, 142290, Russian Federation

Abstract

This chapter considers the fate of a charge somehow injected into homogeneous and regular double-helical DNA chains. The charge motion is described in terms of quantum mechanics, whereas vibrational degrees-of-freedom in DNA are treated both classically and quantum mechanically. A typical charge transfer/transport pattern can physically be viewed as a polaron and/or soliton. A closed analytical expression for charge carrier velocity dependence on electric field in DNA samples has been derived and analyzed in detail.

24.1 INTRODUCTION

In recent times, the problem of charge transfer in the DNA double helix has attracted much attention [1–7]. Since DNA molecules in an equilibrium state have no free charge carriers, one can create electrons and holes (cation radicals) by means of photo-excitation or as a result of chemical reactions [3,8–11]. Free charge carriers in DNA can also be created by doping it, e.g., with metal atoms (M-DNA), i.e., by a procedure similar to that used to create semiconducting alloys [12].

The hypothesis that DNA might have conductive properties was first expressed in [13], soon after Watson and Crick discovered the DNA double-helical structure [14]. As was mentioned in [12], the quasi-one-dimensional DNA polymer that consists of paired pyrimidine and purine bases is similar to the famous stacking structures of low-dimensional organic conductors. A problem that has been discussed for a long time is whether DNA with its aromatic base pairs forming the stack can provide a profile of potential energy that allows a charged particle to move along it for long distances.

Very recently, charge transfer along DNA oligomers started to be observed in direct experimental studies, which was made possible, on the one hand, due to the development of special experimental equipment capable of capturing physico-chemical processes on the nano- and femto-second time scale, and, on the other hand, owing to the advent of sophisticated, high-precision biochemical techniques to produce complexes between

specific nucleotide pairs within DNA oligomers and different organic molecules capable of playing the role of electron donors and acceptors [1–12].

As a result of a series of experiments carried out during the last decade, DNA molecules have been suggested to be conductors [3,5–9]. Given the DNA molecule's ability to self-reproduce and the formation of complex structures, the discovery of its conductive properties has made it a very attractive object, from the point of view of its application in nano-electronics [1]. Of special interest for nano-technology are experiments on charge transfer/transport through synthetic double-helical polynucleotide chains with uniform base sequences, since such chains could act as molecular wires [15].

In this study, we will focus on DNA oligonucleotides with uniform base sequences. Properties of electrons and holes in such oligonucleotides ought to be analogous to those of band charge carriers in solid-state crystals.

The structure of this chapter is as follows. In Section 24.2, the physical model of DNA charge transfer/transport is proposed and its physical foundations are discussed, with special reference to the interaction of a charged particle in a uniform DNA oligomer with vibrational modes of its nucleotides.

In Section 24.3, an energy spectrum of a particle in a rigid nondeformed polynucleotide chain is discussed. It is shown that, in a uniform nucleotide double-helical chain consisting of guanine–cytosine pairs only and a charge a particle moving along the $(G)_n$ strand, there is one band of the particle states permitted energetically. In a regular double-helical chain of guanine–cytosine and thymine–adenine base pairs of $\binom{GT}{CA}_n$ -type with charge transfer occurring along the $(GT)_n$ strand, there are two permitted bands. In the general case, in a regular nucleotide sequence of $(GT_1 \ldots T_{k-1})_n$-type, there are k different energy bands.

In Section 24.4, a Hamiltonian of a polynucleotide chain is considered in continual approximation. It is shown that, in this case, the initial Hamiltonian can be represented in the form of a Fröhlich Hamiltonian that describes interaction between a quantum particle and a phonon field.

In Section 24.5, a case of weak interaction between a particle and the phonons is considered. The energy renormalization of a particle Hamiltonian due to interaction with phonons, as well as the damping of the particle quantum states, are considered.

In Section 24.6, the motion of a charged particle in a uniform polynucleotide chain in an applied electric field is studied. It is shown that there exists a minimum velocity for the particle uniform motion. Thus, Ohm's law breaks down for the case of a uniform polynucleotide chain. The current–voltage plot for such a chain reveals negative differential conductivity.

In Section 24.7, a case of strong interaction between a particle and the chain oscillations is considered. Through a Bogolubov canonical transformation, equations for uniform motion of a particle in a phonon field were found. An expression for the energy of the localized state is derived, which, in this case, represents a soliton or strong-coupling large-radius polaron.

In Section 24.8, the motion of a strongly coupled polaron with large radius in an electric field is studied. Unlike the case of weak coupling, a polaron can move with an arbitrary velocity, moreover there are two branches. The branch that is responsible for lower velocities dominates the current–voltage plot and corresponds to a positive differential

conductivity. The branch that is responsible for higher velocities corresponds to negative differential conductivity. It is also shown that, in the case of strong coupling, there exists some maximum value of electric field at which a uniform motion of polaron is possible.

In Section 24.9, we discuss how to calculate charge mobility in a nucleotide chain at finite temperatures using the Kubo formula. The temperature dependence of hole mobility in a uniform nucleotide sequence is studied.

In Section 24.10, Bloch oscillations of charge in a nucleotide chain in an electrical field are considered. Such oscillations emerge at $T = 0$, and collapse at finite temperatures.

24.2 QUANTUM MECHANICAL MODEL

In the model used below, DNA is considered to be a chain with N sites. Every site represents a nucleotide pair (Fig. 24.1), which is viewed as a harmonic oscillator.

Charge transfer through such a chain is described by the following Hamiltonian [16–18]:

$$H_p = \sum_i \alpha_i^0 |i\rangle\langle i| + \sum_{i,j} v_{ij} |i\rangle\langle j| + \sum_i \alpha_i' q_i |i\rangle\langle i| \tag{1}$$

where: $\alpha_i = \alpha_i^0 + \alpha_i' q_i$ stands for the energy value of a charged particle on i-th site; α_i^0-is the energy of particle on i-th site in rigid chain; $-N/2 \leqslant i, j \leqslant N/2$; v_{ij} are matrix elements describing the particle hopping between i and j-sites; and α_i' is the force constant for the displacement of the normal coordinate q_i in ith site.

In the charge transfer experiments, the charge carrying particles are represented by negatively charged electrons or positively charged holes.

Expression (1) represents a tight-binding Hamiltonian for a particle localized on the i-site and found in the state $|i\rangle$.

In absence of dynamic fluctuations, when $\alpha_i' = 0$, the Hamiltonian (1) describes coherent motion of the charged particle. In the general case, one cannot neglect the on-site motions, and, in our present model this is taken into account by assuming that the site energy α_i is a linear function of the displacement q_i.

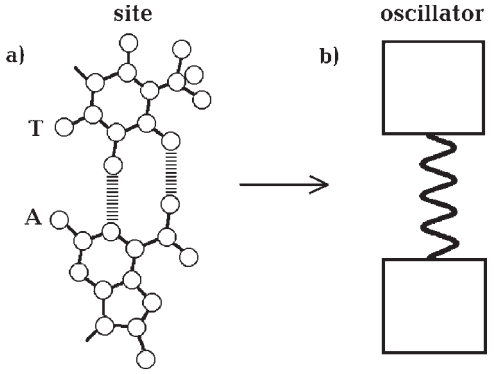


Fig. 24.1. Pairs of Watson–Crick oscillators: (a) nucleotide pair, (b) equivalent oscillator.

In our model, the Hamiltonian H_{osc} that describes DNA oscillations represents the sum of contributions from all the nucleotide pairs each containing one independent harmonic oscillator with the effective mass M_i, and frequency ω_i (Fig. 24.1):

$$H_{osc} = \sum_i \frac{p_i^2}{2M_i} + \sum_i K_i \frac{q_i^2}{2} \tag{2}$$

In the expression (2), p_i and q_i represent the impulse and the displacement of the ith normal coordinate; K_i is the elastic constant of the ith oscillator, and $\omega_i = (K_i / M_i)^{1/2}$.

In the Hamiltonian (1), only the diagonal matrix elements are dependent on q_i. In the general case, the non-diagonal matrix elements should also depend on q_i, and the latter dependences can pay a contribution of the same order as that from the diagonal elements. Moreover, there are examples where the contribution of site shifts into the nondiagonal part of the Hamiltonian is indeed the main one. Such examples include polyacetylene and polydiacetylene, where the strong dependence of the nondiagonal matrix elements v_{ij} on the shifts is due to the presence of alternating chemical bond lengths in these polymers.

In our simplified DNA model, we consider planes of the nucleotide bases as parallel to each other in every moment of time, and distances between the planes of the neighboring site bases as constant (Fig. 24.2). Then, charge transfer/transport in DNA is determined only by the degree of overlapping between the wavefunctions of the neighboring sites. As Fig. 24.2 shows, if the charge transfer occurs along one strand, in view of the rigid geometry, hopping integrals are virtually independent of on-site shifts.

Dependence of the nondiagonal matrix elements on the shifts may arise from other degrees-of-freedom, e.g., from DNA longitudinal deformations. However, the main contribution to charge transfer in DNA is most likely to be provided by the transversal degrees-of-freedom originating from the Watson–Crick hydrogen bonds. In particular, the latter assumption is fully supported by the important role of proton transfer along the Watson–Crick hydrogen bonds, which accompanies a hole transfer along the nucleotide sequence [19,20].

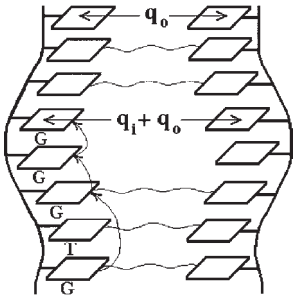


Fig. 24.2. Hole transfer along a DNA chain.

24.3 BAND STRUCTURE OF POLYNUCLEOTIDE CHAINS.

Let us consider a case, where α_i and v_{ij} are independent of the on-site deformations, i.e., a rigid chain. For the uniform chain we can assume $\alpha_i^0 = 0$. Let us also assume that the matrix elements differ from zero only between the nearest neighbors $v_{i,i\pm1} = v$, and the chain contains an even number of sites N. In this section let us assume $\hbar = 1$.

Let us cast the wave function $|\psi\rangle$ of the charged particle in the chain in the following form:

$$|\psi\rangle = \sum_n b_n |n\rangle \tag{3}$$

Under the above assumptions, the Schrödinger equation for the amplitudes b_n has the following form:

$$i\dot{b}_n = v(b_{n+1} + b_{n-1}) \tag{4}$$

The solution of Eq. (4) results in the single-band representation with stationary values of amplitudes $b_n(t)$; and energies of particle W_k:

$$\begin{aligned} &b_n(t) = \mathrm{e}^{-iW_k t} R_{n,k}, \qquad R_{n,k} = \frac{\mathrm{e}^{ikn}}{\sqrt{N+1}}, \\ &W_k = 2v\cos k, \qquad k = \pm\frac{2\pi l}{N}, \qquad l = 0,1,2,...,\frac{N}{2} \end{aligned} \tag{5}$$

The width of the conduction band, determined by (5), is equal to $4v$, and the basic state that corresponds to the conduction band bottom, is responded by energy $(-2v)$.

Let us consider, as an example of a regular strand, a $(GT)_n$-type nucleotide sequence, where G stands for guanine–cytosine pair and T is for thymine–adenine pair. In the eigenvalue equation for the $R_{n,k}$ -function, let us assume that G-values are even, and T-values are odd:

$$\begin{aligned} &n = 2j\text{: } \alpha_n = \alpha_{2j} = 0, \qquad v_{n,n+1} = v_{2j,2j+1} = v_1 \\ &v_{n,n+1} = v_{2j,2j-1} = v_2 \qquad j = 0,\pm1,\pm2,... \\ &n = 2j+1 : \alpha_n = \alpha_{2j+1} = \alpha_T, \qquad v_{2j+1,2j} = v_1 \quad v_{2j+1,2j+2} = v_2 \end{aligned} \tag{6}$$

After substituting (5) and (6) into (4), we obtain the following set of equations for the eigenvalues W:

$$\begin{aligned} W_k R_{2j,k} &= v_1 R_{2j+1,k} + v_2 R_{2j-1,k} \\ W_k R_{2j+1,k} &= \alpha_T R_{2j+1,k} + v_2 R_{2j+2,k} + v_1 R_{2j,k} \end{aligned} \tag{7}$$

If we express its solutions as $R_{2j+1,k} = u_2 \exp[ik(2j + 1)]$ and $R_{2j,k} = u_1 \exp[ik2j]$, we arrive at the following expression for the eigenvalues W_k:

$$W_k^{\pm} = \frac{\alpha_T}{2} + \sqrt{\left(\frac{\alpha_T}{2}\right)^2 + \left(v_1^2 + v_2^2 + 2v_1 v_2 \cos 2k\right)}$$
$$k = \pm \frac{2\pi m}{N}, \qquad m = 0,1,2,..., \frac{N}{4} \tag{8}$$

From (8) it follows that, in regular strands composed out of sites of two different types, there are two bands. Both bands have the same width ΔW equal to:

$$\Delta W = \sqrt{\left(\frac{\alpha_T}{2}\right)^2 + (v_1 + v_2)^2} - \sqrt{\left(\frac{\alpha_T}{2}\right)^2 + (v_1 - v_2)^2} \tag{9}$$

Figure 24.3 shows the band structures of strands of $(G)_n$, $(GT)_n$, $(GTT)_n$, and $(GTTT)_n$-types, depending on the potential difference between the α_Ts of thymine–adenine and guanine–cytosine. Thus, one can see that for regular chains, that contain m sites in an elementary cell, there are m different branches determining their band structure.

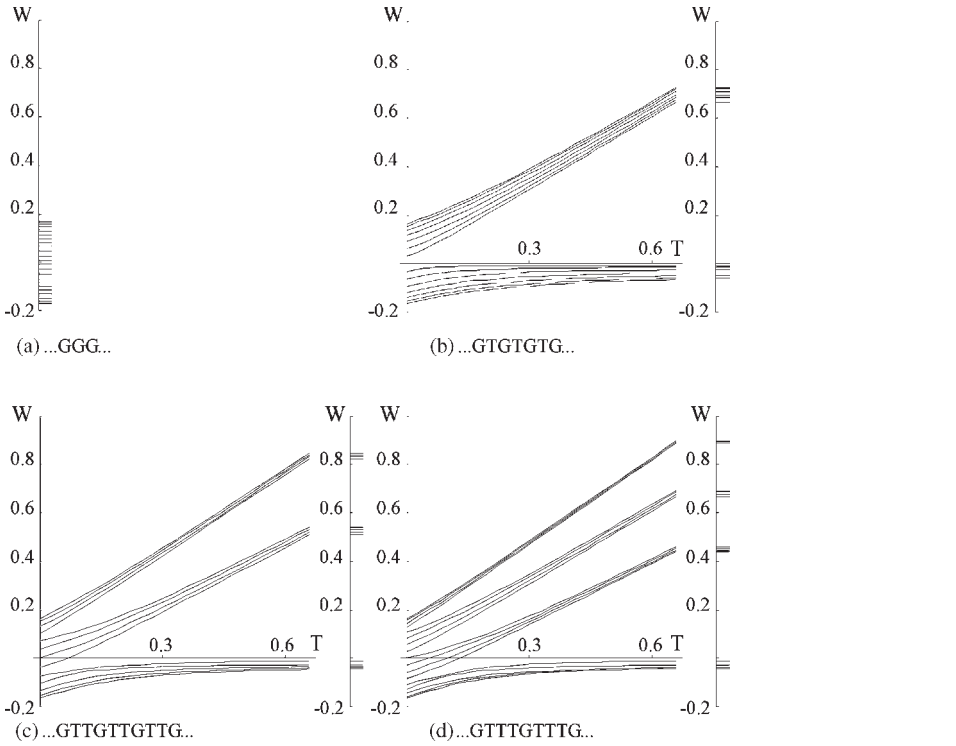


Fig. 24.3. Band structures for nucleotide pair sequences: (a) GGG…strand, (b) GTGTGT…strand, (c) GTTGTTGTT…strand, (d) GTTTGTTT…strand; for W (eV); $v_1 = 0.137$ eV; $v_2 = 0.085$ eV.

24.4 FRÖHLICH HAMILTONIAN IN HOLSTEIN MODEL

According to the band theory discussed in the previous section, a charge placed into a regular nondeformed rigid chain being in a state with a quasi-momentum k will rest in this state for all time. Hence, a rigid chain is possessed of infinite conductivity.

In a deformable chain, to maintain a constant value of the particle speed, generally speaking, an external electric field is necessary, since in the process of motion, the particle will loose energy by giving it out to vibrational degrees-of-freedom. In order to describe this process, let us consider the Hamiltonian (1), which, in the case of an uniform chain, is a Holstein Hamiltonian [21,22]. According to (1), a particle interaction with site deformations has the following form:

$$H_{\text{int}} = \alpha' q_n \tag{10}$$

By introducing dimensionless variables Q_n instead of q_n, using a relation for $q_n = Q_n\sqrt{\hbar/M\omega}$, as well as operators of creation and annihilation a_n^+, a_n:

$$a_n^+ = \frac{1}{\sqrt{2}}\left(-\frac{\mathrm{d}}{\mathrm{d}Q_n} + Q_n\right), \quad a_n = \left(\frac{\mathrm{d}}{\mathrm{d}Q_n} + Q_n\right) \tag{11}$$

we will have the following expressions for the phonon operator H_q and interaction operator H_{int}:

$$H_{\text{int}} = \alpha'\sqrt{\frac{\hbar}{2M\omega}}\left(a_n^+ + a_n\right), \quad H_q = \sum_{n=1}\hbar\omega a_n^+ a_n \tag{12}$$

By assuming that

$$a_n = \frac{1}{\sqrt{N}}\sum_k \beta_k \mathrm{e}^{ikna} \tag{13}$$

(Here, $k = \pm 2\pi l/Na$, $l = 0,1,2,\ldots$, a is distance between nearest sites), let us pass to a continuum limit by introducing a coordinate $x = na$. By assigning:

$$\overline{\alpha} = \alpha'^2\hbar/2M\omega \tag{14}$$

we will finally have the following standard Fröhlich Hamiltonian for Hamiltonian H, determined by (1):

$$\hat{H} = -\frac{\hbar^2}{2\mu}\Delta_x + \frac{\overline{\alpha}^{1/2}}{\sqrt{N}}\sum_k\left(\beta_k \mathrm{e}^{ikx} + c.c.\right) + \sum_k \hbar\omega\beta_k^+\beta_k \tag{15}$$

where

$$\mu = -\hbar^2/2va^2 \tag{16}$$

In the Hamiltonian (15), its first part corresponds to kinetic energy of a particle with effective mass μ; its second part stands for the energy of interaction between the particle and the chain oscillations; and its last part describes the phonon energy of the whole chain; $\Delta_x = d^2/dx^2$ is a Laplacian operator.

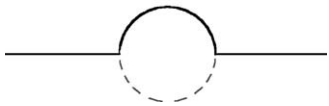
Let us note, that the continuum approximation is accurate in the limit $a \to 0$. It follows from (16), that the particle effective mass would still have a finite value if the value of the product va^2 remained finite. In other words, the passage to the continuum limit means that we assume the conductivity band, cast in (5), to be infinitely broad.

Let us now, first of all, consider the limiting case of weak coupling, when the coupling constant $\overline{\alpha}$ is small, and then another limiting case of strong coupling when the coupling constant $\overline{\alpha}$ is large.

24.5 WEAK COUPLING

If the values of $\overline{\alpha}^{1/2}$ are small compared with $\hbar\omega$, the Hamiltonian (15) describes processes of virtual emission and absorption of quanta $\hbar\omega$, during the charged particle motion along a chain.

Then the ground-state energy of a particle E_k is described by a diagram:



(17)

which corresponds to the following expression:

$$E(k) = \frac{\hbar^2 k^2}{2\mu} + \overline{\alpha} \int G(k - q)\, D(q)\, \frac{dq}{2\pi} \tag{18}$$

Making use of electronic G and phonon D Green functions, namely, their unperturbed expressions G^0 and D^0, to determine the particle energy, we have:

$$E(k) = \varepsilon(k) + \frac{1}{N} \sum_q \frac{\overline{\alpha}}{\varepsilon_k - \varepsilon_{k-q} - \hbar\omega - is} \tag{19}$$

where $\varepsilon_k = k^2/2\mu$ is the energy of the free particle, $s \to 0$. Expression (19) corresponds to the second order of perturbation theory.

If $k = 0$, one considers the case of resting particle, where the shift of the ground state energy ΔE due to creation and emission of phonons is equal to (cf. Eq. (19)):

$$\Delta E = -\overline{\alpha}\, a \sqrt{\mu/2\hbar^3\omega} = -\alpha'^2 \sqrt{\mu a^2/8M^2\hbar\omega^3} \tag{20}$$

In accordance with this, the total frequency of the collisions between the particle and phonons Γ_k, i.e., the relaxation rate of the particle in state k, is equal to:

$$\Gamma_k = 2\mathrm{Im}\, E(k) \tag{21}$$

and, taking into account Eq. (19), has the following form:

$$\Gamma_k = 2\pi \frac{\bar{\alpha}}{N} \sum_q \delta(\varepsilon_k - \varepsilon_{k-q} - \hbar\omega) \tag{22}$$

24.6 PARTICLE MOTION IN AN ELECTRIC FIELD IN A WEAK COUPLING LIMIT

Expression (22) for the frequency of collisions between a particle and the chain oscillations allows us to analyze its motion in an external electric field. Let us consider that, owing to the acceleration in the electric field ξ, the particle should emit phonons, so that the particle energy loss rate is equal to $\hbar\omega\Gamma_k$. Then, the particle energy win rate due to the electric field, if the speed of the particle is $v = \hbar k/\mu$, will be equal to $e\xi v$. Hence, the equilibrium speed of the particle in the electric field is determined by the following equation:

$$\hbar\omega\Gamma_k = e\xi v \tag{23}$$

From (22) it follows that the probability of phonon emission by a particle (Γ) will be nonzero, provided $v \geqslant \sqrt{2\hbar\omega/\mu} = v_{\min}$. Hence, the charge transfer/transport in an ideal DNA chain could be considered as follows. Under exertion of arbitrarily low electric fields, the charge will be moving with acceleration until its speed reaches the value of $v = v_{\min}$. After this, the equilibrium value of the particle velocity is reached as determined by the electric field strength ξ. At $\xi \to 0$, the equilibrium speed tends to infinity, as can be readily seen from Eq. (23). Therefore, at zero temperature, the conductivity of a deformable DNA chain must be infinite. It is possible that such a 'superconductivity' has been observed in DNA experimentally [23].

When passing, in (22), from summation by q to integration, after integrating we have:

$$\Gamma = \frac{\alpha'^2}{\pi\hbar\omega Ma\sqrt{v^2 - 2\hbar\omega/\mu}} \tag{24}$$

Using (23) and (24), the equilibrium speed of the particle will read:

$$v = \sqrt{\hbar\omega/\mu + \sqrt{(\hbar\omega/\mu)^2 + \zeta^2}} \tag{25}$$

$$\zeta = \frac{\hbar\omega}{\mu}(\xi_0/\xi) \qquad \xi_0 = \frac{\mu a}{\pi\hbar\omega}\frac{\alpha'^2}{Me} \tag{26}$$

From (25) and (26) it follows that the speed v decreases, as the electric field strength ξ increases. Since the drift speed v is proportional to the electric current, and the electric field is in turn proportional to the applied voltage, the negative differential resistance region will show up on the current–voltage plots of ideal polynucleotide chains, as determined by (25).

The above results show that DNA specimens could in principle be used to create generators of electric oscillations, as with conventional semiconductors (Section 24.10).

Meanwhile, the above picture of DNA charge transfer in the weak coupling limit assumes an infinitely broad conductivity band, whereas the width of this band should be finite, according to the results of Section 24.3,. Moreover, according to the examples shown in Fig. 24.3, the DNA conductivity band ought to be much narrower than in the conventional wide-band semiconductors. In this case, even before the speed is reached, stark localization of the charged particle in the electric field can occur, which ought to be accompanied by qualitatively new physical phenomena (Section 24.9 and 24.10). Another possible restriction is that the condition of weak coupling may be broken as the velocity of the charged particle increases. Then, another limiting case should be considered, namely strong interaction between charge and chain oscillations.

24.7 CASE OF STRONG COUPLING

In the above case of weak coupling $\overline{\alpha} \to 0$, the number of phonons surrounding the charged particle approaches zero. Therefore, in electric fields, the particle motion is unimpeded, until the condition for Cherenkov emission, $v > v_{\min}$ is fulfilled, and the charged particle can thereafter be transferred with the equilibrium velocity.

In the case of strong coupling, when $\overline{\alpha} \to \infty$, the number of phonons that surround the particle tends to infinity, even when $v = 0$. This is why, however low the velocity of the particle might be, in the limit of strong coupling its motion will always be accompanied by phonon emission.

The method to consistently solve the quantum problem (15) for the strong coupling case $\overline{\alpha}^{1/2} >> \hbar\omega$ was described in [24–26], where the adiabatic approximation and the translation degeneracy of Hamiltonian (15) were taken into account. Using this recipe to solve (15), one might first perform the transition to complex field coordinates:

$$q_k = \varepsilon\,(\beta_k + \beta^+_{-k})/\sqrt{2} \qquad -i\frac{\partial}{\partial q_k} = i(\beta^+_k - \beta_{-k})/\varepsilon\sqrt{2} \tag{27}$$

where ε is a small value, and the Hamiltonian (15) may then be recast in the following form:

$$H = -\frac{\hbar^2}{2\mu}\,\Delta_x + \sqrt{2}\,\frac{\tilde{\alpha}^{1/2}}{\sqrt{N}}\sum_k q_k e^{ikx} + \frac{1}{2}\sum_k \overline{v}\; q_{-k}\, q_k - \frac{\varepsilon^4}{2}\sum_k \overline{v}\;\frac{\partial}{\partial q_{-k}}\;\frac{\partial}{\partial q_k} \tag{28}$$

where $\hbar\omega = \varepsilon^2\overline{v}$ and the last part of (28) corresponds to the kinetic energy of phonons, which tends to zero in the limit of zero oscillator frequency, i.e., at $\varepsilon \to 0$, $\overline{\alpha} = \varepsilon^2\tilde{\alpha}$, $\tilde{\alpha}^{1/2} = \sqrt{1/2M\overline{v}}\;\alpha'/\varepsilon$.

In consistently applying the conventional perturbation theory for the Hamiltonian (28) with respect to the ε-parameter, we arrive at the corresponding wave equations in the following form:

$$(H - E)\,\Psi = 0 \quad (H_0 + \varepsilon H_1 + \varepsilon^2 H_2 + \ldots - E)\,\phi = 0 \tag{29}$$

$$\Psi = \mathrm{e}^{ipq}\,\phi\,(\lambda, Q_k), \qquad \hat{p} = -i\hbar\,\frac{\partial}{\partial q} \tag{30}$$

where q is the x-part that relates to the uniform translational motion; λ is the fluctuating x-part; Q_k are the coordinates of the phonon fields. At the zero approximation:

$$\phi = \varphi\,(\lambda)\Theta\left(\ldots Q_k \ldots\right) \tag{31}$$

$\varphi\,(\lambda)$ should obey the following equation [24]:

$$\left(-\frac{\hbar^2}{2\mu}\,\Delta_\lambda + U(\lambda) - W_0\right)\varphi\,(\lambda) = 0$$

$$U(\lambda) = -\,\frac{2\tilde{\alpha}}{N}\sum_k \frac{\bar{v}}{\bar{v}^2 - (\hbar k c)^2}\int \mathrm{e}^{-ik\lambda'}\,|\varphi\,(\lambda')|^2 \mathrm{d}\lambda' \mathrm{e}^{ik\lambda} \tag{32}$$

where $c = v/\varepsilon$, v is the mean velocity of the progressive motion of the charged particle and its phonon 'cloud'. The total energy E in this case reads:

$$E = W_0 + \frac{1}{2}\sum_k |u_k|^2 \left(\bar{v} + \frac{(\hbar k c)^2}{\bar{v}}\right) \tag{33}$$

$$u_k = -\frac{1}{\sqrt{N}}\;\frac{\sqrt{2}\tilde{\alpha}^{1/2}\bar{v}}{\bar{v}^2 - (\hbar k c)^2}\int \mathrm{e}^{-ik\lambda}\,|\varphi\,(\lambda)|^2 \mathrm{d}\lambda \tag{34}$$

Expressions (33) and (34) actually correspond to the results of the polaron theory in the strong coupling limit.

When $v = 0$, i.e. in the case of the resting polaron, the following Schrödinger equation can be derived from (32):

$$\frac{\hbar^2}{2\mu}\,\Delta_x \varphi\,(x) + \frac{\alpha'^2 a}{M\omega^2}\,|\varphi\,(x)|^2\,\varphi\,(x) + W_0 \varphi\,(x) = 0 \tag{35}$$

which is actually the well-known Holstein model with classical oscillators, instead of quantum phonon fields.

Hence, in the limit of the strong coupling and low velocities, the continuum polaron description is equivalent to the model of soliton excitations in discrete chains.

Equation (35) has the following solution:

$$\varphi\,(x) = \pm\left(\sqrt{2r}\;ch\;\frac{x - x_0}{r}\right)^{-1},\; W_0 = -\,\frac{\hbar^2}{2\mu\,r^2},\; r = \frac{2\hbar^2}{\mu}\;\frac{K}{\alpha'^2 a} \tag{36}$$

where x_0 is an arbitrary constant; r is the radius of the localized state; $\omega = \sqrt{K/M}$.

The semi-classical representation shows that the polaron state (36) is more energetically favorable than the band state. If the absolute value of the polaron total energy $E_0 = 1/3W_0$ is greater than the energy shift $2|v|$ (cf. expression (5)):

$$|E_0| > 2|v| \tag{37}$$

the continuum approximation is broken and small radius polaron state is formed. However, in the quantum mechanical representation, the conditions (37) should be substituted by the following ones (cf. results of Section 24. 5):

$$|E_0| > 2|v| + |\Delta E| \tag{38}$$

where ΔE is given by (20).

