Reviews in Computational Chemistry Volume 18

Reviews in Computational Chemistry Volume 18

Edited by

Kenny B. Lipkowitz and Donald B. Boyd

⟨W⟩ **WILEY-VCH**

Kenny B. Lipkowitz
Department of Chemistry
Indiana University–Purdue University
  at Indianapolis
402 North Blackford Street
Indianapolis, Indiana 46202–3274,
  U.S.A.
lipkowitz@chem.iupui.edu

Donald B. Boyd
Department of Chemistry
Indiana University–Purdue University
  at Indianapolis
402 North Blackford Street
Indianapolis, Indiana 46202–3274,
  U.S.A.
boyd@chem.iupui.edu

Preface

After our first publisher produced our first volume and we were in the process of readying manuscripts for Volume 2, the publisher's executive editor innocently asked us if there was anything in the field of computational chemistry that we had not already covered in Volume 1. We assured him that there was much. The constancy of change was noted centuries ago when Honorat de Bueil, Marquis de Racan (1589–1670) observed that "Nothing in the world lasts, save eternal change." Science changes too. As stated by Emile Duclaux (1840–1904), French biologist and physician and successor to Louis Pasteur in heading the Pasteur Institute, "It is because science is sure of nothing that it is always advancing." Science is able to contribute to the well-being of mankind because it can evolve. Topics in a number of important areas of computational chemistry are the substance of this volume.

Cheminformatics, a term so new that scientists have not yet come to an agreement on how to spell it, is a facet of computational chemistry where the emphasis is on managing digital data and mining the data to extract knowledge. Cheminformatics holds a position at the intersection of several traditional disciplines including chemical information (library science), quantitative structure-property relationships, and computer science as it pertains to managing computers and databases. One powerful way to extract an understanding of the contents of a data set is with clustering methods, whereby the mutual proximity of data points is measured. Clustering can show how much similarity or diversity there is in a data set. Chapter 1 of this volume is a tutorial on clustering methods. The authors, Drs. Geoff M. Downs and John M. Barnard, were educated at the University of Sheffield—the veritable epicenter and fountainhead of cheminformatics. Each clustering method is described along with its strengths and weaknesses. As frequent consultants to pharmaceutical and chemical companies, the authors can knowledgeably point to published examples where real-world research problems were aided by one or more of the clustering methods.

The previous volume of our series, Volume 17, included a chapter on the use of docking for discovery of pharmaceutically interesting ligands. Employed in structure-based ligand design, docking requires a

three-dimensional structure of the receptor, which can be obtained from experiment or modeling. Docking also requires computational techniques for assessing the affinity of small organic molecules to a receptor. These techniques, collectively called scoring functions, attempt to quantitate the favorability of interaction in the ligand–receptor complex. In Chapter 2 of the present volume, Drs. Hans-Joachim Böhm and Martin Stahl give a tutorial on scoring functions. The authors share their considerable experience using scoring functions in drug discovery research at Roche, Basel. Scoring functions can be derived in different ways; they can be (1) based directly on standard force fields, (2) obtained by empirically fitting parameters in selected force field terms to reproduce a set of known binding affinities, or (3) derived by an inverse formulation of the Boltzmann law whereby the frequency of occurrence of an interatomic interaction is presumed to be related to the strength of that interaction. As with most modern computational methods used in pharmaceutical research, viable scoring functions must be quickly computable so that large numbers of ligand–receptor complexes can be evaluated at a speed comparable to the rate at which compounds can be synthesized by combinatorial chemistry. Despite efforts at numerous laboratories, the ''perfect'' scoring function, which would be both extremely accurate and broadly applicable, eludes scientists. Sometimes, several scoring functions can be tried on a given set of molecules, and then the computational chemist can look for a consensus in how the individual molecules are ranked by the scores.* A ligand structure having good scores does not guarantee that the compound will have high affinity when and if the compound is actually synthesized and tested. However, a structure with high rankings (i.e., fits the profile) is more likely to show binding than a randomly selected compound. Chapter 2 summarizes what has been learned about scoring functions and gives an example of how they have been applied to find new ligands in databases of real and/or conceivable (virtual) molecular structures stored on computers.