There is also a further applicability condition for the continuum polaron theory in the strong coupling limit, namely, the continuity condition in its general form, $r >> a$, which can be recast as follows:

$$\sqrt{2\hbar^2 K/\mu} >> \alpha' a$$

and which is in line with (37). Thus, in this case, the polaron approaches may be used to describe solitons in DNA oligomers, and in particular, to calculate DNA soliton mobility.

24.8 PARTICLE MOTION IN AN ELECTRIC FIELD, AT THE LIMIT OF THE STRONG COUPLING

Expressions (32) and (34) become senseless, if the speed v is considered a real quantity, since then they would contain divergent values. In fact, v should be a complex quantity. From the physical point of view, it means that at every non-zero value of the velocity v, the Cherenkov emission of phonons takes place at the following values of the wave vectors: $|k| > \omega/v$. As a result, the v-quantity cannot be considered an eigenvalue of the 'particle + field' Hamiltonian, in other words, it is not a conserved quantity anymore.

Thus, the charged particle motion (i.e., polaron) becomes damped and, in the absence of the external influence, in the $t \to \infty$ limit, the polaron velocity tends to zero.

Below, we will consider a particular case, when the damping due to Cherenkov emission is small. Passing from k-summation to integration in the upper half-plane at $\lambda > \lambda'$ and in the lower half-plane at $\lambda < \lambda'$, we obtain the following expression for $U(\lambda)$ using (32):

$$U(\lambda) = 8\pi\hbar^2\overline{\alpha}\int_{-\infty}^{\infty}\frac{1}{v}\sin\frac{\omega}{v}\left|\lambda - \lambda'\right|.\left|\Psi\left(\lambda'\right)\right|^2 \mathrm{d}\lambda' \tag{39}$$

where v actually stands for the real part of the v-value. Thus, a Schrödinger equation for a moving polaron has the following form:

$$\frac{\mathrm{d}^2\Psi}{\mathrm{d}x^2} + \frac{\kappa'}{v}\int_{-\infty}^{\infty}\sin\frac{\omega}{v}\left|x - \mathrm{x}'\right|.\left|\Psi\left(x'\right)\right|^2 \mathrm{d}x' \cdot \Psi - W'\Psi = 0 \tag{40}$$

where $\kappa' = 4\pi\mu\overline{\alpha}$, $W' = 2\mu W/\hbar^2$.

One may easily verify that Eq. (40) has solution at any v value, and tends to Eq. (35) in the $v \to 0$ limit.

From (32) it follows that the particle energy W may be cast in the following form:

$$W[\Psi] = \frac{\hbar^2}{2\mu} \int_{-\infty}^{\infty} \left| \nabla \Psi \right|^2 \mathrm{d}x - \sum_k \frac{2\hbar\omega\overline{\alpha}}{\hbar^2\omega^2 - k^2 v^2} \cdot \left| \int_{-\infty}^{\infty} \mathrm{e}^{-ikx/\hbar} |\Psi(x)|^2 \mathrm{d}x \right|^2 \tag{41}$$

where $\Psi(x)$ is determined by the solutions of Eq. (40).

Using the well-known relation:

$$\frac{1}{\hbar\omega - kv} = P \frac{1}{\hbar\omega - kv} - i\pi\delta(\hbar\omega - kv) \tag{42}$$

we obtain the following expression for the number of energy quanta, N, which are emitted by a polaron per time unit (emission rate):

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \frac{\mathrm{Im}W}{\hbar}, \quad \mathrm{Im}W = \frac{\overline{\alpha}a}{\hbar v} \left| \int_{-\infty}^{\infty} \mathrm{e}^{-ix\omega/v} |\Psi(x)|^2 \mathrm{d}x \right|^2 \tag{43}$$

The energy loss by a moving polaron, determined by (43), causes a gradual slowing down of its motion and eventually its complete stopping.

Thus, polaron motion with constant velocity is possible only in an externally applied electric field. Then, the polaron equilibrium velocity is determined by the balance of energy, which is acquired by a polaron from the external electric field ξ and lost in the course of Cherenkov emission (43):

$$\hbar\omega \frac{\mathrm{d}N}{\mathrm{d}t} = \omega \mathrm{Im}W = e\xi v \tag{44}$$

From (43) and (44), the following relation between the electric field and the polaron velocity can be obtained:

$$\xi = \frac{\overline{\alpha}a\omega}{\hbar e v^2} \left| \int_{-\infty}^{\infty} \mathrm{e}^{-ix\omega/v} |\Psi(x)|^2 \mathrm{d}x \right|^2 \tag{45}$$

In the general case, the wavefunction $\Psi(x) = \Psi(x;v)$ in (45) should also be velocity-dependent and can be obtained from the solution of Eq. (40).

Now, let us consider several limiting cases. In the $v \to \infty$ limit, Eq. (45) leads to the following expression:

$$v = \sqrt{\frac{\overline{\alpha}a\omega}{\hbar e \xi}} \tag{46}$$

Equation (46) means that, at high polaron velocities, the latter will increase when the external electric field strength decreases, which is evidence of instability of the charge motion. This limit hence corresponds to the onset of negative differential conductivity. Such a velocity area can be used for generation and amplification of oscillations by an external electric field.

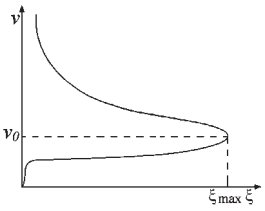


Fig. 24.4. The dependence of the particle velocity v on electric field ξ in the strong coupling limit.

In the opposite limiting case of low velocities $v \to 0$, the wavefunction $\Psi(x)$, which is a solution of Eq. (35), is determined by expression (36). By substituting $\Psi(x)$ from (36) into (45), we arrive at the following formula:

$$\int_{-\infty}^{\infty} e^{-ix\omega/v} |\Psi(x)|^2 \, dx = \frac{\pi}{2} \frac{\omega r/v}{sh(\pi\omega r/2v)} \tag{47}$$

Thus, it follows from (45) that:

$$\xi = \frac{\pi^2}{4} \frac{\overline{\alpha} a \omega^3 r^2}{\hbar e v^4} \frac{1}{sh^2(\pi\omega_r/2v)} \tag{48}$$

It should be noted that (48) tends to (46) in the $v \to \infty$ limit. This circumstance allows us to hope that expression (48), obtained in the low-velocity limit, delivers a qualitatively correct velocity-field relationship for any arbitrary velocity value.

The most important physical conclusion from (48) is the existence of some maximum value of the electric field strength ξ_{max}, at which a polaron motion with a constant velocity is possible (Fig. 24.4). The branch $0 < v < v_0$ corresponds to positive differential conductivity, i.e., to stable polaron motion. Then, the branch $v_0 < v < \infty$ corresponds to negative differential conductivity, i.e., to unstable polaron motion.

As for the case of weak coupling, here we have not taken into account the finiteness of the conduction band width, so that the very possibility of existence of the negative differential conductivity branch is questionable. It also remains unclear whether this negative differential conductivity branch is reachable under usual experimental conditions. In trying to surpass these limitations, the following section will consider charge motion in a discrete chain placed into an external electric field.

24.9 TEMPERATURE DEPENDENCE OF PARTICLE MOBILITY

Above, we have discussed the behavior of a charged particle in an electric field in a deformable chain at $T = 0$. The chain deformation caused by a particle moving in it, results in non-ohmic conductivity. In effect, any actual nucleotide chain always contains some deviations from the ideal structure caused by structural defects. At finite temperatures,

non-zero deviations of the nucleotides from their equilibrium positions occur even without interaction between the charged particle and internal vibrational modes. Therefore, in the presence of this interaction, the particle should move in a random field, caused by temperature fluctuations. The charged particle velocity is dependent on the electric field strength, and this dependence must be ohmic in the limit of a weak electric field.

To calculate charge carrier mobility in nucleotide sequences, we must begin with the Hamiltonian H_ε

$$H_\varepsilon = H + H', \qquad H' = \sum_n e\xi an|n\rangle\langle n| \tag{49}$$

where H is determined by expression (1). In agreement with this, the equations of motion (4) will be recast in the following form:

$$i\hbar\dot{b}_n = v\,(b_{n+1} + b_{n-1}) + e\xi anb_n + \alpha'_n q_n b_n \tag{50}$$

Internal vibrational modes of the sites in (50) are determined by the equations of motion for oscillators in the classical representation of normal modes q_n, affected by an external random force, $A_n(t)$:

$$M_n\ddot{q}_n = -\gamma_n\dot{q}_n - k_n q_n - A_n(t) - \alpha'_n|b_n|^2 \tag{51}$$

where γ_n is the friction factor. Eq. (51) is the conventional Langevin equation with random force $A_n(t)$, satisfying the following statistical properties:

$$\langle A_n(t)\rangle = 0, \qquad \langle A_n(t)A_n(t+t')\rangle = 2k_BT\,\gamma_n\delta_{nm}\delta(t') \tag{52}$$

By solving the simultaneous equations (50) – (52), one could arrive at the following expression for the single-electron current density, $J(t)$:

$$J(t) = N^{-1}\sum_n j_n(t) \tag{53}$$

$$J_n(t) = \frac{ie}{\hbar}\,v\left[b^*_n(t)b_{n+1}(t) - b^*_{n+1}(t)b_n(t)\right] \tag{54}$$

where $j_n(t)$ stands for the single-electron current density at the site number n [18].

The main difficulty of such an approach in the weak electric field case is that the current distributions are hardly convergent to the stationary solutions and require consideration of extremely long chains in the relevant computational experiments. Meanwhile, solving simultaneous differential equations of huge dimensions is still a formidable task, even for the most powerful modern supercomputers.

This is why, instead of directly calculating the current density using Eq. (54), it is much more convenient to estimate the charge mobility from the Kubo formula [27–29]:

$$\mu = \frac{e}{2T} \lim_{\varepsilon \to 0} \varepsilon^2 \int_0^\infty \langle x^2(t) \rangle \exp(-\varepsilon t)\, \mathrm{d}t \tag{55}$$

where T is the temperature; $\langle x^2 \rangle$ represents an ensemble averaging of a charge mean-square shifts x^2 given by the expression :

$$x^2(t) = \langle \Psi(t) | n^2 a^2 | \Psi(t) | \rangle = \sum_n |b_n(t)|^2 n^2 a^2 \tag{56}$$

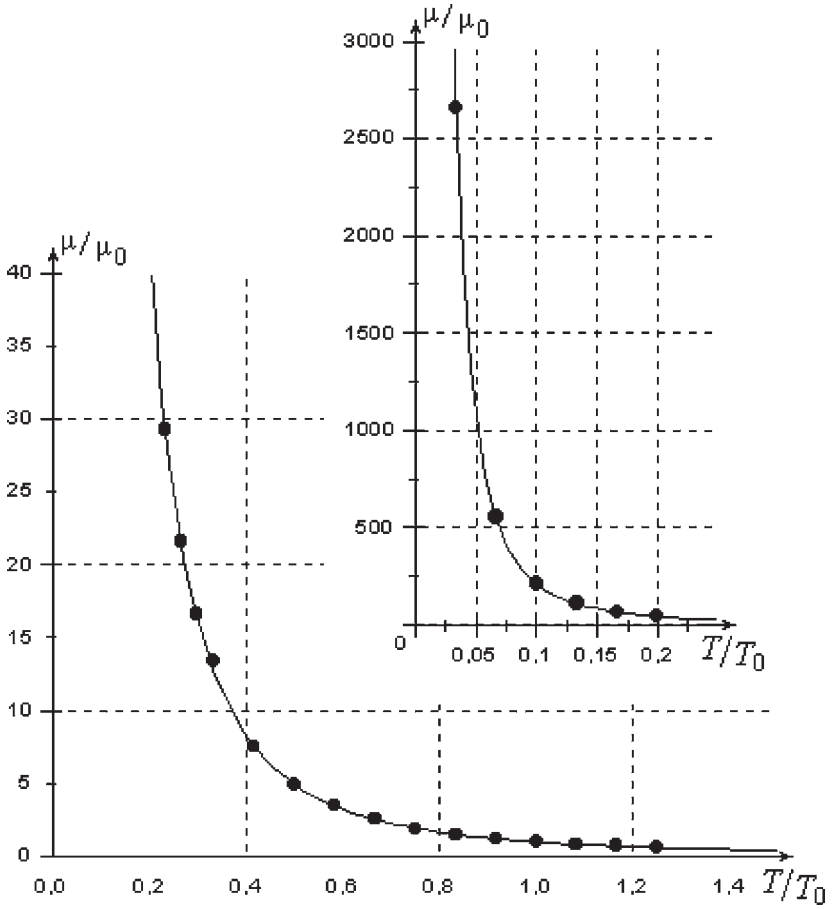


Fig. 24.5. Temperature dependence of hole mobility in DNA; $\mu_0 = \mu(T_0) = 2.87$ cm^2/V s, $T_0 = 300$ K.

Such a charged particle mobility calculation has been carried out in previous work [30], using expressions (50)–(52) and (55), for the hole mobility of a uniform $(G–C)_n$ nucleotide chain, which was estimated to be $\mu \approx 2{,}9$ cm^2/Vs at $T = 300$ K.

Figure 24.5 shows the results of this hole mobility estimation at the same parameter values as in [30], but in the temperature range of 10–375 K [31]. Figure 24.5 exhibits quick growth of the hole mobility with temperature decrease. At $T = 10$ K the mobility value becomes comparable with that of good semiconductors: $\mu(10\ K) = 10^4$ cm^2/Vs. The dependency, shown in Fig. 24.5, is approximated by the following function:

$$\mu/\mu_0 = (T_0/T)^{\gamma} \tag{57}$$

where $\gamma \approx 2{,}3$; μ_0 is mobility at $T_0 = 300$ K.

The increase in hole mobility with the temperature decrease may be qualitatively explained as follows. At every moment of time, a hole is moving in a random potential field as a result of temperature fluctuations, specified by random distribution of the site energies α_n. Since in the present one-dimensional case, there is a localized state due to the random field, an increase in the polaron mobility will occur along with a decrease in the random field amplitude.

According to (1), the latter tendency should also be observable in parallel with the decrease in the $\alpha'_n q_n$ contribution to the total hole energy at a particular site, i.e., along with temperature decrease as well.

Let us note that the described scenario is a rather general one, and may be applied not only to DNA. For example, Ref. [32] has reported on the experimental observation of the polaron mobility temperature dependence similar to that given by expression (57) in carefully purified organic molecular crystals.

Attempts to theoretically explain such a temperature dependence, based upon the Holstein model and the combined Holstein–Peierls model, were presented in Refs [33, 34].

24.10 LOW-TEMPERATURE PHENOMENA. BLOCH OSCILLATIONS

It should be noted that high values of the hole mobility in uniform polynucleotide chains at low temperatures (see Section 24.9 above) opens an opportunity for observation of some physical phenomena analogous to those occurring in conventional semiconductors.

It is well known that, in the case of ideally periodical crystal structure, an electron placed in it performs oscillations in an electric field [35].

In semiconductors with a super-lattice this results in such phenomena as emission of electromagnetic waves ('Bloch oscillator') [36], negative differential and absolute conductivities [37–39], dynamic localization [40–44], and multi-photon absorption [45,46].

We have shown in the section above that a one-dimensional and uniform polynucleotide chain is significantly different from a solid-state crystal, because numerous vibrational modes within the nucleotide pairs induce dynamical structural disorder owing to thermal fluctuations and the motion of injected charged particles.

In this section, we will study the problem of a hole transfer/transport through a nucleotide chain under an electric field, as determined by solution of Eqs (50), (51). In the

absence of interaction between the hole and the chain oscillators ($\alpha'_n = 0$), Eq. (50) has the following solution [47,48]:

$$b_n(t) = \sum_{m=-\infty}^{\infty} b_m(0)\,(-i)^{n-m}\,\mathrm{e}^{-i(n+m)\omega_B t/2}\,J_{n-m}\,(\zeta(t))$$
$$\zeta = (4\nu\sin(\omega_B t/2))/\hbar\omega_B \tag{58}$$

where $\hbar\omega_B = e\xi a$ is the Bloch frequency, J_n is the Bessel function of the first order. The solution (58) describes Bloch oscillations of the hole 'center of mass' $x(t) = \sum |b_n|^2 na$, with frequency ω_B. Using (58), $x(t)$ may be recast in the following form:

$$x(t) = x(0) + \frac{2a\nu}{\hbar\omega_B}\,|S_0|\,(\cos\Theta_0 - \cos\,(\omega_B t + \Theta_0))$$

$$S_0 = \sum_{m=-\infty}^{\infty} b_m^*(0)\,b_{m-1}(0) = |S_0|\mathrm{e}^{i\Theta_0},\; x(0) = a\sum_{m=-\infty}^{\infty} m|b_m(0)|^2 \tag{59}$$

Expression (59) exactly corresponds to the quasi-classical description of charge migration in a crystal lattice, if $b_m(0) \approx b_{m-1}(0)$, when $\Theta_0 = 0$, $S_0 = 1$.

According to (59), the oscillation amplitude $x(t)$ is equal to zero if the charge is localized at only one site in the initial moment. In such a case, the hole density should be symmetrically distributed around the initial position $x = 0$. From the classical point of view, the hole should then move along the field and against the field with accelerations that have equal absolute values and opposite directions. The hole will be gaining energy from the field when moving along the latter and loosing energy when moving in the opposite direction.

Unlike $x(t)$, the mean-square displacement $X^2(t)$ cast in (56) should be non-zero and adopts the following form:

$$X^2(t) = X^2(0) + \frac{a^2}{2}\left(\frac{4\nu}{\hbar\omega_B}\right)^2 \sin^2\frac{\omega_B t}{2}$$
$$X^2(0) = a^2\sum_{m=-\infty}^{\infty} m^2|b_m(0)|^2 \tag{60}$$

Thus, the charge density should undergo coherent oscillations with Bloch periodicity.

The hole motion modeling in a polynucleotide $(\mathrm{G–C})_n$ chain with $\alpha' \neq 0$, according to [31], has led to the following picture.

At the initial moment of time, the charge (the hole) is assumed to be localized, $q_n(0) = \dot{q}_n(0) = 0$, in the center of a chain containing 499 sites. That is, there should be no mass-center oscillation at $\alpha' = 0$. The situation under study is typical, since in the conventional charge transfer experiments a hole is injected into one of the nucleotide pairs as a result of photo-excitation.

Figure 24.6 shows the calculated dependencies $x(t)$ for electric field values $\xi_1 = ea\xi\tau/\hbar$ that start from the minimum value $\xi_{\min} = 8\nu/eaN$, where the chain finiteness does

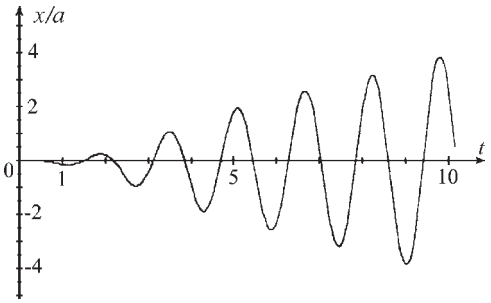


Fig. 24.6. Position expectation dependence on t(ps), for $\xi_1 = ea\xi\tau/\hbar = 0{,}04$.

not yet affect the hole motion, and go till the maximum value $\xi_{\max} = 4v/ea$, which corresponds to the Starck localization of the hole at one nucleotide pair. Thus, interaction between the hole and chain oscillator q_n results in a hole oscillation with a period T, very close to the period of Bloch oscillations $T_B = 2\pi/\omega_B$. Some changes in the friction constants γ and the electron–phonon coupling constants α' should affect the amplitude of these oscillations (which comes to zero at $\alpha' = 0$), but not their period.

The physical mechanism of this phenomenon, according to [31], can be considered as follows. Let us take an infinite rigid chain with $\alpha' = 0$. In this case, the hole localized at the beginning at a site number $n = 0$, will in the course of time result in charge density distributions symmetrical with respect to $n = 0$. The reason for this is that the hole moves in an anti-symmetric potential α_n, where its arrivals at the sites with numbers n, $-n$ occur with equal probabilities. If the electron/phonon coupling is non-zero, $\alpha' \neq 0$, then an initially symmetric density distribution will gradually loose its symmetry, according to Eq. (51) ($A_n(t) = 0$). As a result, the probabilities of charge transfer along the electric field and in the opposite direction become different. It is the latter circumstance that destroys the charge distribution symmetry and leads in turn to a non-zero shift of the hole 'center of mass' $x(t)$. Specifically, at $\alpha' \neq 0$, the absolute values of the hole accelerations along the field and against the field are not equal, and the energy gained by the hole from the field can be emitted in the form of electromagnetic waves, or quanta of nucleotide chain internal vibrations. Let us emphasize that the discussed mechanism is the general one, and could be made observable at any parameter values.

However, qualitative deviations from the above-mentioned behavior of the uniform chain should occur for regular (and non-uniform), DNA base sequences. In this case, the regular Bloch oscillations do not exist even at $T = 0$. According to [31], the reason for this phenomenon is the possibility of the charged particle undergoing transitions between or among the sub-bands formed in the regular sequences (see Section 24.3 above).

At finite temperatures, the Bloch oscillations collapse even in truly uniform chains. Ref. [31] has estimated the temperature range, within which it is still possible to observe the Bloch oscillations.

24.11 CONCLUSIONS

By and large, the results of experimental studies on charge transfer/transport in DNA still remain highly controversial. In the present communication, we did not attempt to solve all the existing problems. Within the frame of our relatively simple dynamical model of DNA, we have just tried to analyze the possible forms of a charged particle motion in an ideal chain, where there are neither disruptions nor deviations from the standard model with plane-parallel stacking of bases. By virtue of the fundamental character of standard model, the obtained results are of general theoretical interest, and can be helpful in formulating somewhat more detailed and sophisticated models. For example, the mobility computation scheme presented and discussed in Section 24.9 is a general one, and can be applied to systems with any number of degrees-of-freedom. For a practical realization of this scheme, one would have to carry out large-scale computations that become more formidable, the greater the number of degrees-of-freedom and model parameters that are being considered. This is why, the specific values presented here, e.g., the values of charge mobility, should be considered as an illustration of the methods in question, which will undoubtedly be refined in the future along with future sophistication of the model itself.

24.12 ACKNOWLEDGEMENTS

The work has been performed with support from the RFBR, grant numbers 03-04-49225, 04-07-90402.

24.13 REFERENCES

1 C. Dekker and M.A. Ratner, Phys. World, 14 (2001) 29.
2 A.A. Voityuk, N. Rösch, M. Bixon and J. Jortner, J. Phys. Chem. B, 104 (2000) 9740.
3 E. Meggers, M.E. Michel-Beyerle and B.J. Giese, J. Am. Chem. Soc., 120 (1998) 12950.
4 J. Jortner, M. Bixon, A.A. Voityuk and N. Rösch, J. Phys. Chem. B, 106 (2002) 7599.
5 M. Bixon, B.J. Giese, S. Wessely, T. Langenbacher, M.E. Michel-Beyerle and J. Jortner, Proc. Natl. Acad. Sci. USA, 96 (1999), 11713.
6 B.J. Giese, S. Wessely, M. Sporman, U. Lindemann, E. Meggers and M.E. Michel-Beyerle, Angew. Chem. Int. Ed., 38 (1999) 996.
7 S. Wessely and B.J. Giese, J. Biomol. Struct. Dynam., special issue, F2 (2000) 293.
8 C.J. Murphy, M.R. Arkin, Y. Jenkins, N.D. Ghatlia, S.H. Bossmann, N.J. Turro and J.K. Barton, Science, 262 (1993) 1025.
9 S.O. Kelley and J.K. Barton, Science, 283 (1999) 375.
10 C.A.M. Seidel, A. Schulz and M.H.M. Sauer, J. Phys. Chem., 100 (1996) 5541.
11 H.A. Wagenknecht, S.R. Rajski, M. Pascaly, E.D.A. Stemp and J.K. Barton, J. Am. Chem. Soc., 123 (2001) 4400.
12 A. Rakitin, P. Aich, C. Papadopoulos, Yu. Kobzar, A.S. Vedeneev, J.S. Lee and J.M. Xu, Phys. Rev. Lett., 86 (2001) 3670.
13 D.D. Eley and D.I. Spivey, Trans. Faraday Soc., 58 (1962) 411.
14 J.D. Watson and F.H. Crick, Nature, 171 (1953) 737.
15 D. Porath, A. Bezryadin, S. de Vries and C. Dekker, Nature, 403 (2000) 635.
16 V.D. Lakhno in V.D. Lakhno and M.N. Ustinin (Eds.), Computers and supercomputers in biology, Moscow-Izhevsk, 2002, pp. 137–171, Chapter 4 (in Russian).

17 N.S. Fialko and V.D. Lakhno, Phys. Lett. A, 278 (2000) 108.
18 N.S. Fialko and V.D. Lakhno, Regular & Chaotic Dynamics, 3 (2002) 299.
19 S. Steenken, Biol. Chem., 378 (1997) 1293.
20 M.G. Debije and W. Bernhard, J. Phys. Chem. B, 104 (2000) 7845–7851.
21 T. Holstein, Annals of Physics, 8 (1959) 325.
22 T. Holstein, Annals of Physics, 8 (1959) 343.
23 A.Y. Kasumov, M. Kociak, S. Gueron, B. Reulei, V.T. Volkov, D.V. Klinov and H. Bouchiat, Science, 291 (2001) 280.
24 N.N. Bogolubov, Ukrainskii Mat. Zh., 2 (1950) 3.
25 V.D Lakhno (Ed.), Polarons & Applications, Wiley, Chichester, 1994.
26 V.D. Lakhno and G.N. Chuev, Rev. Top. Probl., 38 (1995) 273.
27 R. Kubo, J. Phys. Soc. Jpn., 12 (1957) 570.
28 H. Scher and M. Lax, Phys. Rev. B, 7 (1973) 449.
29 J.C. Dyre and T.B. Schroder, Rev. Mod. Phys., 72 (2000) 873.
30 V.D. Lakhno and N.S. Fialko, JETP Lett., 78 (2003) 786.
31 V.D. Lakhno and N.S. Fialko, Pis'ma Zhetf, 79 (2004) 575.
32 N. Karl, Semiconductors, in O. Madelung, M. Schulz and H. Weiss (Eds.), Landolt-Börnstein: numerical data and functional relationship in science and technology, Springer, Berlin, 1985, Group III, 17, Pt.i, p. 106, Chapter 2.
33 V.M. Kenkre, J.D. Andersen, D.H. Dunlap and C.B. Duke, Phys. Rev. Lett., 62 (1989) 1165.
34 K. Hannewald and P.A. Bobbert, Appl. Phys. Lett., 85 (2004) 1535.
35 N.W. Ashcroft and N.D. Mermin, Solid state physics, Holt-Saunders Int., Ed., Philadelphia, 1981.
36 E.E. Mendez and G. Bastard, Phys. Today, 46 (6) (1993) 34.
37 L. Esaki and R. Tsu, IBM J. Res. Dev., 14 (1970) 61.
38 R. Tsu and G.H. Döhler, Phys. Rev. B, 12 (1975) 680.
39 B.J. Keay, S. Zenner, S.J. Allen, K.O. Maranovski, A.C. Cossard, U. Bhattacharya and M.J.W. Rodwell, Phys. Rev. Lett., 75 (1995) 4102.
40 K. Unterrainer, B.J. Keay, M.C. Wanke, D. Leonard, G. Medeiros-Ribeiro, U. Bhattacharya and M.J.W. Rodwell, Phys. Rev. Lett., 76 (1996) 2973.
41 D.H. Danlap and V.M. Kenkre, Phys. Lett. A, 127 (1988) 438.
42 M.H. Shon and H.N. Nazareno, J. Phys. Condens. Matter., 4 (1992) L611.
43 X.G. Zhao, Phys. Lett. A, 167 (1992) 291.
44 M. Holthaus, G.H. Ristow and D.W. Hone, Europhys. Lett., 32 (1995) 241.
45 B.S. Monozon, J.L. Dunn and C.A. Bates, Phys. Rev. B, 50 (1994) 17097.
46 B.S. Monozon, J.L. Dunn and C.A. Bates, J. Phys. Condens. Matter., 8 (1996) 877.
47 A.M. Bouchard and M. Luban, Phys. Rev. B, 52 (1995) 5105.
48 M. Luban, J. Math. Phys., 26 (1985) 2386.

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 25

Electronic structure of DNA derivatives and mimics by density functional theory

Rosa Di Felice and Arrigo Calzolari

National Centre on nanoStructures and bioSystems at Surfaces (S3) of INFM-CNR, c/o Dipartimento di Fisica, Università di Modena e Reggio Emilia, Via Campi 213/A, 41100 Modena, Italy

Abstract

The prospect to extend nanotechnology concepts to self-assembling device architectures has recently driven efforts to demonstrate that biomolecules may perform electrical functions in a non-native environment. One particular research line in this framework has evolved for the measurement of quantum transport through single native and modified DNA molecules. These experiments directly probe the electronic structure of the conducting material: hence, they also stimulated a huge theoretical activity to understand the electronic structure of DNA-based one-dimensional polymers, best explored by density functional theory computations. After a brief summary of existing reviews in the field, we focus in this chapter on collecting our most recent results for two selected periodic polymers constructed as DNA modifications, which may constitute valid alternatives to native DNA for the development of molecular electronics.

25.1 MOTIVATION AND BACKGROUND

The current intense interest in the experimental and theoretical probing of the electronic structure of DNA-based polymers arose in connection to the quest for the development of molecular electronics. We do not intend to discuss the birth of molecular electronics and in particular of DNA electronics: for a recent review on molecular electronics see Ref. [1] that cites the most relevant original contributions and collections in the field to date; other topical reviews focus on DNA electronics [2–4]. We only point out here the salient results that motivated the pursuit of DNA derivatives and mimics beyond native DNA: the desired 'mutants' should exhibit enhanced conductivity and/or other exploitable functions, while maintaining the inherent recognition and structuring treats of native Watson–Crick double helices that are demanded for self-assembling.