In the 1980s when computers were making molecular simulation calculations more feasible, computational chemists readily recognized that accounting for the polarizability of charge distribution in a molecule would become increasingly important for realistically modeling molecular systems. In most force fields, atomic charges are assigned at the beginning of the calculation and then are held fixed during the course of the minimization or simulation. However, we know that atomic charges vary with the electric field produced by the surrounding atoms. Each atom of a molecular system is in the field of all the other atoms; electrostatic interactions are long range (effective to as much as 14 Å), so a change in the molecular geometry will affect atomic charges,

---

*Such a consensus approach is reminiscent of what some computational chemists were doing in the the 1970s and 1980s when they were treating each molecule by not one, but several available semiempirical and ab initio molecular orbital methods, each of which gave different—and less than perfect—predictions of molecular properties.

especially if polar functional groups are present. In Chapter 3, Professors Steven W. Rick and Steven J. Stuart scrutinize the methods that have been devised to account for polarization. These methods include point dipole models, shell models, electronegativity equalization models, and semiempirical models. The test systems commonly used for developing and testing these models have been water, proteins, and nucleic acids. This chapter's comparison of computational models gives valuable guidance to users of molecular simulations.

In Chapter 4, Professors Dmitry V. Matyushov and Gregory A. Voth present a rigorous frontier report on the theory and computational methodologies for describing charge-transfer and electron-transfer reactions that can take place in condensed phases. This field of theory and computation aims to describe processes occurring, for instance, in biological systems and materials science. The chapter focuses on analysis of the activation barrier to charge transfer, especially as it relates to optical spectroscopy. Depending on the degeneracy of the energy states of the donor and acceptor, electron tunneling may occur. This chapter provides a step-by-step statistical mechanical development of the theory describing charge-transfer free energy surfaces. The Marcus–Hush mode of electron transfer consisting of two overlapping parabolas can be extended to the more general case of two free energy surfaces. In the last part of the chapter, the statistical mechanical analysis is applied to the calculation of optical profiles of photon absorption and emission, Franck–Condon factors, intensities, matrix elements, and chromophores.

In Chapter 5, Dr. George R. Famini and Professor Leland Y. Wilson teach about linear free energy relationships (LFERs) using molecular descriptors derived from—or adjuncts to—quantum chemical calculations. Basically, the LFER approach is a way of studying quantitative structure-property relationships (QSPRs). The property in question may be a physical one, such as vapor pressure or solvation free energy, or one related to biological activity (QSAR). Descriptors can be any numerical quantity—calculated or experimental—that represents all or part of a molecular structure. In the LFER approach, the number of descriptors used is relatively low compared to some QSPR/QSAR approaches that involve throwing so many descriptors into the regression analysis that the physical significance of any of these is obscured. These latter approaches are somewhat loosely referred to as "kitchen sink" approaches because the investigator has figuratively thrown everything into the equation including objects as odd as the proverbial kitchen sink. In the LFER approach, the descriptors include quantities that measure molecular dimensions (molecular volume, surface area, ovality), charge distributions (atomic charges, electrostatic potentials), electronic properties (ionization potential, polarizability), and thermodynamic properties (heat of formation). Despite use of the term "linear" in LFER, not all correlations encountered in the physical world are linear. QSPR/QSAR approaches based on regression analysis handle this situation by simply squaring—or taking some other power of—the values of

some descriptors and including them as separate independent variables in the regression equation. In this chapter, the authors discuss statistical procedures and give examples covering a wide variety of LFER applications. Quantum chemists can learn from this chapter how their methods may be employed in one of the most rapidly growing areas of computational chemistry, namely, QSAR.