From the analysis of the experimental literature until 2003, Porath and co-authors [2] concluded that: (i) charges may be transported with poor conductivity in short single DNA molecules or in longer molecules organized in bundles and networks; (ii) charge flow is blocked for long molecules deposited onto 'hard' inorganic substrates. Similar conclusions were drawn by Endres and co-authors [3]. A reasonable advanced hypothesis was that the inability to conduct is not due to the intrinsic electronic structure, but to strong deformations induced by the substrate [5], that completely alter the structure and the regular π-stack. Indeed, two recent experiments devoted to optimize the molecule–electrode contacts and realize current–voltage measurements on suspended DNA polymers by conductive atomic force microscopy (cAFM) [6,7], thus avoiding the contact with the substrate, demonstrated that it is possible to solicit a current flow through short stacks of 26 base pairs in one case [6], and of up to 14 base pairs in the other [7]. We do not wish to suggest that these two experiments can be interpreted in the same way in terms of conduction ability and mechanisms, since the measured currents as a function of voltage are different in both shape and intensity. Yet, they both show the efficiency of avoiding the contact to a substrate.

Given the situation illustrated above, two solutions remain to continue pursuing the route towards DNA-based electronics: either to reduce/avoid substrate-induced deformations (standing-molecule measurements, use of a soft organic buffer, etc.), or to explore stiffer molecules. Along the former possibility, given the fact that currents are detected only for very short DNA fragments and have low intensities even when the problem of substrate contact is avoided [6,7], metal incorporation and base alteration are pursued to enhance the intrinsic conductivity: the CuHy-wire that we describe later encloses both aspects. Along the latter possibility, G4-DNA emerges as a valuable candidate, and also bears the promise of a better conductivity.

25.1.1 G4-wires

G4-DNA is a quadruple helical structure of homoguanilic or guanine-rich sequences, that occurs in a wide variety of natural situations and organisms. It can play an important role in the telomeric region of chromosomes, and is an excellent prototype case to study supramolecular self-assembling. The chemico-physical and biological aspects of G4-DNA were recently reviewed [8]. G4-DNA molecules are currently also attracting interest within the molecular electronics research community [2,9]. This envisaged novel role of G4-DNA is proposed on the basis of some structural features that are expected to influence the electronic characteristics in a device set-up. (i) The peculiar in-plane arrangement of four guanines forming a tetrad, kept together by a double ring of eight hydrogen bonds (see Fig. 25.1), endows the quadruple helices with a higher stiffness than the DNA double helices: as a consequence, a higher persistence length than B-DNA is indeed reported for G4-DNA molecules deposited on a mica surface [9,10]. (ii) The high molecular density in a tiny cylindrical section (diameter 2.3 nm versus 2.1 nm of B-DNA) also induces a highly packed charge density distribution and channels for charge motion [11]. (iii) The experimental evidence that the quadruplex arrangement is stabilized by the presence of metal cations in solution during synthesis [8], stimulates the idea of a metal-wire accomplishment at the core of the molecule, which in principle could mediate charge motion [12].

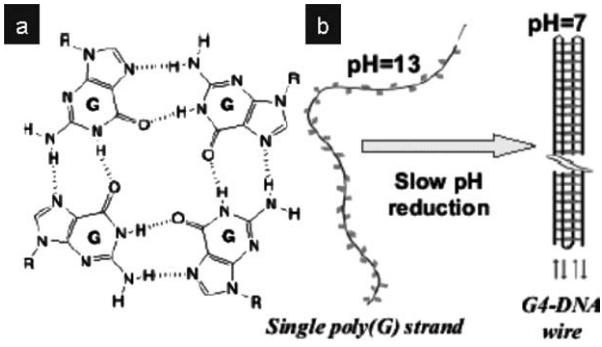


Fig. 25.1. (a) Each plane of the G4-wires is a tetrad of four guanines connected by eight hydrogen bonds. Besides this basic unit, metals of different chemical nature can be incorporated in the inner hole at various locations in the planes or between the planes. (b) Self-assembly process recently demonstrated to obtain long G4-wires from single poly(G) strands [9]. (From Ref. [9], with permission; copyright ©2005 Wiley-VCH.)

(iv) The rotation angle of 30º between adjacent guanines, smaller than in double-stranded guanine-rich DNA sequences, might be an indication of a better $\pi - \pi$ superposition (that is expected to improve with decreasing angle [13]).

Short G4-wires are well known in a variety of self-assembly schemes (e.g., with a sugar–phosphate backbone or without a backbone with the guanines decorated by alkylic chains) and have been characterized with nanoscopic techniques for a while [8,10,14]. Huge efforts have been devoted most recently to the synthesis and characterization of long G4-wires suitable for molecular electronics. The successful demonstration of an efficient synthesis protocol, and the accompanying imaging characterization for such materials [9], are encouraging signals towards G4-based electronics, and motivate our continuous interest in the electronic structure of the guanine tetrameric units and the dependence on external factors such as complexation with metal ions and structural deformations [11,12,15].

25.1.2 CuHy-wire

The manipulation of DNA duplexes with metal ions is explored as a viable trick to replace hydrogen bonding selectivity [16,17] and to impart an electrical function [18] to these biomolecules for exploitation in biotemplated nanotechnology. We remind the reader that the envisaged nanotechnology applications of DNA are not limited to molecular electronics (which is indeed probably the most alien), but extend to several other cases, such as molecular motors [19–21], three-dimensional cages [22,23], simple logical operations [24], and computing machines [25]. Among the newest suggestions, we find that of molecular magnets [17].

Another kind of DNA manipulation is the modification of nucleobases, in particular through insertion of an aromatic ring between the pentagonal and the hexagonal units in purines [16,26,27] but also through other paradigms [28]. This is becoming a very lively field in nucleic acid synthesis: the idea is the exploration of nucleobase mimics as an expansion of the genetic alphabet.

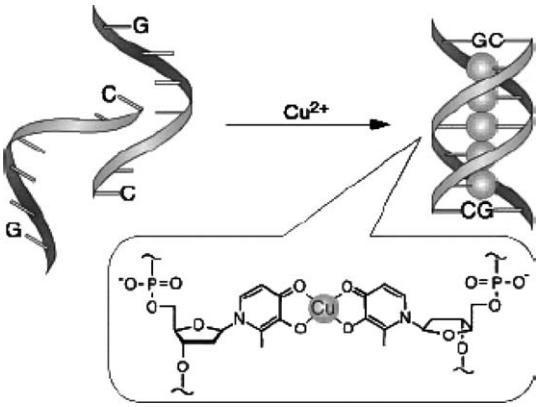


Fig. 25.2. The figure shows the self-assembly process of artificial DNA strands in which some of the native nucleobases (except the terminal ones) can be substituted by hydroxypyridone bases: if Cu(II) ions are present in solution during synthesis, they can be incorporated along the axis and form molecules with a magnetic moment. The inset shows the artificial H–Cu(II)–H pair: the Cu(II) ion in each plane is tightly coordinated by four surrounding electronegative O atoms (see Color Plate 33). (From [17], with permission; copyright ©2003 Science.)

The helical motif studied by Tanaka and co-workers [17] combines metal incorporation with base modification. They succeded in synthesizing double helices in which metal ions are aligned along the axis through alternative base pairing mediated by nucleoside mimics, thus demonstrating that metal coordination can successfully replace the hydrogen-bonding pairing pattern, while preserving the helical conformation and the autorecognition feature (see Fig. 25.2). Alernative chelator nucleosides [29,30] are employed to insert metal ions into double helices: in the experimental work that inspired our theoretical investigation, the chelator is hydroxypyridone (hereby labeled **H**) and the metal is Cu in the Cu(II) oxidation state. Starting from short DNA molecules, Tanaka and co-workers substituted one to five planes of native base pairs with Cu-bridged hydroxypyridone pairs (Fig. 25.2); the altered pairs were decorated by Watson–Crick GC pairs at the edges of the stacks. The most striking result from the experiment is the ferromagnetic behaviour that we address by spin-polarized DFT.

25.1.3 State-of-the-art DFT calculations of DNA structures

The most sophisticated computational techniques to tackle the electronic structure of molecules are based on quantum chemistry methods that write explicitly the electron wave functions in terms of Slater determinants [31]. For what specifically concerns DNA, important information was gained during the years on gas-phase isolated nucleobases and simple assemblies/complexes by applying the Møller–Plesset second-order perturbation theory (MP2) [31] to the ground-state Hartree–Fock (HF) configurations [32–34]. Such studies yield accurate geometries and relative formation energies, as well as excited-state electronic properties including correlation effects. However, they are very cumbersome and thus limited to few-atom systems. Quantum chemistry methods

mostly limited to the HF level were also applied to enquire about the electronic structure and transfer integrals of base and base-pair stacks [35].

With the advent of clear-cut experiments on single DNA molecules, the demand for addressing realistic helical conformations became evermore compelling and scientists turned to density functional theory (DFT) [36,37]. The computational load is drastically quenched with respect to wavefunction-based methods, at the expense of quantitative accuracy and predictivity: in fact, correlation effects are treated only at the mean-field level within an effective potential, and dispersion forces are missing. Despite such drawbacks, DFT is becoming increasingly popular and reliable for investigating the electronic structure of biomolecules and reactive chemistry [38]. The performance of DFT relative to other quantum chemistry methods to investigate nucleic acids was analysed within the framework of understanding charge transfer mechanisms [39]. For the sake of clarity, we wish to point out that in the current level of theory/implementation development, purely DFT studies give access to the electronic structure but do not yield a direct interpretation of charge motion mechanisms. In order to proceed further and to get close to the simulation of experimental data in the field of molecular electronics, progress is required in at least two directions: (i) on one hand, the gap between understanding transport through extended states (as in solids) and describing charge transfer between localized states (as in molecules) should be closed, based on theoretical developments[1]; (ii) on the other hand, the coupling of the electronic structure with intrinsic scattering mechanisms [43,44], vibrations [45], defects and the environment [46], is desired[2].

In the last five years, DFT studies on a variety of DNA double helices and focusing on different physical phenomena (pure electronic structure of frozen systems, effects of water and counter-ions, effects of vibrations, etc.) have been conducted. The majority of DFT codes are inherited from solid-state applications and are commonly applied to periodic systems. In this spirit, periodic infinite double helices obtained by a finite repeat unit were described: Z-DNA with a 12-plane unit [47,48], A-DNA with an 11-plane unit [49–51], B-DNA with a 10-plane unit [52–54]. DFT simulations were also performed for short finite DNA oligomers in a solution environment with water and counter-ions [55], to enquire into hole localization associated with counter-ion motion. A further application of DFT concerns the extraction of transfer integrals from the *ab initio* Hamiltonian matrix elements [56,57], to construct a simplified tight-binding Hamiltonian for transport calculations: the thus computed intra-strand and inter-strand couplings are related to the electronic factor of Marcus' theory, whose range of values in different DNA contexts (sequences, environment, length, etc.) has been critically interpreted in terms of mechanisms for charge motion [35,58,59].

[1] *Ab initio* DFT bandstructure calculations give an equilibrium description of the electronic properties, whereas transport characteristics require an account of non-equilibrium effects. The opposite approach, starting from molecular states has an inherent non-equilibrium nature but was traditionally based on empirical electronic parameters. A significant attempt to bridge the two opposite starting points was recently done by A. Nitzan [40–42]: despite the approximations of his formulation, this is to date the most advanced endeavour to match the Marcus' electron transfer theory with the theory of transport through extended electron states.

[2] Whereas methods to describe electronic correlation [43,44] and electron–phonon coupling [45] in an *ab initio* fashion exist, their application to realistic DNA molecules in a nano-junction setup is presently hindered by the high computational cost. Software and hardware progress will help bypass the current limitations.

Most such contributions up until 2003 have already been commented upon and we refer the reader to recent reviews on DNA-based molecular electronics [2–4].

In just one critical short paragraph, we summarize our personal understanding of the conclusions from DFT studies of infinite DNA wires in different conformations in view of the topics developed in the rest of the chapter, and according to transport paradigms usually employed to describe crystals: namely, by labelling the materials as metal or insulator or semiconductor[3]. DNA polymers behave as 'wide-bandgap semiconductors'. Owing to the well-known failure of DFT to evaluate bandgap values, it is impossible to know the exact entity, but it spans the range between 4 eV and 8 eV. However, the 'bands' do not have exactly the same meaning as in perfect crystals. True coherent dispersive bands with very small bandwidths are only obtained by imposing the helical symmetry, which corresponds to setting up perfect one-dimensional wires [51,54]: this perfection may be very weak under natural circumstances and the environmental perturbations are not obviously manageable as weak perturbations. Alternatively, by imposing only the translational symmetry to build periodic wires, the electronic structure is made of incoherent manifolds stemming from the energy levels of the single bases: each level yields a quasi-flat band; each whole manifold contains a number of such bands equal to the number of planes in the periodicity unit [47,48], and its amplitude should be compared with the bandwidth of the corresponding unfolded band obtained under the helical constraint. For instance, in the symmetric description of poly(G)–poly(C) A-DNA given by Artacho and co-workers [51], the π-HOMO of isolated guanine gives origin to a band with an amplitude of 40 meV and the π-LUMO of isolated cytosine to a band with an amplitude of 270 meV, unfolded into 11 **k** points (because of the modulo-11 periodicity) of the Brillouin zone relative to a single inter-plane separation. In the non-symmetric description of poly(GC)–poly(CG) Z-DNA given by Gervasio and collaborators [47], according to their description, 12 quasi-degenerate states (because of the modulo-12 periodicity) are positioned at the top of the valence band. These states have a π-character and are mostly localized on the guanine nucleobases and basically constitute the HOMO manifold; the 'quasi-degeneracy' claimed by the authors is consistent with the very small amplitude reported above for the symmetric case, although a strict comparison is hindered by the difference in base sequence; Na^+/PO_4^- states appear in the bandgap. Extended states with a good degree of delocalization through the polymer, which is an index of high charge transfer integrals and possibly efficient charge motion, can be found only in the case of uniform sequences; charge continuity is easily destroyed by a defect in sequence continuity [49,51]. These are essentially the properties that one would like to improve to attain DNA-based molecular wires: band dispersion and consequent bandwidth, coherence, high-density continuous orbital channels for charge motion, the ability to insert in a controlled way electron states within the fundamental HOMO–LUMO bandgap.

Other noteworthy DFT studies published very recently and not yet reviewed have turned attention to 'Bandgap tunnelling states in DNA' [53] and to the 'optical conductivity of wet

[3] We warn the reader that these notions employed for the description of conduction mechanisms of conventional materials may not be suitable for flexible one-dimensional DNA polymers. The development of the field of DNA-based electronics may reveal a new conceptual framework to properly represent the effective phenomena for charge motion.

DNA' [52]. Whereas we believe that these works contribute important pieces in the puzzle of understanding DNA conductivity and how this can be manoeuvred through designed actions (both in synthesis and measurements), we do not comment specifically here because they focus on particular aspects that are not directly instructive for our later arguments. The issues presented in the following are instead inspired by the search for other DNA-based materials in the attempt to optimize the bandstructure parameters described above.

25.2 COMPUTATIONAL DETAILS

25.2.1 Method

Our investigations are based on periodic DFT calculations [36], as implemented in the PWSCF codes [60]. The electron–ion interaction is described by non-norm-conserving ultrasoft pseudo-potentials for all the species [61]. The electronic wave functions are expanded on a basis of plane waves (PWs). The main advantages of such a basis set are that it does not suffer the basis set superposition error [62,63], and its completeness is controlled by a unique parameter, namely the kinetic energy cut-off E_c: in all the calculations presented here, E_c=25 *Ry*. The exchange-correlation functional is treated with gradient corrections beyond LDA, according to the PW91 formulation [64]. The ability and accuracy of the method to describe DNA-related systems was extensively tested on the individual bases [13] and on simple aggregates [65–67] before tackling the wire-like assemblies presented here.

25.2.2 Systems

The investigated one-dimensional polymers are simulated by repeated supercells containing several atoms. Orthorhombic supercells of different sizes are employed.

25.2.2.1 G4-wires

G4-wires can be successfully synthesized using both sugar–phosphate bonded strands [9] and lipophylic guanosine monomers [8,14], through different chemical protocols. Nonetheless, the resulting self-assembled product presents the same helical fashion: the core of the wires is constituted of stacked planes of four guanines, named G4s[4], as shown in Fig. 25.3(a). The G-quartets have a square-like symmetry and are separated by 3.4 Å along the stacking direction, with a rotation angle of 30º. Metal ions may be hosted in the central channel, resulting from the supramolecular G4 stacking.

Since the helical motif is independent of the inter-plane backbone linkers, and because of our specific interest in the electronic properties of base assemblies, we focus on the guanines alone, dropping the surrounding medium and the external attached units, as shown in Fig. 25.3(c) for a single tetramer. The external groups are replaced by H atoms

[4] A plane of four Gs in the configuration of Fig. 25.3(c) can be alternatively named G-quartet, or tetrad, or tetramer.

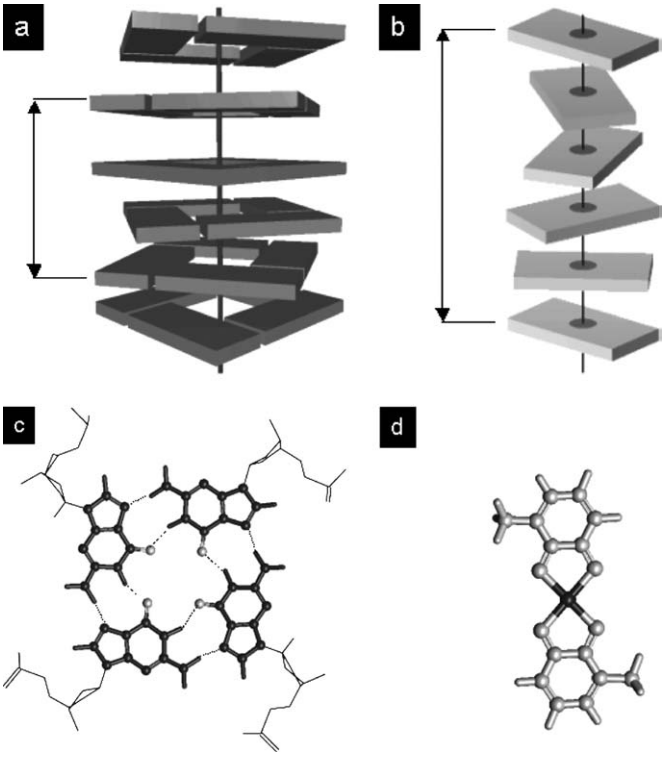


Fig. 25.3. (a) Schematic view of an empty G4-wire. Each plane is represented as a square brick with a central hole, to indicate the square symmetry and the collection of four identical units (the guanines). The overall appearance is that of an empty tube, inside which various metals can be hosted. The three-plane unit building block of the periodic system is identified by the vertical arrow. An atomistic sketch of a tetrameric plane is shown in (c), where the eight hydrogen bonds are clearly marked. Thin sticks represent the attached sugar–phosphate units, not included in the present description. (b) Schematic view of the CuHy-wire. Each plane is represented as a rectangular brick and the dark circles indicate the central Cu(II) ions. The five-plane periodicity unit of the ferromagnetic wire is indicated by the vertical arrow; the anti-ferromagnetic phase has instead a 10-plane periodicity unit [**H–**Cu(II)**–H**] obtained by doubling the pentamer shown here. An atomistic sketch of the **H–**Cu(II)**–H** pair constituting each plane is shown in (d).

to saturate dangling bonds. Further theoretical investigations about G4-wires [68] and native DNA [47] proved *a posteriori* that the inclusion of the external backbone does not significantly affect the electronic properties of the systems.

The G4-wires are simulated by means of periodically repeated supercells (24.3 × 24.3 × 10.2 Å^3), each containing three G4 planes. Taking advantage from the square symmetry of the G-quartets and from the inter-plane rotation angle, three consecutive planes are sufficient to model the whole helix [15], the fourth quartet being the translated image of the first one (Fig. 25.3(a)). The wires are laterally separated by a thick vacuum layer (~16 Å), in order to avoid spurious interactions among neighbour replicas. The isolated G-quartet is simulated in a large supercell, where we include a single tetramer and the same vacuum thickness (~16 Å) in any spatial direction.

The starting atomic positions of guanines are derived by an X-ray analysis of short four-stranded helices [69]. Three metal ions may be added in the central cavity along the wire. We consider the case of the empty quadruple helix, as well as the inclusion of monovalent (K, Ag, Cu) and divalent (Cu) ions within the stack. All the structures are relaxed until the forces on all atoms are lower than 0.03 eV/Å.

In the case of Cu(II) ions, the occurrence of unpaired electrons requires a spin-polarized description, that we treat within the local spin density (LSD) approximation.

25.2.2.2 CuHy-wire

The isolated **H–**Cu(II)**–H** planar building block (Fig. 25.3(d)) is simulated by a supercell $23.5 \times 23.5 \times 13.5$ Å^3 wide: this choice ensures that neighbouring replicas are separated by a vacuum thickness of at least 10 Å, in any spatial direction.

According to the experimental characterization [17], adjacent planes are rotated by 36º, and separated by a distance of 3.7 Å, forming a helical motif very similar to that of DNA in a mixed sequence with Watson–Crick native base pairs. The infinite CuHy-wire is constructed on the basis of these experimental parameters, through periodic replication of the [**H–**Cu(II)**–H**] stacked decamer (Fig. 25.3(c,d)), and the simulation supercell is $23.5 \times 23.5 \times 37.0$ Å^3 wide: this choice ensures that neighboring replicas are separated by a vacuum thickness of at least 10 Å in the directions perpendicular to the wire, whereas there is continuity in the wire direction. Owing to the equivalence of the two **H** bases in each Cu-mediated pair, the helical symmetry is reproduced with five planes instead of the 10 planes as in Watson–Crick pairs formed by complementary bases. However, our simulation unit contains 10 planes in order to describe with the same accuracy both the ferromagnetic (FM) and the anti-ferromagnetic (AM) phases.

The structure of the **H–**Cu(II)**–H** plane is fully relaxed until the atomic forces vanish within 0.03 eV/Å. For the CuHy-wire the geometry is not optimized, because the lack of Van der Waals interactions limits the accuracy of DFT to predict stacked geometries. Instead, we assemble the structures according to the experimental indications, verify that they are characterized by small atomic forces, and focus only on the electronic properties.

For both the monomer and the wire we present spin-polarized LSD electronic structure calculations.

25.3 RESULTS AND DISCUSSION ON G4-WIRES

25.3.1 K(I)-trapping and empty G4-wires

The typical experimental examples of G4-wires are realized by intercalating G-quartets with alkali cations (especially K and Na): the stoichiometry is one ion for every guanine tetrad. Quadruple helices have also been obtained with other metals (e.g., Ca, Ba, Sr, Pb). However, the most studied case is ‘G4 planes with K(I) ions’, because of the occurrence in natural systems [70]. Potassium seems also to be the best stabilizer of the four-handed structure.

The theoretical study of K(I)-trapping wires is crucial to understanding the properties of a realistic system (largely characterized with experimental tools), but also as a reference point for further investigations.

The most accurate X-ray structure available in the literature [69] includes Na ions instead of K. Sodium ions are mainly hosted between consecutive tetrads (inter-plane position), except for the external layers of the finite stacks where they may also be located on G4 planes[5] (intra-plane position). To simulate the K(I)-G4-wire we substitute sodium with potassium in the inter-plane positions from the NDB file UDF062, and let the system relax. In the optimized geometry K(I) ions remain at inter-layer locations, in agreement with the experimental findings [8]. Metal ions end up being bipyramidally coordinated with the eight oxygen atoms (four above and four below) of first-neighbour guanine molecules [11].

In order to investigate the existence of other metastable positions of the ions (as in the case of Na), and to roughly estimate a possible ion mobility through the helix, we performed a set of static (fixed-atom) electronic structure calculations, sampling the positions of the ions labelled as 1 to 7 in Fig. 25.4(a). Positions 1 and 7 mark the inter-plane site, position 4 marks the intra-plane site. The resulting total energies for each configuration are shown in the top panel of Fig. 25.4(b), where the origin of the energy scale is fixed as the minimum obtained value. The plot clearly shows that in the case of K(I) ions the intra-plane position is unfavored by ~ 3.8 eV, with respect to the inter-plane site. Even though a full atom relaxation is expected to reduce this energy barrier, the energy difference between the two configurations is sufficiently high to exclude the possibility of a free ion migration along the wire. These results have been recently more thoroughly inspected by molecular dynamics simulations at finite temperature [71].

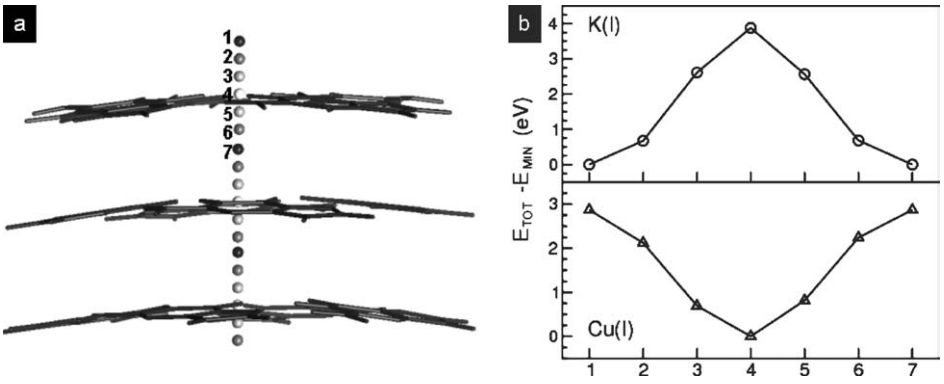


Fig. 25.4. (a) Stick representation of a G4-wire (dark grey frame) where K(I) and Cu(II) ions are located at different positions along the axis, indicated by the inner spheres of different grey intensities. Such positions are not simultaneouly occupied, but exclusive to each other: only one of them (plus the periods by 3.4 Å) for each different calculation. The total energy of the system with the ions at each location is computed to determine the preferred location and an approximate migration energy barrier. (b) Relative total energy, with respect to the minimum total energy, versus the ion location.

[5] See for instance Fig. 1 of Ref. [11].

Another interesting question is: are the metal ions necessary to form and stabilize the quadruple helices? So far, the answer has been absolutely affirmative, because ions have always been detected in characterization studies of G-quartets [8]. In particular, it was thought that charged cations interact electrostatically with the negative electronic distribution of the guanine molecules, adding an attractive contribution to the stability of the overall structure. However, the experimental samples were typically obtained from short oligonucleotides or from a few single monomers. On the other hand, recent experimental results about very long ($> 1\,000$ planes) helices [9] indicate the formation of stable empty G4-wires, i.e., without inner cations.

To gain insights into this problem, we compare the stability[6] of G4-wires with and without metal cations. Our repeated supercell approach describes infinite systems, neglecting finite-size effects at the borders, thus it best simulates the properties of long wires. The analysis of the formation energies (i.e., relative stability) is complicated by the different charge states of the empty (neutral) and K(I)-trapping (charged) G4-wires, and must be done as a function of the electron chemical potential (Fermi level). The method is described elsewhere [15]; here we summarize the results. Both the empty and the K(I)-trapping G4-wires turn out to be energetically favoured with respect to the single guanine. On the other hand, only the K(I)-G4-wire can be favourable with respect to isolated G-quartets when the system Fermi level is pinned very close to the occupied levels, whereas the empty G4-wire is unstable with respect to isolated G-quartets. Apparently, this finding would mean that indeed metal cations are crucial for the formation of the guanine quadruple helices, which is in contrast to recent experimental suggestions. We remark, however, that our simulations neglect several environmental effects (all the backbone, the solution and the counter-ions), which likely have an important role in stability issues. This problem may be circumvented by classical molecular dynamics (MD) simulations that, due to the lower computational burden, can take into account all the surroundings.[7] MD simulations cannot be employed for electronic structure problems, but the available force fields are deemed as sufficiently accurate for a desciption of the energetics of nucleic acids. Indeed, the results of our classical MD investigation [71] indicate that long G4-wires are stable at room temperature both without and with cations (K, Na, Li). Instead, for short chains [71,72] the electrostatic contribution induced by the ion coordination is mandatory to balance the boundary effects and the thermal fluctuations.

We conclude this section with the description of the electronic and conduction properties of the K(I)-G4 wires. As mentioned generically in Section 25.1.3, the translational invariance alone is not sufficient to obtain degenerate and dispersive energy bands [11,15]: symmetric unfolded bands are obtained only under helical constraint [68]. On the other hand, the analysis of the electronic structure shows the formation of gathered manifolds of energy levels [11,15], whose number is related to the number of molecules included in the simulation unit cell. We interpret the spreading of energy levels in a manifold in terms of the formation of an *effective* band-like peak[8] in the density of states

[6] Our simulations are performed at $T = 0$ K, thus the analysis of the stability includes only the enthalpy contribution and neglects the entropic one.

[7] Finite stacks with border effects, the backbone, water and various species of counter-ions.

[8] The fingerprint of an *effective* band-like peak with respect to a true band peak is the loss of pure coherence.

(DOS), which may constitute an *effective* channel for the electronic transport through the helical axis [73]. The DOS of the K(I)-G4 wire (Fig. 25.5(a)) appears similar to those of wide-bandgap semiconductors (E_{gap} = 3.3 eV), with dispersive band-like structures. For instance, the resulting HOMO and LUMO peaks have a finite width (Δ_{HOMO} and Δ_{LUMO}, respectively) of ~ 0.3 eV.

The formation of spread manifolds is a common feature of the DNA-like stacked systems and derives from the interactions among the nucleobases: in fact, a similar DOS curve is also observed in the case of the empty G4-helix [11,15]. Such molecular coupling may be described through the convolution of the single-particle electronic states belonging to the same manifold. Figure 25.6 (a) shows the isosurface plot of the HOMO manifold of the isolated G-quartet. It is the linear superposition of the HOMO states of the single guanine molecules that constitute the tetrad. In this case, the in-plane H-bonding interactions are not able to induce the formation of electronic channels that might support transport. However, if we also consider the out-of-plane interactions (e.g., $\pi - \pi$ coupling) in the stacked wires, we observe the formation of extended orbitals, delocalized along the stacking direction, that may be interpreted as *effective* conduction channels, corresponding to the *effective* dispersive bands of the manifolds. Figure 25.6(b) shows, in the case of the empty tube, how the HOMO states of the single guanines may interact, to sustain a weak *effective* conductivity along the wire.

If these issues are caused by guanine–guanine interaction, what is the effect induced by the inner metals on the electronic structure of the wire? There is not a single answer to this question, because it depends on the specific choice of the metal ions included in the stack (see Section 25.3.2). In the K(I) case, the states of potassium essentially do not interact with those of the guanine stack, limiting the metal/molecule interaction to the electrostatic attraction described above but with no chemical hybridization. Indeed, the

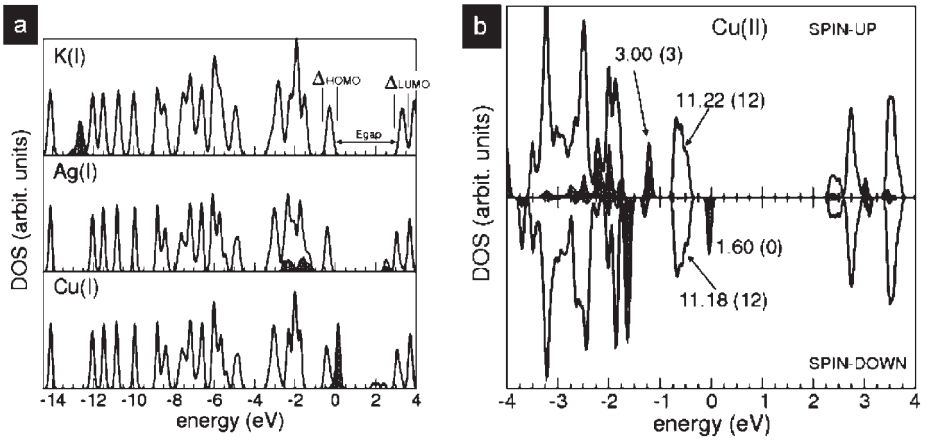


Fig. 25.5. (a) Total (solid black line) and metal-projected (grey area) Density of States for G4-wires containing K (top), Ag (middle), and Cu (bottom) ions in the 15 oxidation state int eh inner cavity. (b) Total (solid black line) and metal-projected (grey area) Density of States for a G4-wire containing Cu ions in the 25 oxidation state in the inner cavity. Spin-up (spin-down) indicate the majority (minority) spin component. The computed occupation numbers are compared with the nominal ones, reported in parentheses. Non-colinearity is not allowed in our approach.

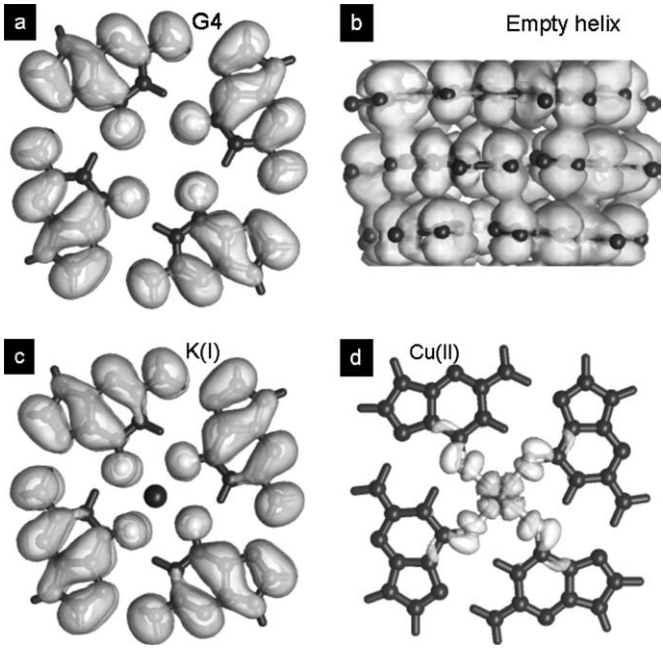


Fig. 25.6. Isosurface plots of the HOMO states for various G4 tetrameric units and an empty G4-wire. (a) π HOMO of a G4 tetrad. (b) π HOMO manifold (convolution of top occupied 12 states) for an empty G4-wire. (c) Plane-projection for a K(I)-G4-wire of the convolution equivalent to that shown in (b): the π nature and the absence of metal localization is clear. (d) Plane-projection of the SOMO manifold (convolution of three partially occupied molecular orbitals) for a Cu(II)-containing G4-wire, where the Cu ions migrate to in-plane positions: the σ nature is evident and the convolution is equally distributed on the 3 planes of the periodicity unit.

close-shell K_{3p} orbitals originate a peak (shaded area in Fig. 25.5(a)) in the low energy range of the projected density of states (PDOS) at 12 eV below the HOMO, hosted in a gap of the guanine PDOS. The convolution of states corresponding to the HOMO manifold is identical to that of the empty wire of Fig. 25.6(b). The comparison between the HOMO states of the isolated G4 (Fig. 25.6(a)) and a plane-projection of the HOMO state of the K(I)-G4 wire (Fig. 25.6(c)), underlines the absence of metal-molecule electronic mixing, as well as of any other contribution of the metal cations to the frontier orbitals. Thus, we conclude that the presence of K(I) ions should not influence the motion of charges along the wire.

25.3.2 What are the effects of metal substitution?

We showed above that K(I) ions trapped in the inner core of G4-wires do not alter the nature of the guanine orbitals and the band structure in the energy range that could be interesting for transport applications. In fact, the K-projected DOS exhibits peaks only 12 eV below the HOMO and at lower energies. Hence, alkali metals are not good 'dopants' for these systems. Or, to put it another way, they are not a good way to manipulate the inherent electronic properties of nucleobase assemblies in a direction

useful to enhance conductivity. The spontaneous question is then: are other metals better performing? Does nature offer ideal candidates that form hybrids with the guanines, possibly yielding partially occupied manifolds? Are transition metals particularly suitable for this purpose, because of their multiple oxidation states and a natural redox activity [74–76] in a number of biological situations?

To answer these questions, we decided to try Ag and Cu. Whereas G4-wires are commonly reported in the presence of alkali ions, there is no known structure with Ag and Cu. Therefore, we had no reliable starting point. Yet, as a conceptual excercise, we tried Ag and Cu ions in the same locations that are stable for K(I) and in the same oxidation state. Although Ag and Cu can have the 2+ and 1+ formal charge states, we chose the latter because we wanted to fix them in the same site that is stable for a 1+ charged alkali ion. For the comparison between Cu(I)-G4-wires and Ag(I)-G4-wires, our main enquiry was: is there a metal better than another to form hybrid electron states with the guanines? If the answer were yes, we would then proceed with the selected metal for more thorough investigations. In such a spirit, the test calculations were not accurate and could be considered as a probing exercise: we just put Ag(I) and Cu(I) ions in the place of K(I), namely, between the tetrameric planes, and computed the DOS without inspecting the stable locations for the new ions. The thus computed DOSs are shown in Fig. 25.5(a) (middle and bottom). With both Ag and Cu the metal-projected DOS contributes peaks at much higher energy than K. Looking in more detail, only Cu ions couple with guanines forming mixed Cu–guanine orbitals, and contributes to the HOMO manifold around the energy origin. On the basis of this observation, we decided to enquire more deeply on Cu-trapping G4-wires.

The deeper analysis was still based on the conceptual hypothesis that Cu ions can be incorporated inside the channel, which is not proven experimentally to date. Since the most likely formal charge state for salt-derived Cu ions would be 2+, we switched to Cu(II) ions. Starting from the inter-plane sites, we relaxed the structures.

The first striking outcome is that the inter-plane site is not even metastable for Cu(II)[9]: upon free atomic relaxation (all mobile atoms), the ions migrate spontaneously away from their initial positions, ending up in the in-plane site, practically co-planar with each G4 tetrad. Similarly to what was explained in Section 25.3.1 for the K(I) ions, we also fixed Cu(I) ions at different locations along the helical axis and computed the total energy of the frozen geometries (see explanation in Section 25.3.1, and Fig. 25.4). The bottom panel in Fig. 4(b) clearly illustrates that the planar site is a local minimum for Cu(I), whereas the inter-plane site is a maximum. This is consistent with what we found by free relaxation for both Cu(I)- and Cu(II)-G4-wires.

We then computed the electronic structure of the Cu(II)-G4-wire for the optimized geometry with in-plane Cu ions, through spin-polarized calculations. The total (solid line) and Cu-projected (shaded area) DOS is shown in Fig. 25.5(b). Without entering the details, we underline the salient features. (i) Cu(II) ions mix efficiently with guanines, forming strong coordination complexes with the neighbouring electronegative O atoms. (ii) There are unpaired electrons in the system, which form the highest singly occupied molecular orbital (SOMO) manifold, which is the spin-down peak at zero energy in Fig. 25.5(b). The occupancy of this peak would be exactly zero if the charge state of the Cu

[9] Indeed, we verified that it is also not metastable for Cu(I).

ions were exactly the nominal 2+ state; indeed, the Cu–guanine mixing induces a reorganization of the charge and a partial occupation of the spin-down peak at zero energy. (iii) This peak is a three-fold multiplet constituted of electron orbitals with the shape of Fig. 25.6(d) and has a σ character. (iv) The spin-up counterpart is the peak at -1.25 eV, fully occupied; the intermediate peak has instead symmetric up and down occupation; the down–up energy splitting of the SOMO is due to the correlation energy cost when putting two electrons with opposite spins on the same localized orbital. The situation is very similar to that described in the literature for the oxidized form of the protein Azurin [74].

At the end of this metal substitution analysis, a clear conclusion does not emerge, but we wish to make three comments to take away.

1. Cu–G4-wires are not proven experimentally and this study remains as a conceptual excercise for the time being, even more so because the stable site for transition metals is more likely to occur outside the helix (attaching to the guanine N_7 atom) than inside.
2. The Cu–guanine mixing is very appealing on one hand because it indicates that such ions are able to 'dope' the system by altering the inherent electronic structure around the fundamental bandgap.
3. On the other hand, the Cu–guanine mixing induces the transition of the HOMO/SOMO from a π to a σ character, thus reducing the inter-plane coupling. Indeed, preliminary calculations indicate that the transfer integral for the Cu(II)-G4-wire decreases abruptly with respect to those for the empty G4-wire.

25.3.3 What are the effects of structural deformations?

Following a recent AFM characterization of novel long G4-wires deposited on mica [9], a possible axial contraction was proposed. The current level of characterization does not provide certain proof of the existence of such strained structures, nor would it be able to explain their origin, and DFT simulations cannot answer these questions because of a limited predictivity for non-covalent interactions (stacking). Yet, DFT can answer the question: what would be the effect of axial compression/extension and accompanying structural deformations on the bandwidths and bandgaps?

To answer this question, we have undertaken the computation of the bandstructure of G4-wires with an empty inner core and with variable stacking distance.[10] In addition, we have also considered empty G4-wires affected by other distortions, such as the variation of the twist angle [68].

We focus the discussion of our results on the most important band-structure parameters, namely, the fundamental HOMO-LUMO bandgap E_{g0} and the bandwidth of the HOMO manifold Δ_{HOMO}.

[10] This study was carried out with no metals inside the channel formed by the guanine tetrads. Whereas the ultimate band structure of real systems would be affected by the incorporation of metals as a synthesis or post-synthesis procedure, we are comforted in this choice by the fact that at least alkali ions do not form electronic hybrids with guanines. This evidence may change when the stacking distance changes, especially when transition metals are present, but here we isolate one effect, namely the variation of the $\pi - \pi$ inter-plane superposition following a variation of geometrical factors.

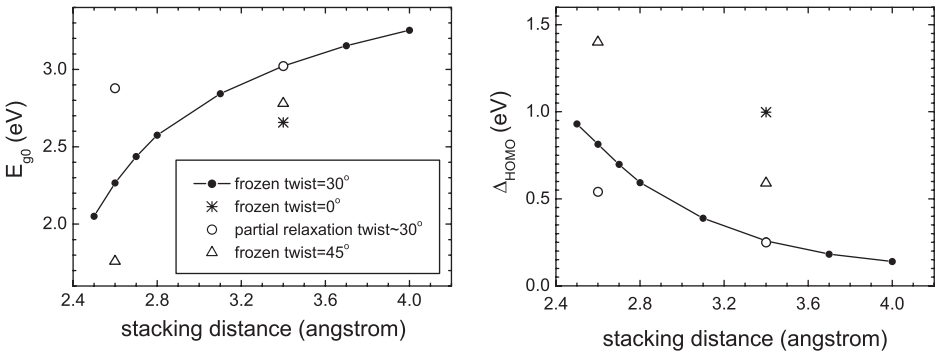


Fig. 25.7. Left: dependence of the fundamental bandgap Eg0 on the stacking distance and on other structural distortions. Right: similar plot for the HOMO bandwidth Δ_{HOMO}. The legend applies to both plots.

For all the examined distortions, the band structure maintains the characteristic of grouping into manifolds, each containing a number of bands equal to the number of guanines in the periodic building block, as explained above [11,15]. Given the definition of E_{g0} (E_{gap}) and Δ_{HOMO} in Fig. 25.5(a), the dependence of such parameters on the stacking distance is shown in Fig. 25.7.

The dots connected by segments (no interpolation) pertain to unrelaxed G4-wires in which all the atomic degrees-of-freedom were frozen in the equilibrium structure described above with a stacking distance of 3.37 Å and a twist angle of 30°, and the inter-plane separation was repeatedly changed between 2.5 Å and 4.0 Å. A saturation behaviour beyond 3.4 Å is clear. At small stacking distance a sharp rise of Δ_{HOMO} and drop of E_{go} is observed. Both these trends are advantageous from the transport viewpoint.

Being aware that it is very unlikely that real systems would remain frozen under strong axial compression, we also show the effect of other structural changes that may accompany such a compression, choosing only an unstrained G4-wire (3.4 Å) and a highly strained wire (2.6 Å) for this analysis. For instance, the circles correspond to G4-wires for which a full structural optimization is attempted within DFT, while keeping the periodicity unit to a fixed length. In the case of the uncompressed wire, the initial structure is already characterized by very small forces, and the structural optimization converges very quickly. In the case of the highly compressed wire, the initial configuration experiences strong forces and therefore the atoms start to deviate very sharply from their positions, showing a tendency to increase the twisting. However, a twisting different from 30 degrees does not respect the modulo-3 periodicity imposed on the system, and would require relaxation of the periodicity contraint: consequently, we cannot determine an optimal configuration for a stacking distance of 2.6 Å, and we represent in Fig. 25.7 the parameters of an intermediate very disordered snapshot. What we wish to remark for the circles is that the very smooth behaviour of the frozen wires can be drastically affected by structural flexibility, in particular in the direction opposite to what would be desired to improve transport: Δ_{HOMO} becomes smaller and E_{g0} becomes larger.

Since we detect a tendency of the compressed G4-wire at 2.6 Å to augment the twisting, we also show results (triangles) for a higher twist of 45°, obtained from frozen simulations with a modulo-2 periodicity. These findings show that even for an ideal stacking distance of

3.4 Å, it is possible to manoeuvre the band structure parameters in a direction useful for improving conductivity, e.g., increase Δ_{HOMO} and decrease E_{g0}, by controlling the rotational flexibility. This is further remarked by the asterisks in Fig. 25.7 that characterize an unstrained wire with eclipsed planes (twist = 0°).

To summarize this section, we note that the band structure parameters can be tuned by structural factors. Both axial compression and twist variation are able to enhance the π - π overlap and increase the bandwidth values. It is possible to show [68] that these results are not compromised by neglect of the backbone in the simulations.

25.4 RESULTS AND DISCUSSION ON THE CuHy-WIRE

Stimulated by the interesting experimental results [17], DFT calculations can probe the electronic structure of the system, inspecting the FM spin alignment but also comparing with the AM case. The additional inspiration for this work, as we noted in the introductory section, is the identification of self-assembling DNA-like wires in which metals can be incorporated in such a way to alter the frontier orbitals and therefore change dramatically the electronic structure, hopefully to improve a wire-like behaviour.

Zhang and co-workers [67] describe the structure and the strong coordination between the Cu(II) ion in a plane and the two surrounding artificial nucleobases. They also present an approximate analysis of the charge redistribution in each plane (relative to free bases) due to the ligand interaction: to this concern, we just note here that the computed Lowdin charge on Cu is smaller than expected in the Cu(II) formal oxidation state.

Then, attention is turned to the electronic characterization of the FM phase. The FM CuHy-wire has a σ-like spin density (Fig. 25.8), notwithstanding the fact that the HOMO of hydroxypyridone has a π character. This means that unpaired electrons in the

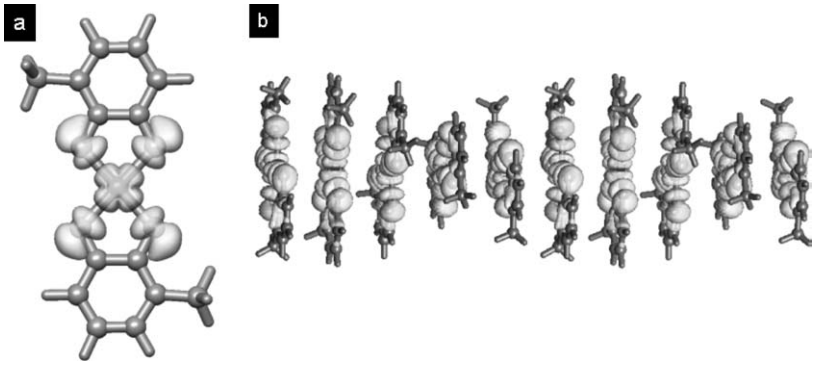


Fig. 25.8. (a) Ball and stick representation of the **H–**Cu(II)**–H** plane, with isosurface plot of the spin density. The light gray spheres identify C, Ñ and O atoms in **H** (compare with Fig. 25.2). The terminal H atoms are shown as simple sticks. The Cu ion is shown in dark grey. (b) Ball and stick representation of the [**H–**Cu(II)**–H**]$_{10}$ decameric building block of the periodic CuHy-wire, with isosurface plot of the wire spin-density in the FM state. The colour code is the same as in panel (a). The spin density is obtained by subtracting the minority-spin charge density from the majority-spin charge density.

system, whose orbitals are in principle the best candidates for a relatively high mobility, are mostly in-plane localized and quench the transfer integrals between adjacent pairs. The authors note, however, that the global HOMO remains π -like, but is doubly occupied: the spin-down peak m_H in Fig. 25.9 has its spin-up companion M_H, which is also occupied and could be exploited for transport only under doping. Furthermore, we note that these peaks m_H and M_H have some degree of Cu**–H** coupling, as desired. Last, at lower energy (1.2 eV below the HOMO) a continuous orbital delocalized through the axis, and with a significant bandwidth, appears. Although the latter cannot be easily excited by the application of an electric field, it may have an important role for mobile charges. The calculations were performed without imposing the helical symmetry: hence, as explained in Section 25.1.3 and described for the G4-wires, the bands are grouped into manifolds with multiplicities determined by the number of molecular units in the periodic building block (each spin-up or spin-down peak is a 10-fold multiplet in this case).

Turning to the AM system, unfortunately DFT calculations do not allow one to assess the relative FM–AM stability on the basis of the energetics. However, one can conclude that both phases are viable. There is no up-down spin splitting in the DOS: all energy levels are doubly occupied, and are represented in Fig. 25.9 by vertical lines. The HOMO manifold is π -like, whereas the first occupied σ orbital occurs ~ 0.25 eV below the HOMO peak and has a shape similar to the spin-density of the FM wire.

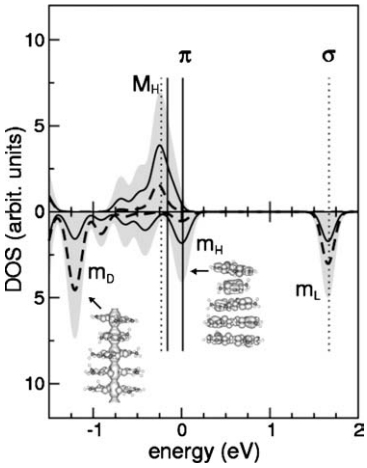


Fig. 25.9. Total (shadowed area), Cu(II)-projected (dashed line), and **H**-projected (solid line) density of states for the FM CuHy-wire. The zero of the energy scale is set at the centre of the highest occupied peak. The upper (lower) plot pertains to the majority-spin (minority-spin) electrons. The m_H minority-spin HOMO is π -like (rightmost inset), and its majority-spin counterpart M_H is also fully occupied: thus, these states do not contribute to the σ spin density. A continuous axial manifold, mostly localized on the metals along the axis, appears at 1.2 eV below the HOMO (leftmost inset). The vertical lines mark the centres of the corresponding DOS peaks for the AM CuHy-wire. Solid lines refer to π states similar to m_H, dashed lines refer to σ states similar to the spin-density of the FM configuration.

It is possible to draw some important general conclusions from this work. Evidently, Cu(II) ions inserted in mimetic base pairs can significantly alter the frontier orbitals, which indirectly means that the route of metal incorporation can be successfully pursued towards a user-designed band-structure behaviour: this is in line with the arguments presented above on the exploration of the effects of various metals inside G4-wires; the effects of Cu are indeed very similar in the CuHy- and G4-wires. Secondly, this particular system can be prepared in a high-spin state: despite the fact that the DFT analysis cannot conclude in favour of the FM phase, it is compatible with it, and supports even more the experimental evidence.

25.5 PERSPECTIVES ON OTHER PROMISING STRATEGIES FOR STRUCTURAL MANIPULATION OF DNA: METAL COMPLEXATION AND AROMATIC EXPANSION

Before summarizing, we briefly sketch our view of other possible modifications of the natural Watson–Crick helix that may enable the use of DNA-related materials in nanotechnology applications (with a particular expectation for molecular electronics). Our view is based on experimental hints and is two-fold.

1. *Towards metallized structures.* In a relatively recent experiment, Rakitin and co-workers [18] claimed that Zn^{2+} ions can substitute for hydrogen-bonding in Watson–Crick pairs, giving as a result a metal–insulator transition in DNA double helices. The 'metal-doped' DNA was called M-DNA. While this experiment has not yet been reproduced by any group, either in the synthesis or in the electrical aspects, it stimulated much discussion and several activities to pursue metal doping in DNA molecules. Given the absence of convincing structural determination, DFT calculations are hindered. However, preliminary calculations were started on metallized GC pairs with various metals (see sketch in Fig. 25.10(a)) [12] and may be continued in the future to predict stacking arrangements. This task is indeed very delicate, and some efforts have already yielded negative outcomes, e.g., the impossibility of obtaining a DFT structure for M-DNA. However, refinements in the experiments and in the theory may prove useful.

   It is also interesting to explore metal complexation with G-quartets. We already said at the end of Section 25.3.2 that transition metal ions added after the synthesis of G4-wires are rather unlikely to penetrate the channel. However, they can find

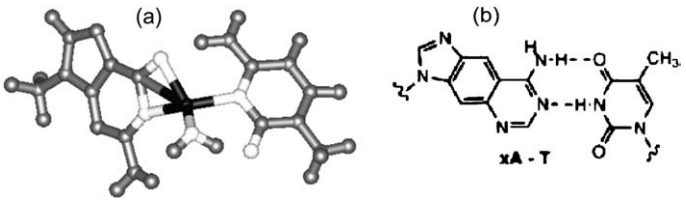


Fig. 25.10. (a) Scheme for an M-DNA modified GC pair. (b) Scheme for an aromatic expanded xAT pair.

accomodation outside the helix and change the electronic structure in a profitable manner.

2. *Aromatic expansion.* The group of Eric Kool has been recently very active in designing and realizing chemical procedures for the alteration of the nucleobases [16,26,27]. The idea of inserting an aromatic ring in the planar structure is indeed very intriguing: in fact, aromatic rings are usually known to facilitate electronic delocalization, both in-plane and out-of-plane. Thus, this particular protocol for nucleobase alteration may strongly enhance the $\pi - \pi$ superposition along double-stranded helices. Stimulated by this idea, we carried out preliminary calculations of inter-plane transfer integrals for natural AT pairs and artificial xAT pairs (see sketch in Fig. 25.10(b)), inspecting the cases where the aromatic expansion is limited to a single strand or done in opposite strands. The transfer integrals were computed with a recently developed technique [77] suitable for interfacing with DFT plane-wave codes; such quantities are equivalent to those obtained by other authors for stacks of native Watson–Crick pairs [35]. We found an increase by up to one order of magnitude for artificial xAT pairs relative to natural AT pairs, which is very promising in the pursuit of charge motion. Expanded GC pairs may even perform much better, and we believe that DFT calculations for such expanded helices with all four altered nucleotides may reveal interesting electronic effects.

In addition to the above perspectives towards novel conformational changes, we finally remark that the potentiality of DFT simulations in relation to DNA-based electronics will greatly benefit from theoretical developments on using the ground-state band structure in non-equilibrium transport formalisms, that will finally allow for *ab initio* simulation of measured current–voltage curves. This goal is rather far away and envisaged in long-term plans (10–15 years), at least for application to a very complex system such as a DNA polymer. However, open avenues accessible in the short-term (3–5 years) are also rather promising to gain insight into the behaviour of mobile charges in DNA fragments, in closer contact with experimental efforts. To mention just the main possibilities that we are pursuing: on one hand, *ab initio* band-structure parameters can be plugged into tight-binding models to compute mesoscopic transport quantities [78]; on the other hand, the software optimization of fully *ab initio* codes for equilibrium quantum transport[11] [79], accompanied by an increase in computer power, may soon enable their application to a simple DNA-based nano-junction, to yield a direct theory–experiment comparison such as that given by Cuniberti and co-workers [78], limited to the linear regime but in a fully *ab initio* fashion.

25.6 SUMMARY

Artificial self-assembled nucleobase systems with native or artificial bases, possibly with a modified helical motif and with incorporated metals, which approximately maintain the

[11] http://www.wannier-transport.org/

nano-size diameter of native DNA but most notably recognition and structure, may be advantageous relative to native DNA for electronic applications. In this chapter, two types of such candidates were presented from the point of view of DFT calculations. The two main messages that we wish to convey are: (i) metals can be effective to the purpose of changing and controlling electronic properties (orbital shape, spin alignment, redox behaviour); (ii) G4 quadruple helices behave like native DNA as 'wide-bandgap semiconductors' with bandwidths and bandgaps highly tunable by interaction with metals and with the environment (e.g., flexibility). Last but not least: although significant limitations must always be taken into account, DFT along the years is proving itself as a valuable tool to describe the electronic structure of nucleotide aggregates and enquire on their possible applications for transporting currents.

25.7 ACKNOWLEDGEMENTS

The results presented in this chapter were enabled by a long-standing collaboration with Anna Garbesi and Elisa Molinari and by their endless suggestions and encouragement. More recently, HouYu Zhang and Manuela Cavallari participated in some aspects of the illustrated research. We also benefited from collaborations and discussions with Simone Alexandre, José Soler, Danny Porath, Joshua Jortner, Shlomo Yitzchaik, Sasha Kotlyar, Julio Gomez, Benjamin Dwir, Eli Kapon, Giovanni Cuniberti, Angel Rubio and Daniele Varsano. The work was funded by the EC under contracts IST-2001-38951 and HPRN-CT-2002-00317, by the Italian Ministry for University and Research (MIUR) under contract FIRB-NOMADE, and by INFM through the Parallel Computing Committee for the allocation of computer time on the supercomputing facilities at CINECA, Bologna, Italy.

25.8 REFERENCES

1 C. Joachim and M.A. Ratner, Proc. Natl. Acad. USA, 102 (2005) 8801.
2 D. Porath, G. Cuniberti and R. Di Felice, Topics Curr. Chem., 237 (2004) 183.
3 R.G. Endres, D.L. Cox and R.R.P. Singh, Rev. Mod. Phys., 76 (2004) 195.
4 M. Di Ventra and M. Zwolak, DNA electronics, in S.H. Nalwa (Ed.), Encyclopaedia of nanoscience and nanotechnology, American Scientific Publishers, CA, 2 (2003) 475.
5 A.Y. Kasumov, D.V. Klinov, P.E. Roche, S. Guéron S and H. Bouchiat, Appl. Phys. Lett., 84 (2004) 1007.
6 H. Cohen, C. Nogues, R. Naaman and D. Porath, Proc. Natl. Acad. Sci. USA, 102 (2005) 11589.
7 B. Xu, P. Zhang, X. Li and N. Tao, Nano Lett., 4 (2004) 1105.
8 J.T. Davis, Angew. Chem. Int. Ed., 43 (2004) 668.
9 A.B. Kotlyar, N. Borovok, T. Molotsky, H. Cohen, E. Shapir and D. Porath, Adv. Mater, 17 (2005) 1901.
10 T. Muir, E. Morales, J. Root, I. Kumar, B. Garcia, C. Vellandi, T. Marsh, E. Henderson and J. Vesenka, J. Vac. Sci. Technol. A, 16 (1998) 1172.
11 A. Calzolari, R. Di Felice, E. Molinari and A. Garbesi, J. Phys. Chem. B, 108 (2004) 2509; ibid, 13058.
12 R. Di Felice, A. Calzolari and H. Zhang, Nanotechnology, 15 (2004) 1256.
13 R. Di Felice, A. Calzolari, E. Molinari and A. Garbesi, Phys. Rev. B, 65 (2002) 045104.
14 G. Gottarelli, G.P. Spada and A. Garbesi, in J.L. Atwood, J.E.D. Davies, D. MacNicol and F. Vögtle (Eds.), Comprehensive supramolecular chemistry, Pergamon, Oxford, 9 (1996) 483–506.
15 A. Calzolari, R. Di Felice, E. Molinari and A. Garbesi, Appl. Phys. Lett., 80 (2002) 3331.
16 E.T. Kool, Acc. Chem. Res., 35 (2002) 936.