In the nineteenth century, the world powerhouses of chemistry were Britain, France, and Germany. In Germany, Justus Liebig founded a chemistry research laboratory at the University of Giessen in 1825. At the University of Göttingen in 1828, Friedrich Wöhler was the first to synthesize an organic compound (urea) from inorganic material. In Karlsruhe, Friedrich August Kekulé organized the first international meeting on chemistry in 1860. Germany's dominance in the chemical and dye industry was legend well into the twentieth century. In the 1920s, German physicists played central roles in the development of quantum mechanics. Erwin Schrödinger formulated the wave function (1926). Werner Heisenberg formulated matrix mechanics (1925) and the uncertainty principle (1927). The German physicist at Göttingen, Max Born, together with the American, J. Robert Oppenheimer, published their oft-used famous approximation (1927). With such a strong background in chemistry and physics, it is not surprising that Germany was a fertile ground where computational chemistry could take root. The first fully automatic, programmable, digital computer was developed by an engineer in Berlin in 1930 for routine numerical calculations. After Germany was liberated from control of the National Socialist German Workers' Party ("Nazi"), peaceful scientific development could be taken up again, notwithstanding the enormous loss of many leading scientists who had fled from the Nazis. More computers were built, and theoretical chemists were granted access to them. In Chapter 6, Professor Dr. Sigrid D. Peyerimhoff masterfully traces the history of computational chemistry in Germany. This chapter complements the historical accounts covering the United States, Britain, France, and Canada, which were covered in prior volumes of this book series.

Finally, as a bonus with this volume, we editors present a perspective on the employment situation for computational chemists. The essay in the appendix reviews the history of the job market, uncovers factors that have affected it positively or negatively, and discusses the current situation. We also analyze recent job advertisements to see where recent growth has occurred and which skills are presently in greatest demand.

We invite our readers to visit the *Reviews in Computational Chemistry* website at http://chem.iupui.edu/rcc/rcc.html. It includes the author and subject indexes, color graphics, errata, and other materials supplementing the chapters. We are delighted to report that the Google search engine (http://www.google.com/) ranks our website among the top hits in a search on the term "computational chemistry". This search engine is becoming popular because it ranks hits in terms of their relevance and frequency of visits. Google

also is very fast and appears to provide a quite complete and up-to-date picture of what information is available on the World Wide Web.

We are also glad to note that our publisher has plans to make our most recent volumes available in an online form through Wiley InterScience. Please check the Web (http://www.interscience.wiley.com/onlinebooks) or contact reference@wiley.com for the latest information. For readers who appreciate the permanence and convenience of bound books, these will, of course, continue.

We thank the authors of this volume for their excellent chapters. Mrs. Joanne Hequembourg Boyd is acknowledged for editorial assistance.

<div align="right">

Donald B. Boyd and Kenny B. Lipkowitz
Indianapolis
January 2002

</div>

Epilogue and Dedication

My association with Ken Lipkowitz began a couple of years after he arrived in Indianapolis in 1977. Ken, trained as an synthetic organic chemist, was a new young assistant professor at Indiana University–Purdue University Indianapolis, and I was a research scientist at Eli Lilly & Company, where I, a quantum chemist by training, had been working in the field of computer-aided drug design for nine years. Ken approached me to learn about computational chemistry. I was glad to help him, and he was an enthusiastic "student". Our first paper together was published in 1980. Unsure whether his career as a fledging computational chemist would lead anywhere, he made a distinction in this and other papers he wrote between his organic persona (Kenneth B. Lipkowitz) and his computational persona (Kenny B. Lipkowitz). Over the subsequent years, he focused his career more and more on computational chemistry and established himself as a highly productive and creative scientist. He has always been a hard-working, amiable, and obliging collaborator and friend.

In the late 1980s, Ken had the idea of initiating a book series on computational chemistry. The field was starting to come into full blossom, but few books for it were being published. Whereas review series on other subjects tended to be of mainly archival value and to remain on library shelves, his inspiration for *Reviews in Computational Chemistry* was to include as many tutorial chapters as possible, so that the books would be more used for teaching and individual study. The chapters would be ones that a professor could give new graduate students to bring them up to speed in a particular topic. The chapters would also be substantive, so that the books would not be just a journal with hard covers. As much as possible, the contents of the books would be material that could not be found in any other source. Ken persuaded me to join him in this endeavor.