17 K. Tanaka, A. Tengeiji, T. Kato, N. Toyama and M. Shionoya, Science, 299 (2003) 1212.
18 A. Rakitin, P. Aich, C. Papadopoulos, Y. Kobzar, A.S. Vedeneev, J.S. Lee and J.M. Xu, Phys. Rev. Lett., 86 (2001) 3670.
19 M. Brucale, G.P. Zuccheri and B. Samorí, Org. Biomol. Chem., 3 (2005) 575.
20 B. Yurke, A.J. Turberfield, A.P. Mills Jr, F.C. Simmel and J.L. Newmann, Nature, 406 (2000) 605.
21 C. Mao, W. Sun, Z. Shen and N.C. Seeman, Nature, 397 (1999) 144.
22 N.C. Seeman, Angew. Chem. Int. Ed., 37 (1998) 3220.
23 N.C. Seeman, Nature, 421 (2003) 427.
24 C. Mao, W. Sun, Z. Shen and N.C. Seeman, Nature, 407 (2000) 493.
25 Y. Benenson, T. Paz-Elizur, R. Adar, E. Keinan, Z. Livneh and E. Shapiro, Nature, 414 (2001) 430.
26 H. Liu, J. Gao, S.R. Lynch, Y.D. Saito, L. Maynard and E.T. Kool, Science, 302 (2003) 868.
27 J. Gao, H. Liu and E.T. Kool, Angew. Chemie Int. Ed., 44 (2005) 3118.
28 L. Zhang, A. Peritz and E. Meggers, J. Am. Chem. Soc., 127 (2005) 4174.
29 H. Weizman and Y. Tor, J. Am. Chem. Soc., 123 (2001) 3375.
30 G.H. Clever, K. Polborn and T. Carell, Angew. Chem. Int. Ed., 44 (2005) 7204.
31 T. Helgaker, P. Jørgensen and J. Olsen, Molecular electronic-structure theory, Wiley, Chichester, 2000.
32 J. Šponer, J. Leszczynski and P. Hobza, J. Phys. Chem., 100 (1996) 1965.
33 J. Šponer, J. Leszczynski and P. Hobza, J. Phys. Chem., 100 (1996) 5590.
34 J.V. Burda, J. Šponer, J. Leszczynski and P. Hobza, J. Phys. Chem. B, 101 (1997) 9670.
35 N.R. Nösch and A.A. Voityuk, Topics Curr. Chem., 237 (2004) 37.
36 R.M. Dreizler and E.K.U. Gross, Density functional theory, an approach to the quantum many body problem, Springer, Berlin, Heidelberg, New York, 1990.
37 C. Fiolhais, F. Nogueira and M. Marques (Eds.), A primer in density functional theory, Lecture notes in physics 620, Springer, Berlin, 2003.
38 R.A. Friesner and B.D. Dunietz, Acc. Chem. Res., 34 (2001) 351.
39 E. Starikov, J. Photochem. Photobiol., 3 (2002) 147.
40 A. Nitzan, Annu. Rev. Phys. Chem., 52 (2001) 681.
41 A. Nitzan, Isr. J. Chem., 42 (2002) 163.
42 J. Jortner, A. Nitzan and M.A. Ratner, in G. Cuniberti, G. Fagas and K. Richter (Eds.), Introducing molecular electronics, Lecture notes in physics, Springer-Verlag, Berlin, 2005, in press.
43 A. Ferretti, A. Calzolari, R. Di Felice, F. Manghi, M.J. Caldas, M. Buongiorno Nardelli and E. Molinari, Phys. Rev. Lett., 94 (2005) 116802.
44 A. Ferretti, A. Calzolari, R. Di Felice and F. Manghi, Phys. Rev. B, 72 (2005) 125114.
45 A. Pecchia, A. Di Carlo, A. Gagliardi, S. Sanna, T. Frauenheim and R. Gutierrez, Nano Lett., 4 (2004) 2109.
46 R. Gutierrez, S. Mandal and G. Cuniberti, Nano Lett., 5 (2005) 1093.
47 F.L. Gervasio, P. Carloni and M. Parrinello, Phys. Rev. Lett., 89 (2002) 108102.
48 F.L. Gervasio, A. Laio, M. Parrinello and M. Boero, Phys. Rev. Lett., 94 (2005) 158103.
49 P.J. de Pablo, F. Moreno-Herrero, J. Colchero, J. Gómez-Herrero, P. Herrero, A.M. Baró, P. Ordejón, J.M. Soler and E. Artacho, Phys. Rev. Lett., 85 (2000) 4992.
50 S.S. Alexandre, E. Artacho, J.M. Soler and H. Chacham, Phys. Rev. Lett., 91 (2003) 108105.
51 E. Artacho, M. Machado, D. Sánchez-Portal, P. Ordejón and J.M. Soler, Mol. Phys., 101 (2003) 1587.
52 A. Höbsch, R.G. Endres, D.L. Cox and R.R.P. Singh, Phys. Rev. Lett., 94 (2005) 178102.
53 H. Wang, J.P. Lewis and O.F. Sankey, Phys. Rev. Lett., 93 (2004) 016401.
54 M. Hjort and S. Stafström, Phys. Rev. Lett., 87 (2001) 228101.
55 R.N. Barnett, C.L. Cleveland, A. Joy, U. Landmann and G.B. Schuster, Science, 294 (2001) 567.
56 C. Adessi, S. Walch and M.P. Anantram, Phys. Rev. B, 67 (2003) R081405.
57 H. Mehrez and M.P. Anantram, Phys. Rev. B, 71 (2005) 115405.
58 M. Bixon, B. Giese, S. Wessely, T. Langenbacher, M.E. Michel-Beyerle and J. Jortner, Proc. Natl. Acad. Sci. USA, 96 (1999) 11713.
59 J. Jortner, M. Bixon, T. Langenbacher and M.E. Michel-Beyerle, Proc. Natl. Acad. Sci. USA, 95 (1998) 12759.
60 S. Baroni, S. de Gironcoli, A. Dal Corso and P. Giannozzi, Plane-wave self-consistent field, http://www.pwscf.org.

61 D. Vanderbilt, Phys. Rev. B, 41 (1990) 7892.
62 N.R. Kestner and J.E. Combariza, Rev. in Comp. Chem., 13 (1999) 99.
63 T. van Mourik, A.K. Wilson, K.A. Peterson, D.E. Woon and T.H. Dunning Jr, Adv. Quant. Chem., 31 (1999) 105.
64 J.P. Perdew, J.A. Chevary, S.H. Vosko, K.A. Jackson, M.R. Pederson, D.J. Singh and C. Fiolhais, Phys. Rev. B, 46 (1992) 6671.
65 A. Calzolari, R. Di Felice, E. Molinari and A. Garbesi, Physica E, 13 (2002) 1236.
66 A. Calzolari, R. Di Felice and E. Molinari, Solid State Commun., 131 (2004) 557.
67 H.Y. Zhang, A. Calzolari and R. Di Felice, J. Phys. Chem. B, 109 (2005) 15345.
68 R. Di Felice, A. Calzolari, A. Garbesi, S.S. Alexandre and J.M. Soler, J. Phys. Chem. B, 109 (2005) 22301.
69 K. Phillips, Z. Dauter, A.I.H. Murchie, D.M.J. Lilley and B. Luisi, J. Mol. Biol., 273 (1997) 171; X-ray structure: NBD ID UDF062.
70 G.N. Parkinson, M.P.N. Lee and S. Neidle, Nature, 417 (2002) 876.
71 M. Cavallari, A. Calzolari, R. Di Felice and A. Garbesi, unpublished (2005).
72 N. Spackova, I. Berger and J. Sponer, J. Am. Chem. Soc., 121 (1999) 5519.
73 A. Calzolari, N. Marzari, I. Souza and M. Buongiorno Nardelli, Phys. Rev. B, 69 (2004) 035108.
74 S. Corni, F. De Rienzo, R. Di Felice and E. Molinari, Int. J. Quantum. Chem., 102 (2005) 328.
72 S. Corni, J. Phys. Chem. B, 109 (2005) 3423.
76 G. Maruccio, A. Biasco, P. Visconti, A. Bramanti, P.P. Pompa, F. Calabi, R. Cingolani, R. Rinaldi, S. Corni, R. Di Felice, E. Molinari, M.P. Verbeet and G.W. Canters, Adv. Mater., 17 (2005) 816.
77 A. Ferretti, A. Ruini, G. Bussi, E. Molinari and M.J. Caldas, Phys. Rev. B, 69 (2004) 205205.
78 G. Cuniberti, L. Craco, D. Porath and C. Dekker, Phys. Rev. B, 65 (2002) R241314.
79 A. Calzolari, I. Souza, N. Marzari and M. Buongiorno Nardelli, Phys. Rev. B, 69 (2004) 035108.

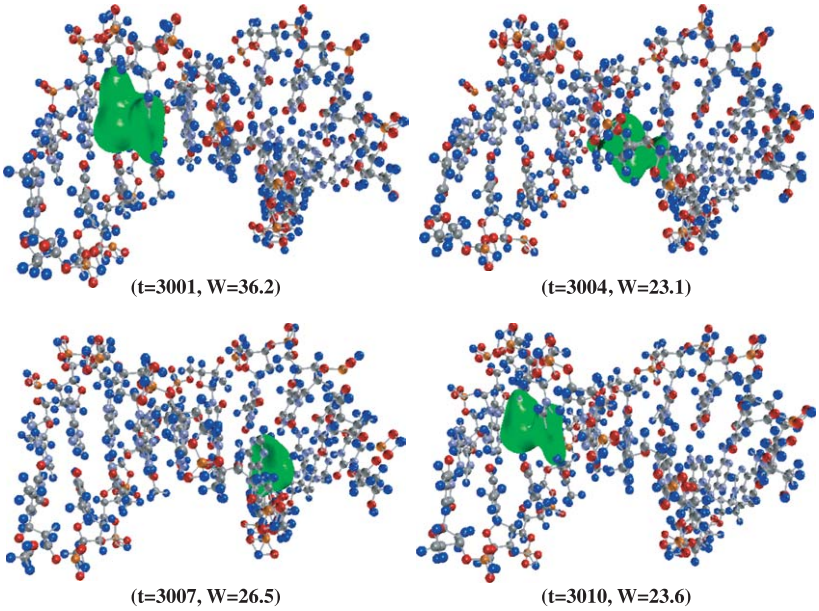


Plate 32. Population density plots for the localised HOMO state as a function of time. The time between snapshots is 1.5 ps (illustration appears on page 417 of this volume).

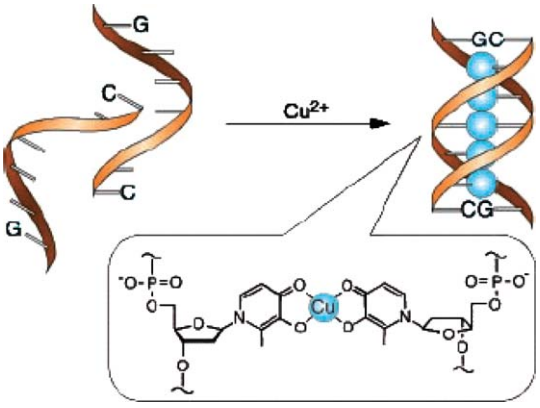


Plate 33. The figure shows the self-assembly process of artificial DNA strands in which some of the native nucleobases (except the terminal ones) can be substituted by hydroxypyridone bases: if Cu(II) ions are present in solution during synthesis, they can be incorporated along the axis and form molecules with a magnetic moment. The inset shows the artificial H–Cu(II)–H pair: the Cu(II) ion in each plane is tightly coordinated by four surrounding electronegative O atoms (illustration appears on page 488 of this volume). (From [17], with permission; copyright ©2003 Science.)

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 26

Embedding method for conductance studies of large molecules

Owen R. Davies* and John E. Inglesfield†

School of Physics and Astronomy, Cardiff University,
Cardiff, CF24 3YB, UK

Abstract

Many large biological molecules like DNA can be split up naturally into sections, and this enables the electronic structure to be calculated piece by piece. Each piece can be 'embedded' on to the section on either side, enabling the electronic structure and transmission to be calculated efficiently, in a time proportional to the size of the molecule. Moreover, the molecule can be embedded on to metal contacts on either side. In this chapter we describe the methods, and present results for different forms of DNA.

26.1 INTRODUCTION

Understanding the electronic structure of biological molecules is fundamental to understanding their properties and behaviour – electronic structure determines the structure and bonding, reactivity, and in the context of this chapter, the electrical conductance. Perhaps the most studied molecule, and certainly one of the most remarkable, is DNA. With its stable double–helix structure and self-assembly properties, DNA has been proposed for numerous practical applications, from DNA computers [1], to biosensors for the pharmaceutical industry [2], and as the building blocks of a nanoscale construction kit [3]. The idea of using DNA as a conductor is quite old, and already in 1962 Eley and Spivey [4] suggested that DNA could transfer electrical charge via π-bonding between consecutive bases in this linear molecule. Debate about the conductance of DNA, and charge transfer down the molecule (not quite the same as conductance) [5], has continued for many years, and there is a wide range of experimental results. Some experiments have shown that DNA is an insulator [6], while others suggest that it can be a wide-band gap semiconductor [7], or a good linear conductor [8]. Charge transfer in DNA is likely to have important biological functions, for example, in oxidizing damage to DNA [9].

* Funded by an EPSRC Studentship
† Corresponding author, J.Inglesfield@astro.cf.ac.uk

In order to understand and resolve the conflicting experimental results, theorists have developed many ways of calculating transport properties of DNA, based on a variety of electronic structure methods. These include tight-binding model Hamiltonians [10], semi-empirical methods such as extended Hückel theory [11–13], which we use in this chapter, and density-functional theory [14] – certainly the most accurate. The effects of structure, base-pair order, temperature and environment are all important, and can be studied theoretically to try to understand the electronic properties of the DNA molecule. It is clear from contradictory reports in the literature that much more work is needed, particularly to study the relationship between structure and conductance.

This chapter describes a way of calculating the electronic structure and electrical conductance of linear molecules like DNA, which is very efficient [13,15]. We shall describe the methodology, and apply it to DNA. The method is based on building up the linear molecule from small units, adding one unit at a time to the system. The electronic structure is calculated within each unit, and this is 'embedded' on to the rest of the system so that the electronic wavefunctions match correctly. The advantage of this approach is that the computer time needed for the calculation of the electronic structure is proportional to the length of the molecule, which means that very large molecules can be studied. Our solution of Schrödinger's equation is in the framework of one-electron theory, so that each electron is moving in the self-consistent field of the others – this means that we are not able to tackle many-body phenomena, such as quantum blockade, where the presence of an electron in the process of conduction inhibits the movement of others, and polaron effects [16,17], where the electron motion is coupled to atomic displacements. However, one-electron theory is invariably the starting point for a more complete approach. The DNA calculations we present here are in a localized orbital representation – in fact we use extended Hückel theory. Our methods can be extended to a fully self-consistent density-functional calculation, and there is no restriction on basis functions – plane waves (or related basis functions like linearized augmented plane waves, LAPWs) can be used, as we shall discuss in the next section.

In this chapter we use atomic units (a.u.), with $e = \hbar = m = 1$; we also use electron-volts for energy, with 1 a.u. = 27.2 eV. The atomic unit of length is the Bohr radius, 0.5292 Å.

26.2 THE EMBEDDING METHOD

26.2.1 The embedding potential

Embedding was developed as a way of solving the one-electron Schrödinger equation in systems that can be split into smaller units [18,19]. An embedding potential, added on to the Hamiltonian for part of the system, allows the Schrödinger equation to be solved for just this part, with the wavefunctions correctly matched onto the surroundings. The method was originally used to calculate the electronic structure of surfaces and interfaces, using an LAPW basis set (LAPWs are essentially plane waves, with atomic solutions matched on to the plane waves inside the atomic cores) [20]. This has been developed into highly accurate self-consistent codes, which provide the electronic structure of a surface (say) [21], by solving the Schrödinger equation explicitly only in the top

layer or two of atoms; the embedding potential, which is an energy-dependent complex potential defined over the interface with the substrate, takes care of matching on to the substrate wavefunctions exactly.

Let us suppose we want to solve the Schrödinger equation in region I, to be embedded on to a substrate region II over boundary S. As a trial function we use an arbitrary function ϕ in region I (ϕ is the function which we shall subsequently optimize), and in region II an exact solution of the Schrödinger equation ψ at a trial energy ε, such that ψ matches in amplitude onto ϕ over S. Substituting into the variational principle, the expectation value of the Hamiltonian H is given by

$$E = \frac{\int_{\mathrm{I}} \mathrm{d}\mathbf{r}\phi^* H\phi + \varepsilon\int_{\mathrm{II}} \mathrm{d}\mathbf{r}\psi^*\psi + \tfrac{1}{2}\int_S \mathrm{d}\mathbf{r}_s\phi^*\left(\frac{\partial\phi}{\partial n_s} - \frac{\partial\psi}{\partial n_s}\right)}{\int_{\mathrm{I}} \mathrm{d}\mathbf{r}\phi^*\phi + \int_{\mathrm{II}} \mathrm{d}\mathbf{r}\psi^*\psi} \tag{1}$$

The first term in the numerator is the expectation value of the Hamiltonian through region I, the region of interest; the second term is the contribution from region II. The final term in the numerator, an integral over the boundary S, contains the difference in normal derivatives on either side of S (measured outwards from I) and comes from the kinetic energy operator in H.

We now introduce G_0, the Green function in region II satisfying the zero normal derivative boundary condition over S. It can be shown using Green's theorem that the surface inverse of G_0 behaves like a generalized logarithmic derivative, relating the amplitude and normal derivative over S of the solution of the Schrödinger equation in II at energy ε [18],

$$\frac{\partial\psi(\mathbf{r}_s)}{\partial n_s} = -2\int_s \mathrm{d}\mathbf{r}'_s G_0^{-1}(\mathbf{r}_s,\mathbf{r}'_s;\varepsilon)\psi(\mathbf{r}'_s) \tag{2}$$

with

$$\int_s \mathrm{d}\mathbf{r}''_s G_0^{-1}(\mathbf{r}_s,\mathbf{r}''_s)G_0(\mathbf{r}''_s,\mathbf{r}'_s) = \delta(\mathbf{r}_s - \mathbf{r}'_s) \tag{3}$$

Following from this, the ψ terms in (1) can be replaced by integrals over interface S involving G_0^{-1}. The variational expression then becomes [18]

$$E = \frac{\int_{\mathrm{I}} \mathrm{d}\mathbf{r}\phi^* H\phi + \tfrac{1}{2}\int_S \mathrm{d}\mathbf{r}_s\phi^*\frac{\partial\phi}{\partial n_s} + \int_S \mathrm{d}\mathbf{r}_s\int_S \mathrm{d}\mathbf{r}'_s\phi^*\left(G_0^{-1} - \varepsilon\frac{\partial G_0^{-1}}{\partial\varepsilon}\right)\phi}{\int_{\mathrm{I}} \mathrm{d}\mathbf{r}\phi^*\phi - \int_S \mathrm{d}\mathbf{r}_s\int_S \mathrm{d}\mathbf{r}'_s\phi^*\frac{\partial G_0^{-1}}{\partial\varepsilon}\phi} \tag{4}$$

This is a variational expression for region I, embedded on to II by G_0^{-1}. G_0^{-1} is the *embedding potential* [18], which we shall subsequently designate as Σ,

$$\Sigma(\mathbf{r}_s,\mathbf{r}'_s) = G_0^{-1}(\mathbf{r}_s,\mathbf{r}'_s) \tag{5}$$

To find ϕ, we may expand it in terms of a basis set $\chi_i(\mathbf{r})$,

$$\phi(\mathbf{r}) = \sum_i a_i \chi_i(\mathbf{r}) \tag{6}$$

and then varying the coefficients to minimize E in (4) we obtain the matrix eigenvalue form of the Schrödinger equation for I embedded into II,

$$\sum_j H_{ij}^{\text{emb}} a_j = E \sum_j S_{ij} a_j \tag{7}$$

The embedded Hamiltonian matrix elements are given by

$$H_{ij}^{\text{emb}} = \int_{\text{I}} \mathrm{d}\mathbf{r} \chi_i^*(\mathbf{r}) \left[-\frac{1}{2}\nabla^2 + V(\mathbf{r}) \right] \chi_j(\mathbf{r}) + \frac{1}{2} \int_S \mathrm{d}\mathbf{r}_s \chi_i^*(\mathbf{r}_s) \frac{\partial \chi_j(\mathbf{r}_s)}{\partial n_s} \tag{8}$$

$$+ \int_S \mathrm{d}\mathbf{r}_s \int_S \mathrm{d}\mathbf{r}'_s \, \chi_i^*(\mathbf{r}_s) \left[\Sigma(\mathbf{r}_s, \mathbf{r}'_s; \varepsilon) + (E - \varepsilon) \frac{\partial \Sigma(\mathbf{r}_s, \mathbf{r}'_s)}{\partial \varepsilon} \right] \chi_j(\mathbf{r}'_s)$$

and S_{ij} is the overlap matrix

$$S_{ij} = \int_{\text{I}} \mathrm{d}\mathbf{r} \chi_i^*(\mathbf{r}) \chi_j(\mathbf{r}) \tag{9}$$

The structure of the embedding method is revealed very clearly in (8). The matrix elements of H^{emb} are evaluated only through region I, and the second term on the right-hand side of (8), involving the normal derivative, ensures hermiticity. The final double integral in H_{ij}^{emb} is the matrix element of Σ evaluated at a trial energy ε, with a first-order correction to give Σ at the correct energy E. Solving (7), with convergence in basis-set size, and iterating so that the trial energy $\varepsilon = E$, gives the wavefunction ϕ which matches in both amplitude and derivative on to the solution of the Schrödinger equation in the substrate. In practical applications, the Green function $G(\mathbf{r},\mathbf{r}';E)$ is evaluated rather than the individual wavefunctions. Expanding this in region I in terms of the basis functions,

$$\mathrm{G}(\mathbf{r},\mathbf{r}';E) = \sum_{ij} G_{ij}(E) \chi_i(\mathbf{r}) \chi_j^*(\mathbf{r}') \tag{10}$$

the Green function matrix satisfies

$$\sum_k (H_{ik}^{\text{emb}} - E S_{ik}) G_{kj} = \delta_{ij} \tag{11}$$

As we specify the energy E in the Green function, the embedding potential is evaluated at this energy, and the energy derivative term in (8) disappears.

The embedding potential is a *self-energy* – a term added to the Hamiltonian to take account of a region of space not treated explicitly. In fact, the self-energy is very frequently referred to in the context of molecule–metal contacts [22], but it is just the same as our embedding potential: the molecule is *embedded* on to the contact. We have seen

that this embedding potential is given by the surface inverse of a Green function (3). It can also be found, without taking a surface inverse, from the zero-amplitude Green function [23]

$$\Sigma(\mathbf{r}_s,\mathbf{r}'_s) = -\frac{1}{4}\frac{\partial^2 \hat{G}}{\partial n_s \partial n'_s}(\mathbf{r}_s,\mathbf{r}'_s) \tag{12}$$

This is the embedding potential in the continuum description–we shall study in Section 26.3.1 the form of the embedding potential with a linear combination of atomic orbitals (LCAO) Hamiltonian.

This embedding formalism has been used to calculate the electronic structure of ideal and defective surfaces, interfaces between semi-infinite crystals, screening at surfaces in electric fields, and so on [19]. Surface electronic structure is very relevant to conductance of large molecules, because the molecule–metal contact is important for contact resistance – electrons in conductance experiments are scattered at the molecule–metal interface. It is useful to consider the local density of states, the spatial distribution of states with energy E,

$$\sigma(\mathbf{r},E) = \sum_i |\phi_i(\mathbf{r})|^2 \delta(E-E_i) \tag{13}$$

where ϕ_i is a wavefunction of the system with energy E_i. This can be found from the Green function, using

$$G(\mathbf{r},\mathbf{r}';E)=\sum_i \frac{\phi_i(\mathbf{r})\phi_i^*(\mathbf{r}')}{E_i-E} \tag{14}$$

hence

$$\sigma(\mathbf{r}, E)= \frac{1}{\pi}\Im \mathrm{m} G(\mathbf{r},\mathbf{r};E+i\varepsilon) \tag{15}$$

where a positive infinitesimal ε is added to the energy. As an example, Fig. 26.1 shows the surface density of states calculated using the embedding method at the (111) surface of Au. The surface density of states in Fig. 26.1 is the local density of states integrated through the surface atoms, and it shows the metal d bands on a smooth background coming from the s and p electrons.

26.2.2 Sub-volume embedding

The methods we use for calculating the electronic structure and conductance of large molecules are based on sub-volume embedding, devised by Crampin *et al.* [24]. Sub-volume embedding can be used to calculate the electronic structure of systems which can be divided up into layers, such as metallic interfaces, but it has an obvious application to linear molecules.

Consider a substrate in the left half-space, with a boundary at the surface S_l–this may be the left-hand bulk metal in an interface calculation, or the left-hand contact in a metal–molecule–metal study (Fig. 26.2). This can be replaced by an embedding potential Σ_l on S_l. We now consider the Schrödinger equation for the atoms between S_l and the

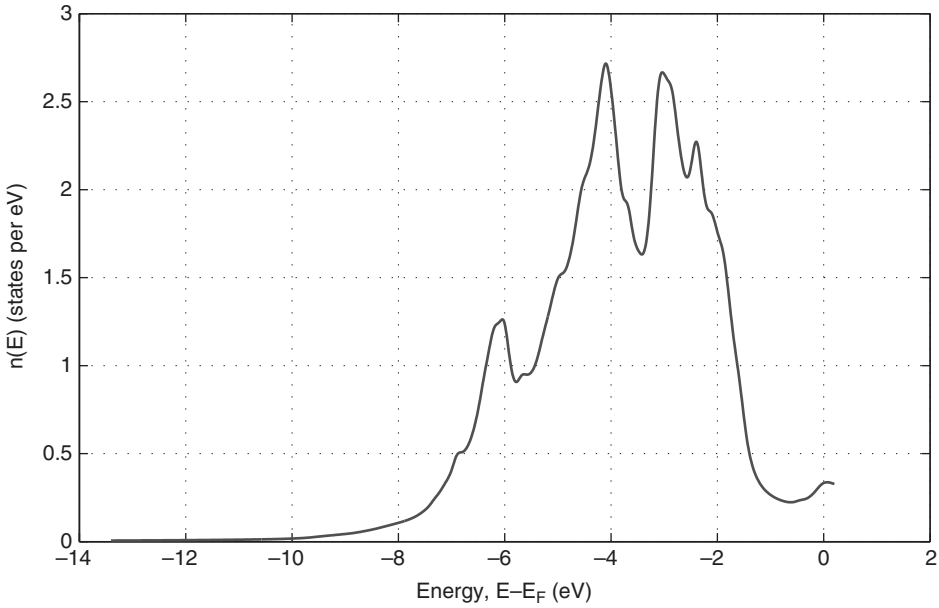


Fig. 26.1. Surface density of states on Au(111), with an energy broadening of 0.005 a.u. (calculated by I. Merrick).

surface S_{l+1} to the right, and find the Green function for this sub-volume, adding Σ_l over the left-hand boundary to the Hamiltonian, and another, presently unspecified, embedding potential Σ over the right-hand boundary. The Green function then satisfies the Schrödinger equation for the sub-volume plus left-hand substrate, together with the boundary condition corresponding to Σ. If we take $\Sigma = 0$, this boundary condition gives G_0 with zero normal derivative on the boundary at S_{l+1}, from (2). The inverse of G_0 over this boundary is $G_0^{-1} = \Sigma_{l+1}$, the embedding potential for a system to the right of S_{l+1}, matching it on the left to the sub-volume plus substrate.

We can continue this procedure to the right, adding sub-volumes in a linear manner, and generating at each stage a new left-hand embedding potential until we reach the final sub-volume between S_{r-1} and S_r (Fig. 26.2). Now the Green function, which we finally require within each sub-volume, has to match both the left and right-hand substrates, which means finding an appropriate right embedding potential as well as a left embedding potential for each sub-volume. So we must also start off at the right-hand side of the system, at the substrate interface S_r, generating a succession of right-hand embedding potentials as we proceed sub-volume by sub-volume to the left. Having a right and left embedding potential for each sub-volume, we can then solve the Schrödinger equation within the sub-volume once again to find the fully matched Green function. This gives us a method for solving the Schrödinger equation for any system which can be broken down into adjacent sub-volumes, scaling linearly with their number. Moreover, we have complete freedom of choice for the basis functions within each sub-volume. This has been developed into a fully self-consistent program using LAPWs

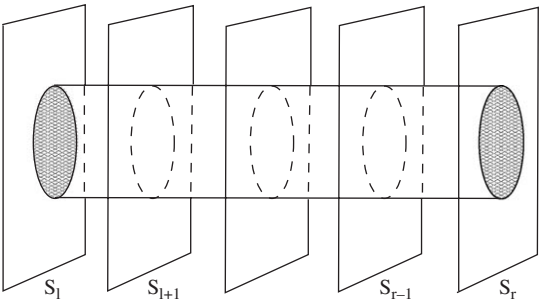


Fig. 26.2. The system embedded on to left and right substrates at S_l and S_r is divided into sub-volumes, and embedding potentials found on intermediate planes, $S_{l+1}...S_{r-1}$

for surface and interface applications [25], and we shall see shortly how it can be used for DNA (say) between metal contacts, in the LCAO version of the method.

26.2.3 Transmission and the embedding potential

The transmission through a region such as the 'molecule' depicted in Fig. 26.3, between left and right-hand substrates, can be expressed in terms of the embedding potentials for contacting the substrates. There is a well-known result [26] that the total transmission through a nanostructure is given by

$$T = 4\mathrm{Tr}[G_{lr}\Im\mathrm{m}\Sigma_r G_{rl}^{*}\Im\mathrm{m}\Sigma_l] \tag{16}$$

where Σ_l, Σ_r are the embedding potentials which embed the nanostructure on to the left- and right-hand substrates, respectively, and G_{lr} is the Green function between the embedding surfaces S_l and S_r (Fig. 26.3). The trace is over the quantum numbers of the conduction channels in each substrate. This formula is mainly used in a localized orbital representation [27,28], and Σ_{llr} called a self-energy, but Wortmann *et al.* [29] recently derived it in the context of embedding, and used it to study the transmission of metallic interfaces, and field emission from surfaces.

There are several ways of deriving (16) [28,29]. We shall start off from the scattering result [30] that if wave $\phi(\mathbf{r})$ is incident on the left contact from the bulk (Fig. 26.3), then the full wavefunction everywhere to the right is given by

$$\psi(\mathbf{r}) = 2i\int_{S_l} \mathrm{d}\mathbf{r}_s \int_{S_l} \mathrm{d}\mathbf{r}_s' G(\mathbf{r},\mathbf{r}_s)\Im\mathrm{m}\Sigma_l(\mathbf{r}_s,\mathbf{r}_s')\phi(\mathbf{r}_s') \tag{17}$$

where G is the Green function for the whole system, that is, the molecule or other nanostructure embedded on to the contacts on either side.

We now define scattering channels crossing the left-hand boundary S_l into the molecule in terms of the embedding potential over S_l. We take the channel functions to be the eigenfunctions of $\Im\mathrm{m}\Sigma_l$ at each energy, satisfying

$$\int_S \mathrm{d}\mathbf{r}_s'\Im\mathrm{m}\Sigma_l(\mathbf{r}_s,\mathbf{r}_s')\psi_i(\mathbf{r}_s') = \lambda_i\psi_i(\mathbf{r}_s) \tag{18}$$

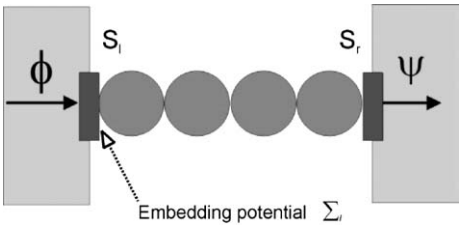


Fig. 26.3. The molecule is embedded on to metal contacts over S_l and S_r. ϕ is the incident wave from the left-hand contact, and ψ is the full wavefunction everywhere to the right of S_l.

(Alternatively, channels may be defined as normal modes of the *transmission* of the molecule [31].) Although this gives the channel functions only over S_l, this uniquely defines the corresponding wavefunctions throughout the whole of the left-hand substrate. It is convenient to take the incident amplitude $\phi(\mathbf{r}_s)$ in (17) to be one of these left-hand channel functions, say $\phi_i(\mathbf{r}_s)$. The scattered wave $\psi(\mathbf{r})$ is also projected onto one of the right-hand channel functions $\phi_j(\mathbf{r}_s)$, an eigenfunction of the right-hand embedding potential. Substituting into (17), we then have the transmission amplitude between these channel functions given by

$$t_{ji} = 2i\int_{S_r} \mathrm{d}\mathbf{r}_s \int_{S_l} \mathrm{d}\mathbf{r}_s' \int_{S_l} \mathrm{d}\mathbf{r}_s'' \phi_j(\mathbf{r}_s) G(\mathbf{r}_s,\mathbf{r}_s') \Im\mathrm{m}\Sigma_l(\mathbf{r}_s',\mathbf{r}_s'') \phi_i(\mathbf{r}_s'') \tag{19}$$

As $\Im\mathrm{m}\Sigma_l$ is diagonal in the basis of the left-hand channel functions, this can be written symbolically as

$$t_{ji} = 2iG_{rl}(j,i)\Im\mathrm{m}\Sigma_l(i,i) \tag{20}$$

Similarly, the transmission amplitude between the right-hand and left-hand channels is given by

$$t_{ij} = 2iG_{lr}(i,j)\Im\mathrm{m}\Sigma_r(j,j) \tag{21}$$

Taking into account the flux in each channel, the transmission probability from channel i on the left to channel j on the right, and vice versa, may be written as

$$T_{ji} = t_{ji}t_{ij}^* = 4G_{rl}(j,i)\Im\mathrm{m}\Sigma_l(i,i)G_{lr}^*(i,j)\Im\mathrm{m}\Sigma_r(j,j) \tag{22}$$

Summing over all channels, $T = \Sigma_{i,j}T_{i,j}$, we obtain (16) for the total transmission through the molecule.

A very interesting point about formula (22) is that the transmission is between the asymptotic systems on either side, in other words between the substrates, rather than the transmission between the embedding boundaries S_l and S_r [29]. This means that we can choose *any* boundaries on which to determine the embedding potentials and Green function, and we should always obtain the same T. As an illustration of this point, we follow Wortmann *et al.* [29] and consider transmission across a one-dimensional

potential step at $z = 0$, such that the wave-vector is k_1 in the region to the left of the boundary, and k_2 to the right. If we bring S_l and S_r together at a point $z < 0$, the Green function G_{lr} is given by

$$G_{lr} = \frac{2ie^{-ik_1z}}{k_1}\left(\frac{k_1\cos k_1z + ik_2\sin k_1z}{k_1 + k_2}\right) \tag{23}$$

and the left and right embedding potentials by

$$\Sigma_l = -\frac{ik_1}{2},\ \Sigma_r = -\frac{ik_1}{2}\left(\frac{k_2\cos k_1z + ik_1\sin k_1z}{k_1\cos k_1z + ik_2\sin k_1z}\right) \tag{24}$$

Substituting into (16) we obtain

$$T = \frac{4k_1k_2}{(k_1+k_2)^2} \tag{25}$$

which is indeed the correct transmission across the potential step. But note that we have obtained this with S_l and S_r at the same point, on the same side of the potential step!

The consequence of this remarkable result is that to obtain the transmission between the substrates, hence the conductance, in the sub-volume embedding method we can apply (16) at any *single* intermediate embedding plane. We have seen that at each of these planes, labelled $S_{l+1} \cdots S_{r-1}$ in Fig. 26.2, we find both a left and a right embedding potential, and with the Green function on this plane this enables us to find T.

If we wish, we may still choose S_l and S_r in (16) to be the embedding surfaces on the left and right substrates, i.e., the same as S_l and S_r in Fig. 26.2. Then we have to find G_{lr} going from one sub-volume to another, whereas sub-volume embedding as described in Section 26.2.2 gives the Green function only within each sub-volume. In fact, Green function methods can be used to build up such a Green function from the separate Green functions. Suppose we wish to find $G(1,4)$, the Green function between surface 1 in sub-volume A and surface 4 in C (Fig. 26.4). We first find Green functions for each sub-volume, satisfying the following boundary conditions:

- G_A satisfies the boundary condition that it is embedded over 1 on to the appropriate substrate – the rest of the molecule, or the metal contact – and over 2 it has zero normal derivative;
- G_B has zero normal derivative over 2 and 3;
- G_C like G_A, is embedded on to the substrate over 4, and has zero normal derivative on 3.

As shown by García-Moliner and his co-workers [32], the matching Green function method can be used to find the Green function connecting surfaces 1 and 4 in the elegant form,

$$G(1,4) = G_A(1,2)\,[G_A(2,2)+G_B(2,2)]^{-1}G_B(2,3) \tag{26}$$

$$\times\,[G_B(3,3)+G_C(3,3)]^{-1}G_C(3,4)$$

In this equation, like in (3) and (16), the Green functions are expanded over the surfaces in terms of two-dimensional basis sets, and the inverses in (26) imply matrix inversion.

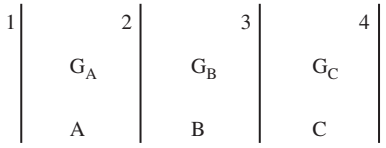


Fig. 26.4. Sub-volumes A, B and C are joined together over surfaces 2 and 3, and the matching Green function method is used to find $G(1,4)$.

The procedure given in (26) can be extended to give the Green function propagating across as many sub-volumes as we need.

26.3 LOCALIZED ORBITALS AND EMBEDDING

We now turn to applying sub-volume embedding to systems described by localized orbital basis functions, with a tight-binding or LCAO Hamiltonian. The formalism of Section 26.2 is based on functions of the continuous spatial variable, and although localized basis functions can be used to expand such functions, this may be dangerous, as the embedding and matching Green function techniques we have described depend on an accurate representation of Green functions over surfaces. Localized orbitals may not have the variational freedom to describe these surface properties accurately, and it is best to use plane-wave-type bases if we are using the methods of Section 26.2. In this section we shall see how equivalent formulae to those developed above can be found for localized orbitals in the framework of LCAO.

26.3.1 The embedding potential as a self-energy

We first consider the LCAO version of embedding [33], working with the Hamiltonian and overlap matrices evaluated with localized basis functions. The system is separated into two groups of orbitals, and the Hamiltonian and overlap matrices written in block matrix form as

$$H = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}, \; S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \tag{27}$$

H_{11} and H_{22} are the sub-matrices of the Hamiltonian within each group; the interaction between the orbital groups is described by H_{12}. Similarly, S_{11} and S_{22} are the overlaps within groups 1 and 2 individually, and S_{12} is the overlap between the orbitals in 1 and 2. We can think of the orbitals in each group as being centred in region I and II, respectively. The Green function for the whole system can be similarly written in the same block matrix form as in (27),

$$G = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} \tag{28}$$

Our aim is to find G_{11} in terms of H_{11} and S_{11}, plus an embedding term to take care of the rest of H and S.

To do this, we consider the block-diagonal Green function G^0 given by inverting $(H-ES)$ without the coupling terms H_{12} and S_{12},

$$G^0 = \begin{pmatrix} G_{11}^0 & 0 \\ 0 & G_{22}^0 \end{pmatrix} = \begin{pmatrix} H_{11}-S_{11} & 0 \\ 0 & H_{22}-S_{22} \end{pmatrix}^{-1} \tag{29}$$

To go from G^0 to G we include the off-diagonal terms as a perturbation δ,

$$\delta = \begin{pmatrix} 0 & H_{12}-ES_{12} \\ H_{21}-ES_{21} & 0 \end{pmatrix} \tag{30}$$

using Dyson's equation

$$G = G^0 - G^0 \delta G \tag{31}$$

(The minus sign in (31) comes from our convention that G satisfies the equation $(H-ES)G=1$.) Expanding (31) gives

$$G = G^0 - G^0 \delta G^0 + G^0 \delta G^0 \delta G \tag{32}$$

and we take the matrix elements of this equation between group 1 orbitals to give

$$G_{11} = G_{11}^0 + G_{11}^0 \delta_{12} G_{22}^0 \delta_{21} G_{11} \tag{33}$$

Comparing (33) with (31), we see that we can treat the perturbation on G_{11}^0 to be $-\delta_{12} G_{22}^0 \delta_{21}$, with matrix elements only between the group 1 orbitals. This is the embedding potential,

$$\Sigma_{11} = -(H_{12}-ES_{12}) G_{22}^0 (H_{21}-ES_{21}) \tag{34}$$

which embeds the Hamiltonian in the space of group 1 orbitals on to the rest of the system. By adding this term on to the Hamiltonian of the group 1 orbitals we obtain an effective Hamiltonian,

$$H_{11}^{\text{eff}} = H_{11} - (H_{12}-ES_{12}) G_{22}^0 (H_{21}-ES_{21}) \tag{35}$$

from which we can find G_{11}

$$G_{11} = (H_{11}^{\text{eff}} - ES_{11})^{-1} \tag{36}$$

working entirely within the group 1 orbitals.

This embedding potential within the LCAO framework was derived by Baraff and Schlüter [33], who were interested in the relationship between embedding and Dyson's equation. The self-energy [22], which has been used for many years in studies of tunnelling in the scanning tunnelling microscope [34] and in molecular transport [28], to include the effect of the metal contacts, is exactly the same as the embedding potential.

26.3.2 Localized orbitals and sub-volume embedding

The LCAO embedding potential can be used to build up the Green function in a molecule using the sub-volume embedding method described in Section 26.2.2 [13]. We shall illustrate this by considering three sections of a linear molecule such as DNA, as shown in Fig. 26.5.

We start off by calculating the Green function G^0_{11} in section 1 by itself, from which we can find the embedding potential for embedding section 2 on to 1 (34),

$$\Sigma^2_l = -(H_{21} - ES_{21})G^0_{11}(H_{12} - ES_{12}) \tag{37}$$

Our notation is that Σ^2_l indicates the embedding potential for section 2 (acting entirely in the space of the orbitals in 2), embedding on to the left-hand side of this section. This is added on to the Hamiltonian for section 2 to give an effective Hamiltonian from which the Green function for this section embedded to the left, $\widetilde{G}_{22}$, can be calculated. We now continue to build up the chain from left to right, so that the left-hand embedding potential for section 3 is given by

$$\Sigma^3_l = -(H_{32} - ES_{32})\widetilde{G}_{22}(H_{23} - ES_{23}) \tag{38}$$

Using $\widetilde{G}_{22}$ in (38) means that the effects of all the previous sections to the left of section 3 are included in this embedding potential.

We now repeat the process, but this time building up the molecule from right to left, so that

$$\Sigma^2_r = -(H_{23} - ES_{23})G^0_{33}(H_{32} - ES_{32}) \tag{39}$$

and

$$\Sigma^1_r = -(H_{12} - ES_{12})\overline{G}_{22}(H_{21} - ES_{21}) \tag{40}$$

where $\overline{G}$ is the Green function of a section embedded on the right. Once we have finished adding sub-units to the left, we have the left and right embedding potentials for all

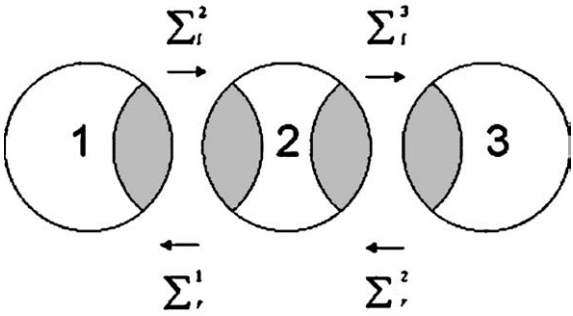


Fig. 26.5. Schematic representation showing three sections of DNA, with embedding potentials Σ_l and Σ_r which embed to the left and right.

sections. We then simply add these embedding potentials to the unperturbed Hamiltonian for each section, and in this way we are able to calculate the Green function for each section of the molecule, no matter where it lies in the chain, with the effect of all other sections in the molecule taken into account. So, in this example we have

$$G_{22} = (H_{22}+\Sigma_l^2+\Sigma_r^2 - ES_{22})^{-1} \tag{41}$$

for the Green function of the middle section.

This method can be applied directly to any molecule with a linear sequence, with the assumption that only neighbouring sections have direct orbital overlap. This method of 'growing' molecules is related to methods of diagonalizing Hamiltonians with tridiagonal block structure [35], and to the recursive Green function algorithm used by Lake *et al.* [36] in their studies of transport through layered semiconductor structures. In Section 26.3.4 we describe how we embed metal contacts on to the molecule in the same way. Of course, many biological molecules, such as proteins, though having linear structure, are folded back on themselves. This adds extra interactions between sections that are not nearest neighbours, which cannot be treated in this approach.

26.3.3 Density of states

We must be careful when calculating the density of states of the molecule to take account of the overlap between orbitals. Showing how we do this in the sub-volume embedding method illustrates how we can constructively use the properties of Green functions, in particular Dyson's equation. The density of states of the molecule $n(E)$ can be found from the local density of states (15) by integrating over the whole system,

$$n(E)=\frac{1}{\pi}\int \mathrm{d}\,\mathbf{r}\,\Im\mathrm{m}G(\mathbf{r},\mathbf{r};E+i\varepsilon) \tag{42}$$

Expanding G in terms of the basis functions, which are used to determine the Hamiltonian and overlap matrix elements in (27), we have

$$G(\mathbf{r},\mathbf{r}';E) = \sum_{ij} G_{ij}(E)\chi_i(\mathbf{r})\chi_j(\mathbf{r}') \tag{43}$$

and substituting into (42) gives the density of states as

$$n(E)=\frac{1}{\pi}\Im\mathrm{m}\mathrm{Tr}(GS) \tag{44}$$

The trace of this product matrix can be written as

$$\mathrm{Tr}(GS) = \sum_{(n,i)(m,j)} G_{(n,i)(m,j)} S_{(m,j)(n,i)} \tag{45}$$

where n and m label the sections of the molecule (as in Fig. 26.5), with i and j labelling the orbitals in n and m respectively. However, there are only contributions to the sum when $m = n, n-1$, or, $n+1$ due to the short range of the overlap. Thus we can rewrite the trace as

$$\mathrm{Tr}(GS) = \sum_{(n,i)}\left(\sum_{(n,j)} G_{(n,i)(n,j)} S_{(n,j)(n,i)} + \sum_{(n-1,j)} G_{(n,i)(n-1,j)} S_{(n-1,j)(n,i)} + \sum_{(n+1,j)} G_{(n,i)(n+1,j)} S_{(n+1,j)(n,i)}\right) \tag{46}$$

We find the first term in (46) directly from our embedding procedure – we have seen how we obtain the Green function between all orbitals within the same section from (41). The third term involves $G_{(n,i)(n+1,j)}$, which is the Green function between orbitals in one section with those in the next section. This can be found by treating the link between sections n and n+1 as the perturbation δ in Dyson's equation (31),

$$\delta_{(n,i)(n+1,j)} = H_{(n,i)(n+1,j)} - ES_{(n,i)(n+1,j)} \tag{47}$$

The unperturbed system then consists of section n embedded to the left and n+1 embedded to the right, and from Dyson's equation we have

$$G_{(n,i)(n+1,j)} = -\sum_{k,l} \tilde{G}_{(n,i)(n,k)} \delta_{(n,k)(n+1,l)} G_{(n+1,l)(n+1,j)} \tag{48}$$

where as before, $\tilde{G}$ implies the Green function embedded only on the left. Both Green functions on the right-hand side of (48) have already been determined. The second term in (46) counts the contribution between the current section and the previous. However these same contributions have already been calculated when $m = n$+1, therefore they can be included by multiplying the third term by a factor of two.

In the sub-volume embedding method with continuous Green functions, described in Section 26.2.2, the basis functions are used only *within* each section, so that the total density of states is simply calculated by integrating (15) through each sub-volume and summing over sub-volumes–the Green function going from one section to its neighbours does not contribute.

26.3.4 Transmission through the molecule

The transmission through the molecule can be calculated from the formula (16), which we discussed in Section 26.2.3. We remind ourselves that T is the transmission between the substrates, and includes the contact resistances at the metal–molecule junctions.

In Section 26.2.3 we saw that the embedding potentials in the transmission formula need not be at the molecule–substrate contacts, but can be defined anywhere in the molecule. However, we have not yet explored the consequences of this in the localized orbital representation, and in our DNA work we have taken Σ_l and Σ_r in (16) to be the embedding potentials linking the molecule to the left- and right-hand contacts. This is in line with previous molecular work, in which the contact self-energies are used in the transmission formula.

We first find the Green function which propagates an electron between the contacting orbitals in the molecule. To emphasize that this Green function is for the molecule embedded on to the contacts, let us refer to it by G^{con}. This can be found using, once again, Dyson's equation,

$$G^{\text{con}}_{lr} = G_{lr} - G_{ll}\Sigma_l G^{\text{con}}_{lr} - G_{lr}\Sigma_r G^{\text{con}}_{rr} \tag{49}$$

in which the embedding potentials Σ_l and Σ_r linking the left and right metal reservoirs to the molecule are treated as perturbations on the Green function for the free molecule G.

This equation couples $G_{lr}^{\rm con}$ to $G_{rr}^{\rm con}$, for which Dyson's equation gives us a second coupled equation

$$G_{rr}^{\rm con} = G_{rr} - G_{rl}\Sigma_l G_{lr}^{\rm con} - G_{rr}\Sigma_r G_{rr}^{\rm con} \tag{50}$$

We already know G_{ll} and G_{rr}– these are the Green functions for the left-hand and right-hand contacting orbitals, respectively, in the free molecule. Each of these Green functions involves orbitals in the same section of the molecule.

To find G_{lr}, which couples the left-hand and right-hand contacting orbitals when these are in different sections, we extend the method given by (48), which gives the Green function between orbitals in neighbouring sections in terms of known Green functions in the same section. Suppose the left-hand contact is at section 1, and the right-hand contact at section 4, as in Fig. 26.6. The first step is to find $\tilde{G}_{12}$, the Green function between section 1 and 2 embedded to the left with the links between 2 and 3 broken,

$$\tilde{G}_{12} = -\tilde{G}_{11}\,\delta_{12}\,\tilde{G}_{22} \tag{51}$$

Here, as before, $\tilde{G}_{11}$ is the Green function for section 1 embedded to the left and uncoupled to 2, and $\tilde{G}_{22}$ is similarly left-embedded but uncoupled to 3. Continuing this process, and with the same notation, we have

$$\tilde{G}_{13} = -\tilde{G}_{12}\,\delta_{23}\,\tilde{G}_{33} \tag{52}$$

The final step is to go to section 4, which is coupled to the rest of the molecule,

$$G_{14} = -\tilde{G}_{13}\,\delta_{34}\,G_{44} \tag{53}$$

so the generalization of (48) is given by

$$G_{14} = -\tilde{G}_{11}\,\delta_{12}\,\tilde{G}_{22}\delta_{23}\tilde{G}_{33}\delta_{34}G_{44} \tag{54}$$

In this way we can find G_{lr} in terms of Green functions within each section, which we have already determined. This expression is the LCAO analogue to the matching Green function expression (26). Substituting into (49) and (50), we can solve these coupled equations to obtain the Green function linking the contacting orbitals in the presence of the contacts,

$$G_{lr}^{\rm con} = [1+G_{ll}\Sigma_l - G_{lr}\Sigma_r(1+G_{rr}\Sigma_r)^{-1}G_{rl}\Sigma_l]^{-1}\times[G_{lr} - G_{lr}\Sigma_r(1+G_{rr}\Sigma_r)^{-1}G_{rr}] \tag{55}$$

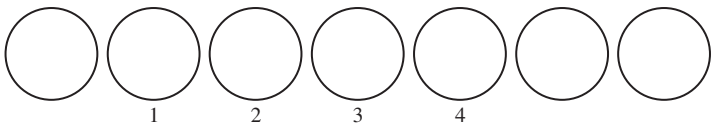


Fig. 26.6. Schematic representation of a molecule in which orbitals in section 1 are coupled to the left-hand contact, and those in section 4 to the right-hand contact.

We now turn to evaluating the embedding potentials $\Sigma_{l/r}$ in (55). Accurate embedding potentials for a molecule coupled to a metal surface have not yet been calculated in the continuum description (Section 26.2.1), because of the difficulty of finding a suitable representation of $G(\mathbf{r},\mathbf{r}')$ in the molecule–surface geometry. The best that has been done so far is to treat a small molecule embedded on to jellium [37]. However, we can go further in the LCAO methods we are considering here. Some workers have used a small cluster of atoms to represent the metal contact [12] and to derive the corresponding embedding potential or self-energy, but we prefer to use a semi-infinite metal reservoir [28,31], and use our knowledge of the surface density of states on the atoms at the surface (Fig. 26.1) to build up the embedding potential.

The calculations we present in this chapter all involve contact between a single atom on the molecule with a single atom on the metal surface. The simplest model is to consider a single *s*-orbital $\phi_s(r)$ on the metal surface atom, projecting the surface density of states (as shown in Fig. 26.1) on to this orbital. From this, the imaginary part of the Green function on the metal atom is given by

$$\Im\mathrm{m}G_s^{\mathrm{metal}}(E) = \pi n_s(E) \tag{56}$$

The real part of the Green function can then be found from the Kramers–Kronig relation, and is given by

$$\Re\mathrm{e}G_s^{\mathrm{metal}}(E)=\int_{-\infty}^{+\infty}\mathrm{d}E'\,\frac{n_s(E')}{E'-E} \tag{57}$$

The embedding potential may then be obtained from (34). We also use the *s*, *p* and *d* components of the surface density of states to give us a more accurate representation of the surface atom Green function, writing

$$G^{\mathrm{metal}}(E)(r,r')=\sum_L G_L^0\phi_L(r)\phi_L(r') \tag{58}$$

where *L* is the angular momentum. This expression ignores cross-terms coming from the lack of spherical symmetry on the surface atom.

This method of obtaining the embedding potential for embedding the molecule on to the metal contact can certainly be improved–there is no obstacle to including the off-diagonal components of the surface Green function in (58), and a Green function involving more than one surface atom can be found quite easily, for multi-atom contact. Up to now we have not mentioned self-consistency – the way that the electrons respond to the electron–electron interaction. This will certainly be important at the metal–molecule contact, and if we want the transmission to include the full contact resistance, it is necessary to include those metal atoms whose potential is affected by the contact in the molecule part of the calculation [31].

26.4 EMBEDDING STUDIES OF DNA

DNA is the linear molecule *par excellence*–it divides naturally into sub-volumes for our Green function embedding method, each section consisting of a base pair with the

associated sugars and phosphates in the backbone. As is discussed elsewhere in this volume, there have been numerous, and frequently conflicting, experimental measurements of the conductance of DNA. The present consensus, supported by calculations, seems to be that poly(G)–poly(C) DNA is a semiconductor with a bandgap of several electron-volts [7,14]. Let us see how our embedding calculations contribute to these DNA studies.

26.4.1 Extended Hückel theory

We use the LCAO formalism described in Section 26.3, using extended Hückel theory for the Hamiltonian matrix elements. This method, which we choose because of its simplicity, was originally developed for calculating the electronic structure of organic molecules [38]. It has recently been used to calculate the Hamiltonian matrix elements for calculations of the electrical conductance through molecular wires, with success in describing experimental results [11].

Extended Hückel theory is a minimum basis set method, in which an orthonormal atomic orbital basis set is chosen for each atom I. The diagonal Hamiltonian matrix elements are given by the empirical valence shell ionization potentials. The non-diagonal elements between two atomic orbitals α,β centred on atoms I,J (where $I \neq J$) are approximated by setting them proportional to the corresponding overlap, $S_{\alpha I,\beta J}$, of the two atomic orbitals, where

$$H_{\alpha I,\beta J} = K_{\alpha I,\beta J} S_{\alpha I,\beta J} \tag{59}$$

$K_{\alpha I,\beta J}$ only depends on the on-site energies, $E_{\alpha I}$, and $E_{\beta J}$. There are different prescriptions for representing $K_{\alpha I,\beta J}$. Following Cerdá [39], we use

$$K_{\alpha I,\beta J} = K_{EHT} \frac{E_{\alpha I}+E_{\beta J}}{2} \tag{60}$$

where we take the dimensionless constant $K_{EHT} = 1.75$.

In calculating the overlap integrals, each atomic orbital is approximated by a single Slater-type orbital, given by a product of a radial wavefunction and a spherical harmonic,

$$\chi(r,\theta,\phi) = R_{N,\eta}(r) Y_{l,m}(\theta,\phi) \tag{61}$$

where

$$R_{N,\eta}(r) = \frac{1}{\sqrt{(2N)!}} (2\eta)^{N+\frac{1}{2}} \mathrm{e}^{-\eta r} r^{N-1} \tag{62}$$

In this formula, which includes the normalization, N is the principle quantum number, and η is the Slater exponent. For d-orbitals the radial part of the wave-function is expressed in terms of two Slater functions with different Slater exponents. The parameters which we use are tabulated elsewhere [13]. In our applications of extended Hückel theory, we apply a cut-off of 8 a.u. to the overlap integrals, so that there is significant overlap only between neighbouring sections in our application to DNA.

26.4.2 Density of states

We are now ready to do the actual calculations, and we begin with the density of states of DNA. We shall give results for two different forms of DNA, both of which are 12 base-pairs long. The first is DNA with the base-pair sequence (CGTAGATCTACG) (only one base in each pair is listed, as the complementary bases are given by C (cytosine)–G (guanine) and A (adenine) – T (thymine)). The spatial coordinates were obtained from X-ray diffraction experiments [40], but as this does not detect hydrogen, the computer program ViewerLite [41] was used to add H atoms to the dangling bonds of the structure. This structure is quite spatially disordered, and elsewhere [13] we consider the effects of optimizing its structural energy using the AMBER force field [42] within the HyperChem molecular dynamics computer package [43]. The second molecule we investigate is poly(G)–poly(C) DNA, which has guanine bases on one strand with the complementary cytosine bases on the other strand (Fig. 26.7). This molecule is particularly interesting, since both experiment [7] and theory [14] suggest that this sequence of base pairs gives the highest conductivity. Again we use HyperChem to simulate the molecule and determine its structure.