I have viewed an editor's prime duties to set high standards and to heed the needs of both readers and authors. Hence, every effort has been made to produce volumes of the highest quality. It has been a keen pleasure working with authors who take exceptional pride in their workmanship. The expertise and hard work of many authors have been essential for producing books of

sustained usefulness in learning, teaching, and research. With this volume, the eighteenth, more than 7300 pages have been published since the series began in 1990. More than 200 authors have contributed the chapters. Appreciating the value of these chapters, scientists and libraries around the world have purchased more than 13,000 copies of the books since the series began.

My vision of computational chemistry, as embodied in this book series as well as in the Gordon Conference on Computational Chemistry that I initiated, was that there were synergies to be gained by juxtaposing all the various methodologies available to computational chemists. Thus, computational chemistry is more than quantum chemistry, more than molecular modeling, more than simulations, more than molecular design. Versatility is possible when scientists can draw from their toolbox the most appropriate methodologies for modeling molecules and data. Important goals of this book series have been to nurture the development of the field of computational chemistry, advance its recognition, strengthen its foundations, expand its dimensions, aid practitioners working in the field, and assist newcomers wanting to enter the field.

However, it is now time for me to rest my keyboard-weary hands. I wish Ken and his new co-editors every success as the book series continues. Ken could not have paid me a higher compliment than by enlisting not one, but two, excellent people to carry on the work I did. I have every confidence that as computational chemistry continues to evolve, its spectrum of methods and applications will further expand and increase in brilliance.

Dedication

With completion of this, my final, volume, I am reminded of my blessings to live in a country conceived by the Founding Fathers of the United States of America. Nothing would have been possible for me without the selflessness and devotion of Howard Milton Boyd, Ph.G., B.S., M.S. Nothing would have been worthwhile without the following:

<div align="center">

Andy
Cynthia
Douglas
Drew
Elisabeth
Emma
Joanne
Mary
Richard
Susanne

</div>

<div align="right">

Donald B. Boyd
Indianapolis
January 2002

</div>

Contents

Appendix. Examination of the Employment Environment for Computational Chemistry   **293**

Donald B. Boyd and Kenny B. Lipkowitz

Contributors

John M. Barnard, Barnard Chemical Information Ltd., 46 Uppergate Road, Stannington, Sheffield S6 6BX, United Kingdom (Electronic mail: barnard@bci.gb.com)

Hans-Joachim Böhm, F. Hoffmann-La Roche AG, Pharmaceuticals Division, Chemical Technologies, CH-4070 Basel, Switzerland (Electronic mail: hans-joachim.boehm@roche.com)

Donald B. Boyd, Department of Chemistry, Indiana University–Purdue University at Indianapolis (IUPUI), 402 North Blackford Street, Indianapolis, Indiana 46202-3274, U.S.A. (Electronic mail: boyd@chem.iupui.edu)

Geoff M. Downs, Barnard Chemical Information Ltd., 46 Uppergate Road, Stannington, Sheffield S6 6BX, United Kingdom (Electronic mail: geoff@bci.gb.com)

George R. Famini, RDA International Cooperative Programs Division, United States Army Soldier and Biological Chemical Command, 5183 Blackhawk Road, Aberdeen Proving Ground, Maryland 21010-5424, U.S.A. (Electronic mail: george.famini@sbccom.apgea.army.mil)

Kenny B. Lipkowitz, Department of Chemistry, Indiana University–Purdue University at Indianapolis (IUPUI), 402 North Blackford Street, Indianapolis, Indiana 46202-3274, U.S.A. (Electronic mail: lipkowitz@chem.iupui.edu)

Dmitry V. Matyushov, Department of Chemistry and Biochemistry, Arizona State University, P. O. Box 871604, Tempe, Arizona 85287-1604, U.S.A. (Electronic mail: dmitrym@asu.edu)

Sigrid D. Peyerimhoff, Institut für Physikalische und Theoretische Chemie, Universität Bonn, Wegelerstrasse 12, D-53115 Bonn, Germany (Electronic mail: unt000@uni-bonn.de)