We now apply our embedding method, as described in Section 26.3.3, to calculate the density of states of DNA. The results for poly(G)-poly(C) DNA are shown in Fig. 26.8, the densities of states for the other structures being very similar. The very peaky structure of the density of states has been slightly broadened by adding a small imaginary part of 0.005 a.u. to the energy in (42). Taking the highest occupied molecular orbital (HOMO) at -0.435 a.u. and the lowest unoccupied molecular orbital (LUMO) at -0.327 a.u. gives a bandgap for DNA of 0.108 a.u., 2.9 eV. Most literature values vary between 1.12 eV and 3.2 eV [14,44], so we can be quite satisfied with this result (but

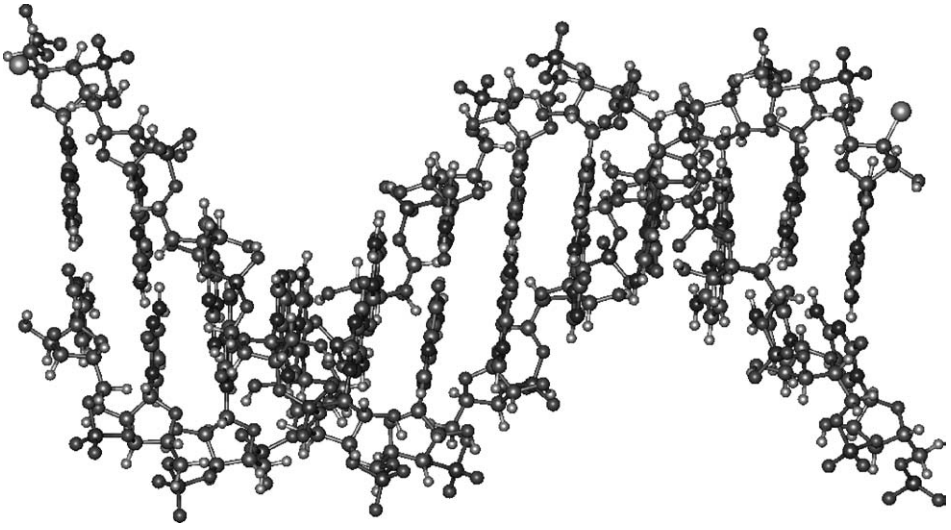


Fig. 26.7. 12 base-pair poly(G)-poly(C) DNA molecule. The double ring guanine bases lie consecutively along the same backbone strand, with the complementary single ring cytosine bases along the opposite strand (see Colour Plate 34) .

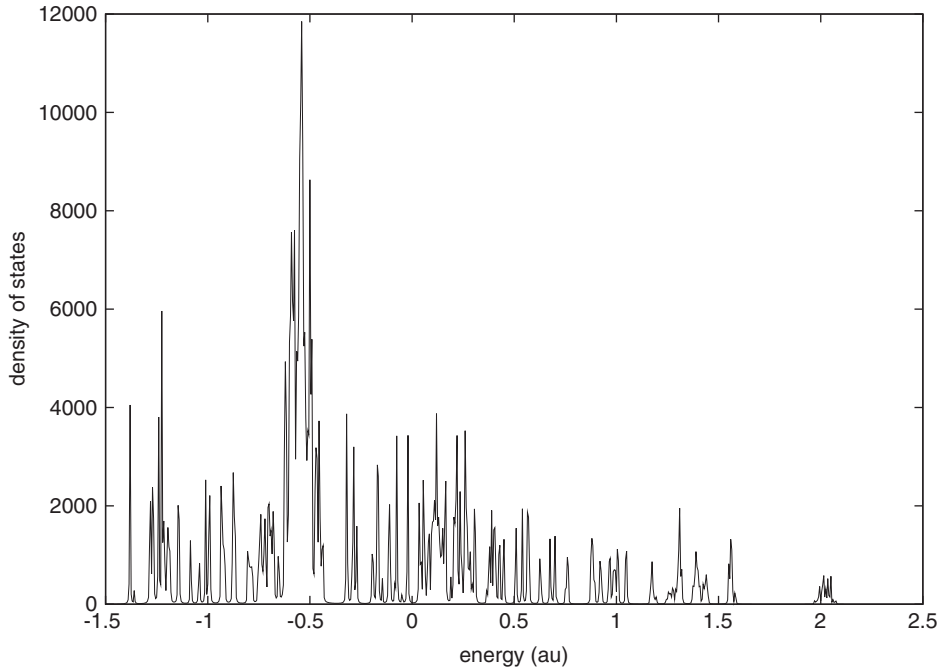


Fig. 26.8. Density of states for poly(G)–poly(C) DNA dodecamer. HOMO is at −0.435 a.u. and LUMO at −0.327 a.u.)

see below for a qualification!). The bandgap we calculate for the whole molecule is actually the same as for a single GC pair; this contrasts with the self-consistent density functional calculations of Lewis *et al.* [44], who found a bandgap of 3.37 eV for a single GC section, and a narrower bandgap of 1.40 eV for a 10-base molecule. Lewis *et al.* [44] also reported a valence bandwidth of 1.1 a.u., in good agreement with our result of 0.9 a.u. However, their conduction bandwidth is smaller by a factor of two, an effect in extended Hückel theory that we have also encountered when dealing with small molecules.

Identifying the HOMO and LUMO is of course crucial for determining the electrical properties of molecules, and the values given above were determined by integrating the density of states using the trapezium rule. However, this is quite tricky, because of the peaked structure, and energy broadening is not the whole answer because this broadens states into the bandgap, introducing additional uncertainties. We can in fact perform the density of states integration very accurately using contour integration of the Green function, which is analytic in the upper-half complex energy plane. Using (44), the total number of states up to the Fermi energy is given by

$$N = \int_0^{E_F} \mathrm{d}E n(E) = \frac{1}{\pi}\, \Im\mathrm{m} \int_0^{E_F} \mathrm{d}E \mathrm{Tr}(G(E+i\varepsilon)S) \tag{63}$$

where ε is infinitesimal. As the Green function only has singularities along the real axis, Cauchy's theorem tells us that the integration in (63), indicated by contour 1 in Fig. 26.9, can be deformed into the semi-circle, contour 2, with a much greater imaginary part over most of the energy range. Changing the variable from E to ϕ (Fig. 26.9), the number of states up to E_F is then given by

$$N = \frac{1}{2\pi}\int_0^{\pi} d\phi[\sin\phi \Im m\mathrm{Tr}(GS) - \cos\phi \mathrm{ReTr}(GS)] \tag{64}$$

Gaussian integration is an appropriate way of performing the integration in (64), as it concentrates the sampling points at the ends of the range, and from Fig. 26.9 this is where the density of states is least broadened; over most of the energy range of the integral the density of states is very much broadened, and 128 energy points are sufficient for an excellent degree of accuracy.

Using this method, we find that the HOMO for poly(G)-poly(C) DNA, with 24 extra electrons transferred to the phosphate groups to account for the acidity of the molecule, lies at -0.4356 a.u., on the third state in the quasi-continuum of states below the large bandgap in Fig. 26.8. This makes the DNA effectively a conductor. On physical grounds, we would in fact expect the HOMO to be the state at the bottom of the band gap–a band gap above the HOMO stabilizes the structure, and in any case the more accurate density functional calculations of Lewis *et al.* [44] place the HOMO at the bottom of this gap. The discrepancy between the position of the HOMO and the bottom of the bandgap, corresponding to 4 electrons out of 2794, may be due simply to the inadequacies of extended Hückel. The method is not self-consistent, and we believe that this may allow two states to drop from the conduction band into the valence band.

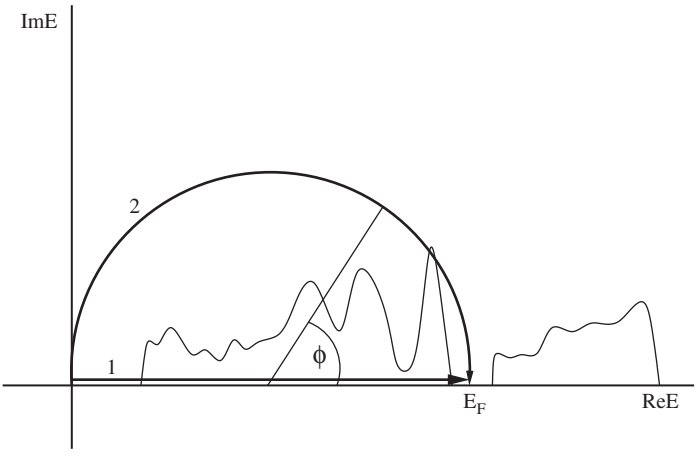


Fig. 26.9. The integration along 1 can be deformed into an integration along the semi-circle 2 in the upper-half energy plane.

26.4.3 Transmission and conductance

When calculating the transmission and conductance, it is a problem to know where to put the relative Fermi energies in the metal contacts and molecule (this problem does not arise, of course, in a self-consistent calculation, where the potentials adjust themselves in the course of the calculation [27]). In the transmission results, which we first discuss in this section, we allow the contact Fermi energy to track the energy in the molecule at which we evaluate the transmission.

The transmission across the molecule for the mixed base DNA, with the experimental X-ray structure, is shown in Fig. 26.10. In this calculation the molecule was contacted on to Cu surfaces, using a single Cu orbital in contact with a base orbital on each end of the molecule. It can immediately be seen that the transmission is very peaky. This is due to the finite size of the system, which gives rise to discrete states through which transmission occurs, rather than bands. Pendry *et al.* [45] have found that disorder in one-dimensional systems, like DNA, leads to extremely peaky transmission as a function of energy, a result which is consistent with the results of Fig. 26.10. There are isolated states with appreciable transmission, as high as 0.44, but for the majority of states the transmission is tiny, between 10^{-14} and 10^{-6} . The width of the individual transmission peaks varies, but is generally very narrow, of the order of 10^{-6} a.u.–the peak width comes from the interaction of the DNA with the continuum states in the electrodes.

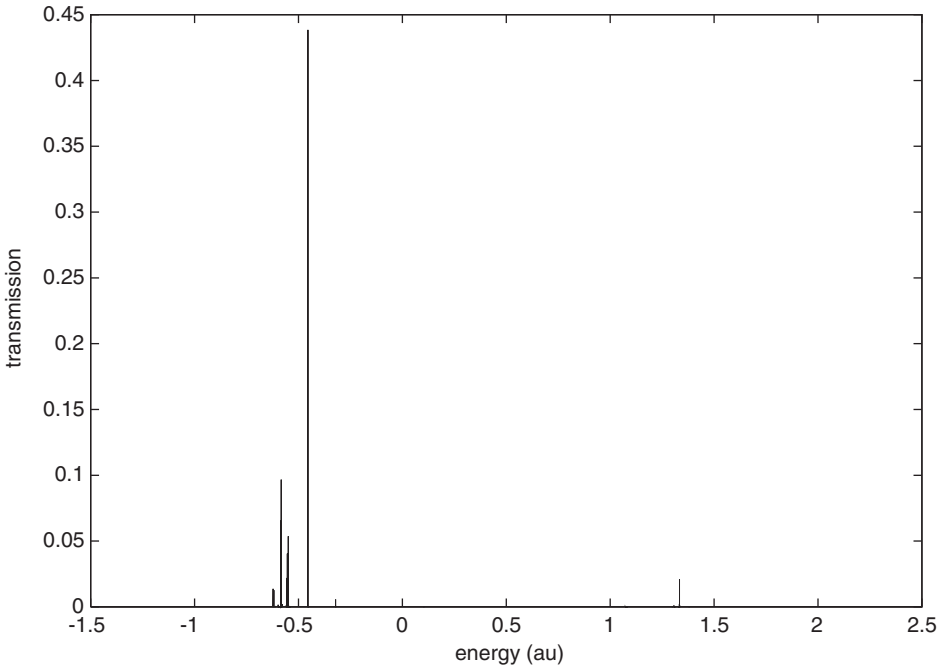


Fig. 26.10. Transmission of mixed-base DNA as a function of energy.

In the case of poly(G)–poly(C) DNA, shown in Fig. 26.11, there are many states with remarkably good transmission, of order unity. The transmission of poly(G)–poly(C) is much greater than that of mixed-base DNA, and this is in agreement with experimental results [46] and previous theory [14]. However, for nearly all energies the transmission is still effectively zero, presumably because of residual disorder in the energy-minimized structure. What is significant for conductance is the fact that there are groups of states around the bandgap with large transmission. Adessi *et al.* [47] have studied the effect of the environment of the counter-ions on the transmission properties of this form of DNA–their results show narrow bands of transmission rather than broadened discrete states, presumably because they consider an infinite system.

We now calculate the current-voltage characteristics of poly(G)–poly(C) DNA contacting gold. When a finite voltage V is applied across the metallic contacts, the current (at $T = 0$K) is given by the energy integral from the Landauer formula [48]

$$I = 2e/h \int_{\mu_r}^{\mu_l} T(E)\mathrm{d}(E) \tag{65}$$

where $\mu_{l/r}$ are the chemical potentials in the left and right metallic contacts, with $\mu_l - \mu_r = V$. In these *I/V* calculations, the Fermi energy of the molecule was placed in the middle of the large bandgap above -0.43 a.u.; this energy was taken as the mid-point for the

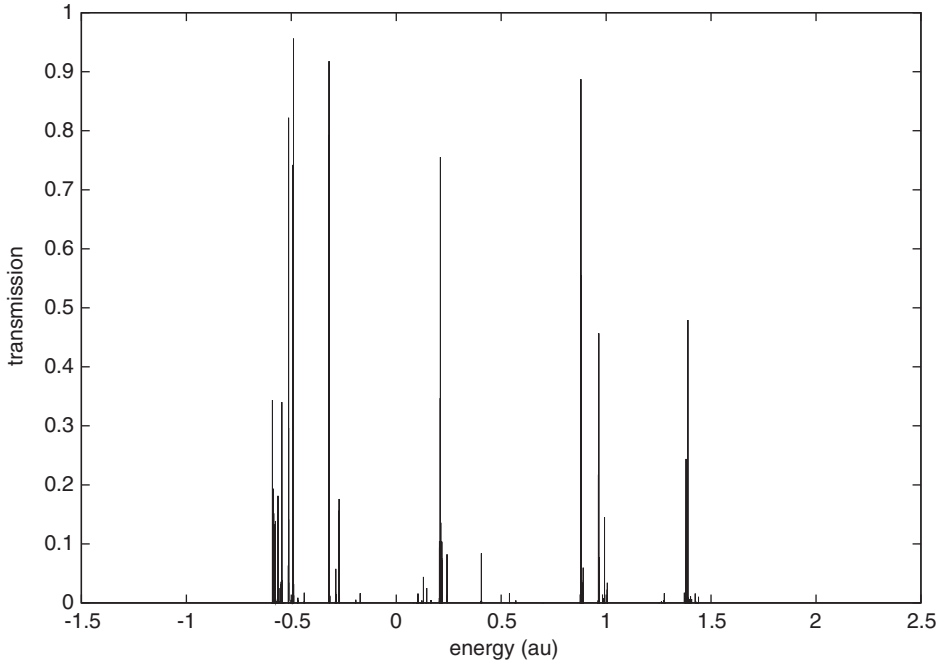


Fig. 26.11. Transmission of poly(G)-poly(C) DNA as a function of energy.

integral in (65), with the limits extending equally in energy into the valence and conduction bands. When calculating the current at voltage V, we should calculate the transmission $T(E)$ with the molecule embedded on to the Au contacts with their Fermi energies shifted by $\pm eV/2$. However, since the calculation of current involves an integral over energy, we would need to repeat the transmission calculations for each energy point, shifting the contact Fermi energies each time. We therefore fix the Fermi energy of the left and right metal contact at ± 2V, which is in the middle of the typical voltage range.

For a more realistic representation of experiment we substitute a sulphur atom for a hydrogen on sugars at each end of the molecule, and make contact between these sulphurs and the Au surfaces. In these calculations, we use full atom-to-atom contact between the Au and the S atoms, projecting the s, p and d components of the Au surface density of states on to the Au contact orbitals to find the Au Green function, as in (58).

The results for the current-voltage characteristics of poly(G)–poly(C) DNA are shown in Fig. 26.12. We see that there is a threshold voltage of 3 V, making it a wide bandgap semiconductor, with a typical current in the nanoampère range. The experimental situation is confusing, but recent work on poly(G)–poly(C) DNA shows semiconducting behaviour, with steps in the current similar to what we (and others) have found [7].

Green function and embedding techniques have enabled us to calculate the electronic structure and transport properties of a linear molecule in an efficient manner. There are other completely different linear scaling schemes – for example, the well-known SIESTA

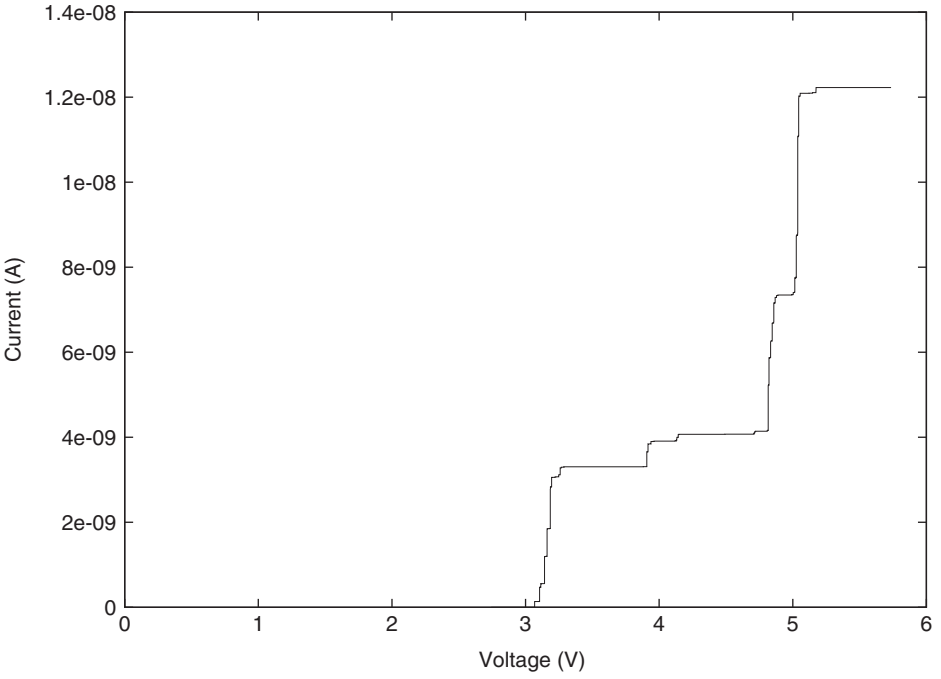


Fig. 26.12. Current-voltage characteristics for poly(G)-poly(C) DNA.

program [49] – in which the computer time for calculating total energy, say, is proportional to the size of the system, but for finding the Green function in linear molecules, the subvolume embedding methods are straightforward and efficient. In this chapter we have also shown how the ubiquitous self-energy is the same as our embedding potential, a concept with wide applicability.

26.5 REFERENCES

1 Y. Benenson, T. Paz-Elizur, R. Adar, Z. Livneh and E. Shapiro, Nature, 414 (2001) 430.
2 J. Fritz, M.K. Baller, H.P. Lang, H. Rothuizen, P. Vettiger, E. Meyer, H.-J. Güntherodt, Ch. Gerber and J.K. Gimzewski, Science, 288 (2000) 316.
3 N.C. Seeman, Nature, 421 (2003) 427.
4 D.D. Eley and D.I. Spivey, Trans. Faraday Soc., 58 (1962) 411.
5 C.J. Murphy, M.R. Arkin, Y. Jenkins, N.D. Ghatlia, S.H. Bossmann, N.J. Turro and J.K. Barton, Science, 262 (1993) 1025.
6 E. Braun, Y. Eichen, U. Sivan and G. Ben-Yoseph, Nature, 391 (1998) 775.
7 D. Porath, A. Bezryadin, S. de Vries and C. Dekker, Nature, 403 (2000) 635.
8 H.W. Fink and C. Schonenberger, Nature, 398 (1999) 407.
9 M.E. Nunez, D.B. Hall and J.K. Barton, Chemistry and Biology, 6 (1999) 85.
10 S. Roche, Phys. Rev. Lett., 91 (2003) 108101.
11 J.G. Kushmerick, D.B. Holt, J.C. Yang, J. Naciri, M.H. Moore and R. Shashidhar, Phys. Rev. Lett., 89 (2002) 086802.
12 T. Tada, M. Kondo and K. Yoshizawa, Chem. Phys. Chem. 4 (2003) 1256.
13 O.R. Davies and J.E. Inglesfield, Phys. Rev. B, 69 (2004) 195110.
14 M. Hjort and S. Stafström, Phys. Rev. Lett., 87 (2001) 228101.
15 O.R. Davies and J.E. Inglesfield, Prog. Surf. Sci., 74 (2003) 161.
16 H. Ness, S.A. Shevlin and A.J. Fisher, Phys. Rev. B, 63 (2001) 125422.
17 G.P. Triberis, C. Simserides and V.C. Karavolas, J. Phys. Condens. Matter, 17 (2005) 2681.
18 J.E. Inglesfield, J. Phys. C, 14 (1981) 3795.
19 J.E. Inglesfield, Comp. Phys. Commun.,137 (2001) 89.
20 J.E. Inglesfield and G.A. Benesh, Phys. Rev. B, 37 (1988) 6682.
21 H. Ishida, Phys. Rev. B, 63 (2001) 165409.
22 C. Caroli, R. Combescot, D. Lederer, P. Nozières and D. Saint-James, J. Phys. C: Solid St. Phys., 4 (1971) 2598.
23 A.J. Fisher, J. Phys. Condens. Matter, 2 (1990) 6079.
24 S. Crampin, J.B.A.N. van Hoof, M. Nekovee and J.E. Inglesfield, J. Phys.: Condens. Matter, 4 (1992) 1475.
25 J.B.A.N. van Hoof, The embedding method, thesis, Katholieke Universiteit Nijmegen, Nijmegen, 1996.
26 A. Levy Yeyati and M. Büttiker, Phys. Rev. B, 52 (1995) 14 360.
27 M. Brandbyge, N. Kobayashi and M. Tsukada, Phys. Rev. B, 60 (1999) 17 064.
28 P.S. Damle, A.W. Ghosh and S. Datta, Phys. Rev. B, 64 (2001) 201403.
29 D. Wortmann, H. Ishida and S. Blügel, Phys. Rev. B, 66 (2002) 075113.
30 J.E. Inglesfield, S. Crampin and H. Ishida, Phys. Rev. B, 71 (2005) 155120
31 J. Heurich, J.C. Cuevas, W. Wenzel and G. Schön, Phys. Rev. B, 88 (2002) 256803.
32 F. García-Moliner and V.R. Velasco, Theory of single and multiple interfaces, World Scientific, Singapore, 1992.
33 G.A. Baraff and M. Schlüter, J. Phys. C, 19 (1986) 4383.
34 J. Ferrer, A. Martín-Rodero and F. Flores, Phys. Rev. B, 38 (1988) 10 113.
35 M.P. Anantram and T.R. Govindan, Phys. Rev. B, 58 (1998) 4882.
36 R. Lake, G. Klimeck, R.C. Bowen and D. Jovanovic, J. Appl. Phys., 81 (1997) 7845.
37 M.I. Trioni, S. Marcotulio, G. Santoro, V. Bortolani, G. Palumbo and G.P. Brivio, Phys. Rev. B, 58 (1998) 11 043.

38 R. Hoffmann, J. Chem. Phys., 39 (1963) 1397.
39 J. Cerdá and F. Soria, Phys. Rev. B, 61 (2000) 7965.
40 G.A. Leonard and W.N. Hunter, J. Molec. Biol., 234 (1993) 198.
41 ViewerLite 4.2, Accelrys Inc.
42 W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell and P.A. Kollmann, J. Am. Chem. Soc., 117 (1995) 5179.
43 HyperChem 7.1, Hypercube Inc.
44 J.P. Lewis, P. Ordejón and O.F. Sankey, Phys. Rev. B, 55 (1997) 6880.
45 J.B. Pendry, A. Mackinnon and P.J. Roberts, Proc. Roy. Soc. London A, 437 (1992) 67.
46 C. Dekker and M.A. Ratner, Physics World, 14 (2001) 29.
47 Ch. Adessi, S. Walch and M.P. Anantram, Phys. Rev. B, 67 (2003) 081405.
48 R. Landauer, IBM J. Res. Dev., 1 (1957) 223.
49 J.M. Soler, E. Artacho, J.D. Gale, A. García, J. Junquera, P. Ordeón and D. Sánchez-Portal, J. Phys. Condens. Matter, 14 (2002) 2745.

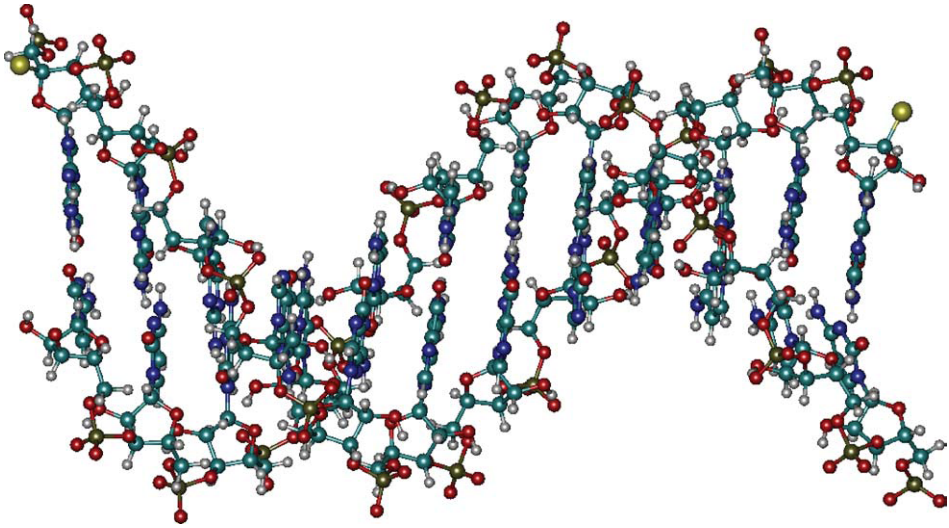


Plate 34. 12 base-pair poly(G)-poly(C) DNA molecule. The double ring guanine bases lie consecutively along the same backbone strand, with the complementary single ring cytosine bases along the opposite strand (illustration appears on page 526 of this volume).

Modern Methods for Theoretical Physical Chemistry of Biopolymers
Edited by E.B. Starikov, J.P. Lewis and S. Tanaka

CHAPTER 27

Ballistic conductance for all-atom models of native and chemically modified DNA: a review of a Kubo-formula-based approach

Ewgeni B. Starikov[1], Shigenori Tanaka[2], Noriyuki Kurita[3], Yasuo Sengoku[3], Takayuki Natsume[3], Aina Quintilla[1] and Wolfgang Wenzel[1]

[1]*Research Center Karlsruhe, Institute for Nanotechnology, D-76021 Karlsruhe, Germany*
[2]*Graduate School of Science and Technology, Kobe University, 1-1, Rokkodai, Nada, Kobe 657-8501, Japan*
[3]*Department of Knowledge-Based Information Engineering, Toyohashi University of Technology, Toyohashi, Aichi 441-8580, Japan*

Abstract

A novel approach to estimate DNA conductance based upon the Kubo formula is presented and discussed. Using this approach, the effects of base-pair mismatches, different conformational changes and base-pair sequence on DNA electrical properties can be investigated. The new approach makes possible very fast estimation of conductance spectra for oligonucleotides with hundreds of base pairs and can easily be extended to treat arbitrary chemical modifications of DNA.

27.1 INTRODUCTION

The problem of DNA conductivity has a long history, but is still heavily debated to date (for the most recent review, see, for example, [1,2]). As one of the most fascinating biological substances with many possible biotechnological applications, DNA is anyway a highly popular object of both experimental and theoretical research efforts. Nevertheless, no consensus has yet emerged as to the fundamental mechanisms of electronic transport through DNA and its specifics in a variety of experimental environments. The chemical complexity of this polymeric molecule, influence of its monomer (base) sequence and counter-ion–water surrounding, render any study on DNA electrical conductance an immensely complicated

task, and thus call for continued development of theoretical methods that help elucidate DNA conductance properties on the basis of as realistic models as possible.

Several approaches to study DNA conductivity have emerged, most recently Refs [3–7], in particular, various methods based upon simplified (tight-binding) and all-atom models of DNA. In the former techniques, separate nucleotide bases or base pairs are considered as 'sites' with some 'site energies' corresponding physically to their ionization potentials. Along with this, Watson–Crick hydrogen-bonding and/or base stacking in DNA are described as 'hopping integrals'. However, these models neglect, by definition, all the atomistic details and are therefore drastically limited in their resolution of the underlying electronic structure of the system involved.

On the other hand, all-atom techniques represent DNA structure in a realistic atomically resolved picture. As a result, they enable one to derive the corresponding Hamiltonian from established quantum chemical techniques. The extended Hückel method and density-functional theory (DFT) are commonly used for this purpose [8–11].

Meanwhile, it should be noted that neither class of these models allows any straightforward accounting for the influence of counter-ion–water surrounding of DNA.

The simplicity of the tight-binding approaches permits consideration of extremely long finite oligonucleotide duplexes, which is apparently a computationally formidable task for competing all-atom methods (still, a maximum of 12 base pairs was taken into account in the most recent study (see [9,10] and the corresponding chapter in this book). Alternatively, infinite homopolynucleotide duplexes poly(dG)–poly(dC) and poly(dA)–poly(dT) can be investigated with DFT at the all-atom level to evaluate the one-electron energy band structure [11]. The latter study suggests that such infinite polymers *themselves* (without additives) could *at the most* be rather poor, wide-gap semiconductors. After calculating the electron structure of the all-atom DNA models, their DNA conductance is usually estimated either with the Landauer–Büttiker formalism (see, for example, [9,10,12]) or via charge tunneling probability [11].

To sum up, there is presently a definite lack of efficient theoretical/computational methods to treat very long and finite DNA duplexes, although such techniques promise significant progress in designing novel DNA-based nano-devices [13]. Hence, we would like to propose here such an approach to estimate electrical conductance for finite DNA duplexes of arbitrary length, which could also be of use when studying DNA complexes with any possible inorganic/organic materials.

This novel formulation estimates the ballistic transport through finite atomically resolved DNA fragments within the framework of the Kubo formalism, recast in its Green function representation (see, for example, [14,15]). Such an approach permits extremely fast evaluation of DNA conductance, with the computation time scaling linearly versus DNA duplex length.

27.2 METHOD OF CALCULATION

The main idea of the present contribution is to modify the approach advocated in the work by J. A. Verges [15] to estimate DNA conductivity. First, let us recall that the work [15] considers a square two-dimensional lattice of identical 'sites' influenced by external

random magnetic fields and diagonal disorder. Specifically, his system consists of a semi-infinite left-side lead ideally bound to the lattice of interest and of the same right-side lead. The system in question could thus be represented as a quasi-one-dimensional row of 'layers' coupled to each other, with both edge 'layers' being ideally coupled to the leads.

Following the work [15], the corresponding Hamiltonian of the pure disordered lattice without electrodes and external magnetic field can be cast in the following form:

$$\hat{H}_0 = \sum_i \sum_{j=1}^{M_i} \varepsilon_{i,j} \hat{c}^{\dagger}_{i,j} \hat{c}_{i,j} - t \sum_i \sum_{j=1}^{M_{i-1}} (\hat{c}^{\dagger}_{i,j} \hat{c}_{i+1,j} + \hat{c}^{\dagger}_{i,j} \hat{c}_{i,j+1} + \hat{c}^{\dagger}_{i+1,j} \hat{c}_{i,j} + \hat{c}^{\dagger}_{i,j+1} \hat{c}_{i,j}) \tag{1}$$

where sums in i are numbering the layers ($1 \leqslant i \leqslant L$) and those in j are the sites within each layer; $-t$ is the uniform hopping energy within and between the layers (which could be set equal to unity without any loss of generality); $\varepsilon_{i,j}$ are random site energies; $\hat{c}^{\dagger}_{i,j}$ creates an electron/hole on site (i, j) and $c_{i,j}$ annihilates it, M_i stands for a variable layer width.

After attaching both the first and the Lth layers of the sample described by the Hamiltonian (1) to ideal electrodes (leads) and applying electric potential V to it, one can formulate the Kubo theory (theory of linear electric response). The work [15] makes the simplest (but seminal) assumption that the distribution of the applied electrical potential in the whole system is just *stepwise* – all the sample and right lead sites ($i \geqslant 1$) are shifted by a constant electric potential energy of $-eV$, whereas all the left lead sites ($i < 1$) remain at zero potential (see Fig. 27.1). This dictates the following choice for a position operator (here the position along the whole system is meant):

$$\hat{x} = \sum_{i \geqslant 1} \sum_{j=1}^{M_i} \hat{c}^{\dagger}_{i,j} \hat{c}_{i,j} \tag{2}$$

and the corresponding perturbation Hamiltonian,

$$\hat{H}_1 = -eV \cos(\omega t) \hat{x} \tag{3}$$

where the simple cosine dependence of the applied voltage is assumed. Owing to this perturbation, the charge Q could be transferred from the right lead to the left one through the stationary eigenstates of the quantum system, and therefore some time-dependent current

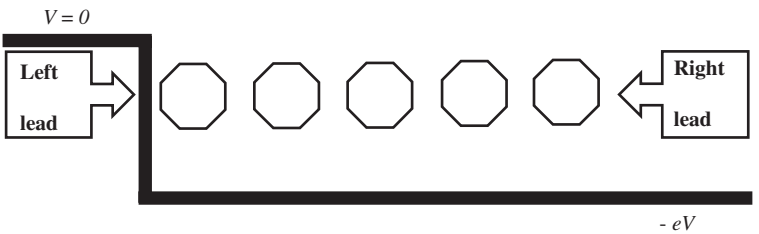


Fig. 27.1. The main assumption of the Verges method [15] – a stepwise distribution of the applied electric potential.

I arises. Then, the formulation of the corresponding quantum mechanical operators for the charge and current is straightforward:

$$\hat{Q} = -e\sum_{i\geq 1}\sum_{j=1}^{M_i}\hat{c}_{i,j}^{\dagger}\hat{c}_{i,j} = -e\hat{x} \tag{4}$$

$$i\hbar\hat{I} = [(\hat{H}_0 + \hat{H}_1)]\hat{Q} = -e[(\hat{H}_0 + \hat{H}_1)\hat{x}] = -ei\hbar\hat{v}_x \tag{5}$$

where $\hat{v}_x$ is the velocity operator.

Within the frame of the first-order perturbation theory, one could now estimate the average current:

$$\langle\hat{I}\rangle = \sum_{\alpha} f_\alpha \langle \Psi_\alpha |\hat{I}| \Psi_\alpha\rangle \tag{6}$$

where Ψ_α are the perturbed wavefunctions, f_α their occupancies, as well as the conductance of the quantum system, using the conventional form of Ohm's law:

$$G = \lim_{\omega\to 0}\frac{\langle\hat{I}\rangle}{V} = -e^2\hbar\pi\lim_{\omega\to 0}\sum_{\alpha,\beta}\left|\langle\alpha|\hat{v}_x|\beta\rangle\right|^2\frac{f_\alpha - f_\beta}{\varepsilon_\alpha - \varepsilon_\beta}\,\delta(\varepsilon_\alpha - \varepsilon_\beta - \hbar\omega) \tag{7}$$

Meanwhile, from the computational standpoint, it would be much more convenient to recast Eq. (7) in its Green function form [14,15], which will in effect be mathematically equivalent to the expression for *G* independently derived within the framework of the Landauer–Büttiker formalism (see, for example, [12]) and used in the works [9,10]:

$$G = 2\left(\frac{e^2}{h}\right)\mathrm{Tr}[(i\hbar\hat{v}_x)\mathrm{Im}\hat{\Gamma}(E)(i\hbar\hat{v}_x)\mathrm{Im}\hat{\Gamma}(E)] \tag{8}$$

where $\mathrm{Im}\hat{\Gamma}(E)$ is determined by the advanced and retarded Green functions:

$$\mathrm{Im}\hat{\Gamma}(E) = \frac{1}{2i}\left[\hat{\Gamma}^R(E) - \hat{\Gamma}^A(E)\right] \tag{9}$$

and *E* stands for the Fermi energy of the system.

To make computations with Eqs (8) and (9), we first need to know the velocity of the charge flow through the DNA duplex under study, as well as the full Green function of the latter. Interestingly, the stepwise nature of the electric potential distribution we adopted here enables us to enormously simplify the expression for the charge velocity operator, which could in effect be cast in the following form [15]:

$$\hat{v}_x = -\frac{\mathrm{t}}{i\hbar}\sum_{j=1}^{m}(\hat{c}_{0,j}^{\dagger}\hat{c}_{1,j} - \hat{c}_{1,j}^{\dagger}\hat{c}_{0,j}) \tag{10}$$

where *m* is the number of modes in (width of) the left infinite lead.

Equation (10) is very important in Verges theory [15], since it shows that only knowledge of the charge flux through the contact between the left lead and the sample is required to evaluate the total current through the quantum system, provided the electric potential distribution is (at least, approximately) stepwise. This facilitates the whole procedure to compute *G*.

Moreover, Eq. (10) reveals that, to evaluate G, only Green function matrix elements at layers $i = 0$ (left lead) and $i = 1$ (the layer of the system making direct contact with the left lead) are really needed. This is another step toward the computational procedure facilitation.

Indeed, the choice of the method to calculate the Green function is now straightforward. First, we use the Dyson equation to back-iterate the Green function layer by layer, starting from the right lead:

$$\Sigma_{i-1} = \mathbf{V}_{i-1,i}(E\mathbf{I} - \mathbf{V}_{i,j} - \Sigma_i)^{-1}\mathbf{V}_{i,i-1} \tag{11}$$

whereby self-energy matrix Σ_{i-1} is calculated, if the one for the next layer is known, together with the Hamiltonian sub-matrices describing intra-layer $\mathbf{V}_{i,i}$ and interlayer $\mathbf{V}_{i-1,i}(\mathbf{V}_{i,i-1} = \mathbf{V}^{\dagger}_{i,i-1})$ interactions.

After completing the iteration (11), we obtain the self-energy matrix Σ_r of the right lead plus the sample, and, combining it with the self-energy matrix Σ_l of the left lead, write the following equation to obtain the Green function matrix elements necessary to evaluate the desired conductance:

$$[E\mathbf{I} - \mathbf{H} - \Sigma_l(E) - \Sigma_r(E)]\hat{\Gamma}(E) = \mathbf{I} \tag{12}$$

where **H** is the matrix of the 'layer 0 plus layer 1' system Hamiltonian.

In applying Verges theory for evaluation of DNA conductivity as outlined above, we note that the role of the 'layers' is taken over by DNA nucleotide pairs. This allows us to transfer the maximum of the theoretical inferences from the work [15] to our case. The most important difference arises here, because in the Verges system the orthogonal basis of 'atomic' orbitals (AO) is used for each of the 'sites' involved (every Verges 'site' carries only one AO). In our DNA model, we generalize this approach to treat nonorthogonal AO basis sets on atomic sites capable of containing more than one AO. The latter ones arise naturally in our system consisting of realistic molecular fragments, which require detailed quantum chemical treatment for their consistent description. A formal solution to the problem of electric transport through molecules using nonorthogonal AO basis sets was discussed in detail in the book [12]. Here we only mention how the above equations will be recast in the new AO basis.

Specifically, we use an extended Hückel model describing the electronic structure of DNA, that is, here we do not explicitly account for electron–electron correlations even in the Pauli sense. We have previously demonstrated this to be a reasonable approximation for rather long DNA molecules [16]. The Hückel parameters for the present study were the same as in the 'HUCKEL-IV' routine package for computational molecular electronics [17]. Then, the most important changes will be introduced into Eqs (11) and (12) of the original Verges theory, where the identity matrix I must be explicitly substituted by overlap matrices S with elements $S_{i,j} = \langle \varphi_i | \varphi_j \rangle$, where φ_i are nonorthogonal atomic orbital wavefunctions of the DNA nucleotide pairs.

Thus, based upon the above ideas, we are able to propose the following algorithm to estimate DNA conductance:

- We divide the given DNA duplex into nucleotide pairs (in our algorithm, they could be the conventional Watson–Crick or mismatched ones, the sequence of the pairs can be arbitrary).

- The leftmost and rightmost nucleotide pairs are covalently attached to sulfur atoms. In the present work we assume that each sulfur atom establishes a single covalent bond to the pertinent deoxyribose O3′–atom (the S–O3′ bond length is 1.74 Å, the C3′–O3′–S bond angle is 109.5°, the C4′–C3′–O3′–S torsion angle is 180°).
- Each of the above-mentioned sulfur atoms is in turn attached to a fragment of the gold [111] surface represented by a triangle of Au atoms, so that all three S–Au distances are 2.53 Å, whereas the Au–Au bond lengths in the triangle are 2.885 Å. Our computer program automatically determines how the Au triangles should be placed with respect to the DNA duplex involved. Specifically, the approximation to the double-helical axis is evaluated first, in that the differences between the *X, Y* and *Z* coordinates of the proximal sulfur atoms are estimated. The axis with the greatest difference is then declared to be the axis of interest, so that the Au triangles are finally attached perpendicularly with respect to this axis. With this in mind, some preliminary work, using a clever molecular viewer software, is still necessary to orientate the DNA duplex in the correct way.
- Both the Au triangles are considered firmly attached to ideal electrodes, so that the latter ones can have a maximum of 27 modes, that is, three times the number of valence electrons of the Au atom.
- Like in [15], we assume that the potential V due to a voltage bias applied to both the electrodes has a very simple stepwise distribution, that is, the potential is zero in the left lead up to the left Au triangle, whereas it is shifted by an electric potential energy of $-eV$ in all the nucleotide pairs, as well as in the right Au triangle together with the right lead.
- The above assumption on the electric potential distribution ensures that the knowledge of the charge flux in the contact between the left Au triangle and its neighboring nucleotide base pair would be enough to estimate the total current through the whole system. This is why we need only Green functions of the left Au triangle and its neighboring nucleotide base pair for our estimations using Eq. (6).
- We start from the right end by obtaining the self-energy of the ideal right-side lead, using well-known analytical expressions (see, for example, [12,15]). This self-energy is used as an input to the Dyson equation, which is then iterated from the right-side Au triangle up to the leftmost nucleotide pair (see Eq. (11)), together with the analytical calculation of the self-energy of the ideal left-side lead. The latter self-energy corrects the Hamiltonian of the left-side Au triangle, whereas the Hamiltonian of the leftmost nucleotide is corrected by the self-energy of the resting nucleotide pairs, together with the whole right-side electrode. Finally, the total Green function of the system is calculated using both of these corrected Hamiltonians forming a block matrix according to Eq. (12).
- Apart from the above Green function, to use the Kubo formula we also need the matrix representation of the charge velocity operator [15]. In our nonorthogonal basis representation, this operator is also a special block matrix including matrix elements for interaction between the left-side Au triangle and the leftmost nucleotide pair. The resulting conductance is purely ballistic and corresponds to $T = 0$ K temperature, since no phonon scattering is taken into account here.

The conductance calculated with the above algorithm is naturally represented as the number of 'channels' open within DNA duplexes for the charge transport (for more elaboration on these 'channels' see, for example, [18]), or, in other words, the number of the transmitted 'conductivity quanta' (e^2/h, where e is the charge of electron and h is Planck constant) [12]. Thus, to estimate DNA transmittance, one would simply need to divide this number of 'channels'/'quanta' by the maximum number of modes in the electrodes.

Some advantages and disadvantages of our proposed approach are listed below.

- The most important approximation of our method lies in the assumption of the electrical potential distribution with the DNA duplex. Actually, the latter assumption restricts the energy spectrum of conduction that can be considered to be in the 1–2 eV range around the Fermi level of the neutral isolated DNA duplex. Larger voltage biases might severely affect the electrical potential energy profile, which would make our approximation fully inapplicable. Moreover, placing a molecule between two electrodes physically means that this molecule is situated within a capacitor. Everybody who is familiar with the basic principles of electrical engineering, might argue that our 'stepwise' approximation is physically inconsistent, since there must be a linear decrease/increase of the electrical potential between the two leads of a capacitor. Nevertheless, if the distance between the capacitor leads is long enough, the slope of the potential linear decrease/increase can be considered to be negligible, and this, indeed, encourages us to safely treat conductivities of very long DNA molecules on the basis of our 'stepwise' approximation.
- Anyway, this same assumption is necessary to noticeably accelerate the calculation process, in that we need to walk *only once* through the DNA duplex when iteratively calculating its Green function. Meanwhile, the analogous method by Inglesfield *et al.* [9,10] requires *three* iterative walks of this sort.
- Hence, we are capable of quickly calculating plots of DNA conduction/transmission versus Fermi energy in a relatively narrow energy range around the Fermi level of the neutral isolated DNA duplex. The energy step we use to reconstruct the Fermi energy spectrum of DNA conduction is 0.01 eV, just one half of the kT energy (k – Boltzmann constant) at room temperature. Here we deal with DNA *hole* conductivity only (at energies lower than the Fermi levels of the neutral isolated molecules under study).
- The iterative Green function methods are well known to be possessed of intrinsic numerical instabilities (see, for example, [19]). We treat this problem as follows: if at some Fermi energy the calculated conduction is too high or negative, we scan the nearest vicinity of this energy with a smaller energy step of 0.001 eV to estimate the borders of the numerically unstable region. The conductance at the Fermi energy of interest is then evaluated through linear interpolation using the two revealed border values. Finally, for those who nonetheless still tend to think of our method as a case of doubt, we might also add that the computational efficiency and overall relevance of the original Verges method has already been carefully tested on generalized lattices–and excellent, very encouraging results have been obtained [20].

Our technique has been implemented in a serial C-code (this program also requires the well-known linear algebra library BLAS/LAPACK) and tested on different computer

platforms [21]. The most elaborate part of our present algorithm is the iterative calculation of the DNA Green function using Eq. (11) (successive Hamiltonian matrix inversions), which is carried out *only once*. Since the method presented by Inglesfield *et al*. requires *three* such calculations at every value of Fermi energy [9,10], we anticipate that it will work approximately *three* times as slow as ours.

27.3 DUPLEXES UNDER STUDY

The above approach has already been applied to evaluate conduction spectra of different DNA oligomer duplexes [21], so that here we only wish to show one example of the application of our method.

Specifically, among other systems, the following DNA duplexes with 15 base pairs, the conventional one and those containing base pair mismatches (only DNA themselves, stripped of all other ingredients, were subject to the calculations) will be considered here (see also [21] for more details):

- The conventional DNA duplex d(5′-AGTACAGTCATCGCG-3′), charge transfer properties of which were experimentally studied in the work [22]. From here on, we denote this molecule as 15BP;
- This same duplex, but with the AC mismatch at the forth position from the 5′-end, also studied in the work [22]. From here on, we denote this molecule as 15BP-AC.
- This same duplex, but with the GA mismatch at the seventh position from the 5′-end, also studied in the work [22]. From here on, we denote this molecule as 15BP-GA.

As to the latter DNA duplex group, the experimental finding of interest for the present theoretical study is that the amount of charge transferred through the DNA duplex decreases to one-third when comparing the conventional oligonucleotide with that containing the A–C mismatch. On the other hand, it is only slightly changed when comparing the conventional oligonucleotide with that containing the G–A mismatch [22]. We assume here that such an observable situation could be connected with alterations in conductive properties of the DNA oligonucleotide in question, which ought to be introduced together with the pair mismatches, and is thus of great interest to test this hypothesis.

For this purpose, we have generated atomic coordinates of each of the three above-mentioned duplexes, put them into spherical water hulls (containing so many Na+ counter-ions as to fully saturate the DNA negative charge) and minimized the energy of these hydrated complexes using an AMBER force field [23], as implemented within the TINKER routine package [24]. The resulting structures have been used to carry out the conductance calculations (only DNA themselves, stripped of all other ingredients, were subject to the conductivity calculations).

27.4 RESULTS AND DISCUSSION

Figure 27.2 gives the transmittance spectrum of 15BP-AC and 15BP-GA, in comparison with that of 15-BP, in the energy region of their valence bands.

To exclude the effect of the electrode orientation, we have made it identical for all the three duplexes involved [21].

Thus, what is seen in Fig. 27.2 is that the highest-energy (−12.45 eV) transmittance peak belongs to the 15BP-GA spectrum, but its height is very low (0.04) and it is very narrow (0.02 eV). The second peak (−12.51 eV) belongs to the 15BP spectrum and is somewhat wider (0.04 eV), as well as one order of magnitude higher (0.6). The third peak (−12.52 eV) is again very narrow (0.02 eV) and approximately two times as low as the second one (0.33), but still one order of magnitude higher than the first one. This third peak belongs to the 15BP-AC spectrum.

Molecular charge transfer (MCT) data of the kind reported in [22] is often interpreted as a tunneling process to a large extent assisted by fluctuations from intramolecular vibrations and solvent environment. The conventional basis for such considerations is provided by Marcus theory [25], which considers the MCT rate k_{MCT} as a product of the form $\mathrm{K_{el}K}_n\omega$, where ω stands for effective vibrational frequency, K_n denotes the so-called nuclear (or Franck–Condon) factor. The electronic factor K_{el} in this product could be interpreted *either as* coupling between the electronic systems of the charge donor and acceptor (described by the hopping integral in the tight-binding representation) *or as* the electronic transmission coefficient.

The above duality allows inference of the exact linear mathematical mapping of the Marcus theory parameter K_{el} onto the conduction/transmission coefficient known from Landauer–Büttiker or Kubo electronic transport theories [26,27]. Physically, this means

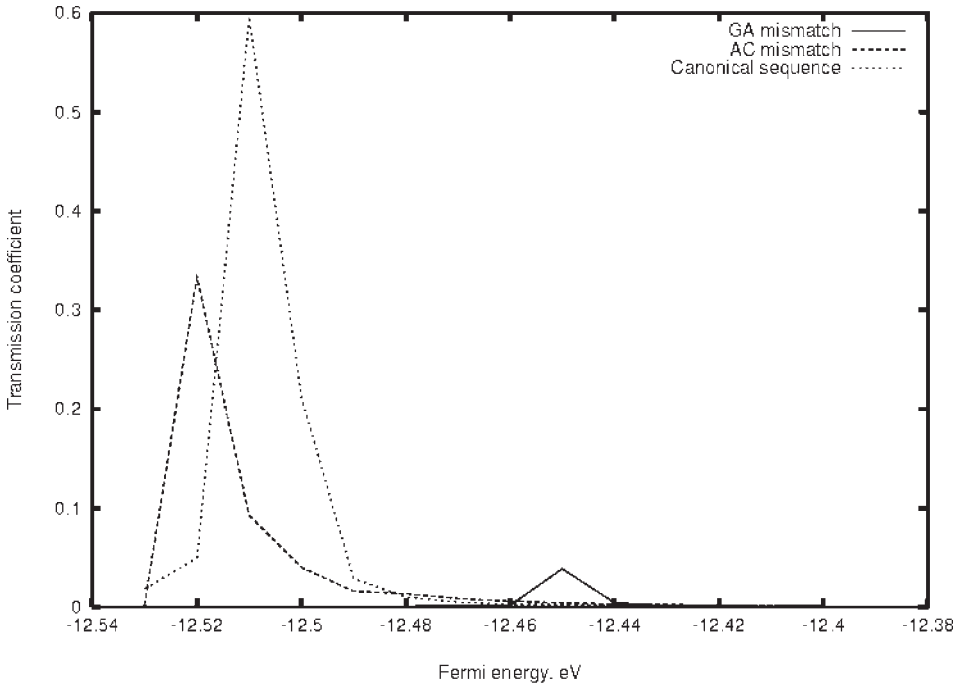


Fig. 27.2. Comparison of 15BP, 15BP-AC and 15BP-GA transmittance spectra in the energy range of their valence bands at the zero bias voltage.

that MCT could be theoretically treated as charge transport in a (semi)conductor under the influence of inelastic scattering, where the latter one is connected with the fluctuations resulting from low-frequency intramolecular vibrations and/or dynamics of the solvent environment. In the systems with inelastic scattering, the overall conductance can in general be cast as a sum of the ballistic and incoherent components [28], which are, in conventional MCT terminology, known as superexchange and hopping terms, respectively.

Detailed numerical studies [26] have shown that the incoherent component can both enhance and diminish the overall electronic transmission (overall donor–acceptor electronic coupling), depending on the interplay between the local tunneling probabilities (δ_i) and local inelastic scattering strengths (ζ_i) on every *i*th site of the system in question. The parameter ζ_i can physically be considered a probability of the quantum mechanical event that the incident charge will be inelastically scattered on the *i*th site, or, in other words, will undergo a phase-breaking (dephasing) process on it. Specifically, assuming that all the sites in the system are identical, an increase in ζ would diminish the total transmission K_{el} in a strong tunneling regime ($\delta \approx 1$), whereas K_{el} would be enhanced with the ζ increase in a weak tunneling regime ($\delta \approx 0$). Interestingly, the intermediate tunneling regime exhibits a non-monotonous behavior of K_{el} as a function of ζ [26], with increasing K_{el} up to some critical point in ζ and then decreasing.

This ballistic/superexchange component is calculated in the present work, so that, based upon our results and the experimental data from [22], we can qualitatively estimate the tunneling regime for the normal mismatched DNA duplexes involved. Bearing the above reasoning in mind, we might speculate that the conventional 15BP system is in the intermediate tunneling regime, and AC mismatch ought to drive the 15BP DNA duplex to the strong tunneling regime, whereas a GA mismatch would cause weak tunneling. This seems to be a likely assumption, since G is known to be a better 'hole trap' than A, T and C, and thus can cause somewhat more efficient dephasing (see, for example [29]).

Thus, we may tentatively conclude that inelastic scattering processes seem to be capable of significantly enhancing the 15BP-GA transmittance at its (−12.45 eV) peak to make it practically equal to the 15BP (−12.51 eV) peak. At the same time, inelastic scattering processes ought to slightly diminish the 15BP-AC transmittance at its (−12.52 eV) peak, to make it about one third of the normal 15BP transmission. This could serve as a tentative rationalization of the experimental data in [22], although further detailed work (far beyond the scope of the present chapter) is necessary to draw final conclusions.

Finally, our calculations [21] show that the DNA fundamental gap lies in a definite range of about 2–3 eV, in full accordance with experimental data (for the most recent account see, for example, [30]) and with earlier theoretical estimations [9–11,16,31]. The fact that we are capable of reproducing a DNA electronic structure in a decent accordance with other theoretical results fully justifies our most critical approximation – the stepwise shape of the potential *V* due to a voltage bias applied to both the electrodes.

27.5 CONCLUSIONS

Our results suggest that isolated DNA itself is at best a kind of poor semiconductor, if not just a trivial insulator. This finding is in accordance with previous theoretical treatises.

Chemically, one could still select some special charge donor and/or acceptor moieties, which, being properly attached to a DNA duplex, would use its specific channels to noticeably facilitate the charge transfer between them. Our above results clearly demonstrate the presence of such channels in different DNA duplexes. External effects, as well as internal structural alterations, should play a significant role in tuning the effectiveness of the DNA charge transfer channels.

Moreover, we feel that application of the approach presented here would result in a noticeably faster estimation of the DNA conductance spectrum than the already available methods of the same class. This should enable us to efficiently estimate conductance spectra for DNA duplexes of arbitrary length (up to hundreds of nucleotide pairs, if not even longer), as well as their complexes with diverse organic/inorganic molecular or polymeric agents. This latter feature is of crucial significance for the modern theoretical nano-bioscience.

27.6 ACKNOWLEDGEMENTS

The authors acknowledge financial support from the 'Volkswagen Stiftung' and 'Deutsche Forschungsgemeinschaft' foundations (grant WE 1863 14-1), as well as from the Japanese Council for Science and Technology (ACT-JST project '*Ab Initio* Approach to Nanoscale Dynamics of DNA'), and express their gratitude to the Supercomputing Facility of the Research Center Jülich, Germany, for generous allocation of supercomputer time.

27.7 REFERENCES

1 R.G. Endres, D.L. Cox and R.R.P. Singh, Rev. Mod. Phys., 76 (2004) 195.
2 D. Porath, G. Cuniberti and R. Di Felice, Top. Curr. Chem., 237 (2004) 183.
3 P. Carpena, P. Bernaola-Galvan, P.Ch. Ivanov and H.E. Stanley, Nature, 418 (2002) 955.
4 P. Carpena, P. Bernaola-Galvan, P.Ch. Ivanov and H.E. Stanley, Nature, 421 (2003) 764.
5 S. Roche, Phys. Rev. Lett., 91 (2003) 108101.
6 S. Roche, D. Bicout, E. Macia and E. Kats, Phys. Rev. Lett., 91 (2003) 228101.
7 S. Roche, D. Bicout, E. Macia and E. Kats, Phys. Rev. Lett., 92 (2004) 109901.
8 Ch. Adessi, S. Walch and M.P. Anantram, Phys. Rev. B, 67 (2003) 081405.
9 O.R. Davies and J.E. Inglesfield, Prog. Surf. Sci., 74 (2003) 161.
10 O.R. Davies and J.E. Inglesfield, Phys. Rev. B, 69 (2004) 195110.
11 H. Wang, J.P. Lewis and O.F. Sankey, Phys. Rev. Lett., 93 (2004) 016401.
12 S. Datta, Electronic transport in mesoscopic systems, Cambridge University Press, Cambridge, 1995.
13 Ch.M. Niemeyer, B. Ceyhan, M. Noyong and U. Simon, Biochem. Biophys. Res. Comm., 311 (2003) 995.
14 P. A. Lee and D.S. Fisher, Phys. Rev. Lett., 47 (1981) 882.
15 J.A. Verges, Comput. Phys. Comm., 118 (1999) 71.
16 E.B. Starikov, Phys. Chem. Chem. Phys., 4 (2002) 4523.
17 F. Zahid, HUCKEL-IV 2.0, http://nanohub.purdue.edu/, 2002.
18 J.C. Cuevas, J. Heurich, F. Pauly, W. Wenzel and G. Schön, Phase Transitions 77 (2004) 175.
19 T.J. Godin and R. Haydock, Comput. Phys. Comm., 64 (1991) 123.
20 J.A. Verges, Comput. Phys. Comm., 127 (2000) 268.
21 E.B. Starikov, S. Tanaka, N. Kurita, Y. Sengoku, T. Natsume and W. Wenzel, Eur. Phys. J., E18 (2005) 437.
22 S.O. Kelley, E.M. Boon, J.K. Barton, N.M. Jackson and M.G. Hill, Nucleic Acids Res., 27 (1999) 4830.

23 W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, Jr, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell and P.A. Kollman, J. Am. Chem. Soc., 117 (1995) 5179.
24 J.W. Ponder, TINKER 3.9, http://dasher.wustl.edu/tinker/, 2003.
25 R.A. Marcus and N. Sutin, Biochim. Biophys. Acta, 811 (1985) 265.
26 X.-Q. Li and Y.-J. Yan, J. Chem. Phys., 115 (2001) 4169.
27 Y.-J. Yan and H.-Y. Zhang, J. Theor. Comp. Chem., 1 (2002) 225.
28 J.L. D'Amato and H.M. Pastawski, Phys. Rev. B, 41 (1990) 7411.
29 M. Bixon and J. Jortner, J. Phys. Chem. B, 104 (2000) 3906.
30 C. Nogues, S.R. Cohen, S.S. Daube and R. Naaman, Phys. Chem. Chem. Phys., 6 (2004) 4459.
31 M. Hjort and S. Stafström, Phys. Rev. Lett., 87 (2001) 228101.