Steven W. Rick, Department of Chemistry, University of New Orleans, New Orleans, Louisiana 70148, U.S.A. (Electronic mail: srick@uno.edu)

Martin Stahl, F. Hoffmann-La Roche AG, Pharmaceuticals Division, Chemical Technologies, CH-4070 Basel, Switzerland (Electronic mail: martin.stahl@roche.com)

Steven J. Stuart, Department of Chemistry, Hunter Laboratory, Clemson University, Clemson, South Carolina 29634-0973, U.S.A. (Electronic mail: ss@clemson.edu)

Gregory A. Voth, Department of Chemistry and Henry Eyring Center for Theoretical Chemistry, University of Utah, 315 South 1400 East, Salt Lake City, Utah 84112-0850, U.S.A. (Electronic mail: voth@chemistry.utah.edu)

Leland Y. Wilson, Department of Chemistry and Biochemistry, La Sierra University, Riverside, California 92515, U.S.A. (Electronic mail: hlwilson@ urs2.net)

Contributors to Previous Volumes*

Volume 1 (1990)

David Feller and **Ernest R. Davidson,**[†] Basis Sets for Ab Initio Molecular Orbital Calculations and Intermolecular Interactions.

James J. P. Stewart,[‡] Semiempirical Molecular Orbital Methods.

Clifford E. Dykstra,[¶] **Joseph D. Augspurger, Bernard Kirtman,** and **David J. Malik,** Properties of Molecules by Direct Calculation.

Ernest L. Plummer, The Application of Quantitative Design Strategies in Pesticide Design.

Peter C. Jurs, Chemometrics and Multivariate Analysis in Analytical Chemistry.

Yvonne C. Martin, Mark G. Bures, and **Peter Willett,** Searching Databases of Three-Dimensional Structures.

Paul G. Mezey, Molecular Surfaces.

Terry P. Lybrand,[§] Computer Simulation of Biomolecular Systems Using Molecular Dynamics and Free Energy Perturbation Methods.

Donald B. Boyd, Aspects of Molecular Modeling.

---

*Where appropriate and available, the current affiliation of the senior or corresponding author is given here as a convenience to our readers.

[†]Current address: Department of Chemistry, University of Washington, Seattle, Washington 98195.

[‡]Current address: 15210 Paddington Circle, Colorado Springs, Colorado 80921-2512 (Electronic mail: jstewart@fai.com).

[¶]Current address: Department of Chemistry, Indiana University–Purdue University at Inidanapolis, Indianapolis, Indiana 46202 (Electronic mail: dykstra@chem.iupui.edu).

[§]Current address: Department of Chemistry, Vanderbilt University, Nashville, Tennessee 37212 (Electronic mail: lybrand@structbio.vanderbilt.edu).

Donald B. Boyd, Successes of Computer-Assisted Molecular Design.

Ernest R. Davidson, Perspectives on Ab Initio Calculations.

Volume 2   (1991)

Andrew R. Leach,* A Survey of Methods for Searching the Conformational Space of Small and Medium-Sized Molecules.

John M. Troyer and **Fred E. Cohen,** Simplified Models for Understanding and Predicting Protein Structure.

J. Phillip Bowen and **Norman L. Allinger,** Molecular Mechanics: The Art and Science of Parameterization.

Uri Dinur and **Arnold T. Hagler,** New Approaches to Empirical Force Fields.

Steve Scheiner,[†] Calculating the Properties of Hydrogen Bonds by Ab Initio Methods.

Donald E. Williams, Net Atomic Charge and Multipole Models for the Ab Initio Molecular Electric Potential.

Peter Politzer and **Jane S. Murray,** Molecular Electrostatic Potentials and Chemical Reactivity.

Michael C. Zerner, Semiempirical Molecular Orbital Methods.

Lowell H. Hall and **Lemont B. Kier,** The Molecular Connectivity Chi Indexes and Kappa Shape Indexes in Structure–Property Modeling.

I. B. Bersuker[‡] and **A. S. Dimoglo,** The Electron-Topological Approach to the QSAR Problem.

Donald B. Boyd, The Computational Chemistry Literature.

---

[*]Current address: GlaxoSmithKline, Greenford, Middlesex, UB6 0HE, U.K. (Electronic mail: arl22958@ggr.co.uk).

[†]Current address: Department of Chemistry and Biochemistry, Utah State University, Logan, Utah 84322 (Electronic mail: scheiner@cc.usu.edu).

[‡]Current address: College of Pharmacy, The University of Texas, Austin, Texas 78712 (Electronic mail: bersuker@eeyore.cm.utexas.edu).

Volume 3 (1992)

Tamar Schlick, Optimization Methods in Computational Chemistry.

Harold A. Scheraga, Predicting Three-Dimensional Structures of Oligopeptides.

Andrew E. Torda and **Wilfred F. van Gunsteren,** Molecular Modeling Using NMR Data.

David F. V. Lewis, Computer-Assisted Methods in the Evaluation of Chemical Toxicity.

Volume 4 (1993)

Jerzy Cioslowski, Ab Initio Calculations on Large Molecules: Methodology and Applications.

Michael L. McKee and **Michael Page,** Computing Reaction Pathways on Molecular Potential Energy Surfaces.

Robert M. Whitnell and **Kent R. Wilson,** Computational Molecular Dynamics of Chemical Reactions in Solution.

Roger L. DeKock, Jeffry D. Madura, Frank Rioux, and **Joseph Casanova,** Computational Chemistry in the Undergraduate Curriculum.

Volume 5 (1994)

John D. Bolcer and **Robert B. Hermann,** The Development of Computational Chemistry in the United States.

Rodney J. Bartlett and **John F. Stanton,** Applications of Post-Hartree–Fock Methods: A Tutorial.

Steven M. Bachrach,* Population Analysis and Electron Densities from Quantum Mechanics.

*Current address: Department of Chemistry, Trinity University, San Antonio, Texas 78212 (Electronic mail: steven.bachrach@trinity.edu).

Jeffry D. Madura, * **Malcolm E. Davis, Michael K. Gilson, Rebecca C. Wade, Brock A. Luty,** and **J. Andrew McCammon,** Biological Applications of Electrostatic Calculations and Brownian Dynamics Simulations.

K. V. Damodaran and **Kenneth M. Merz Jr.,** Computer Simulation of Lipid Systems.

Jeffrey M. Blaney[†] and **J. Scott Dixon,** Distance Geometry in Molecular Modeling.

Lisa M. Balbes, S. Wayne Mascarella, and **Donald B. Boyd,** A Perspective of Modern Methods in Computer-Aided Drug Design.

Volume 6   (1995)

Christopher J. Cramer and **Donald G. Truhlar,** Continuum Solvation Models: Classical and Quantum Mechanical Implementations.

Clark R. Landis, Daniel M. Root, and **Thomas Cleveland,** Molecular Mechanics Force Fields for Modeling Inorganic and Organometallic Compounds.

Vassilios Galiatsatos, Computational Methods for Modeling Polymers: An Introduction.

Rick A. Kendall,[‡] **Robert J. Harrison, Rik J. Littlefield,** and **Martyn F. Guest,** High Performance Computing in Computational Chemistry: Methods and Machines.

Donald B. Boyd, Molecular Modeling Software in Use: Publication Trends.

Eiji Ōsawa and **Kenny B. Lipkowitz,** Appendix: Published Force Field Parameters.

[*]Current address: Department of Chemistry and Biochemistry, Duquesne University, Pittsburgh, Pennsylvania 15282-1530 (Electronic mail: madura@duq.edu).

[†]Current address: Structural GenomiX, 10505 Roselle St., San Diego, California 92120 (Electronic mail: jeff_blaney@stromix.com).

[‡]Current address: Scalable Computing Laboratory, Ames Laboratory, Wilhelm Hall, Ames, Iowa 50011 (Electronic mail: rickyk@scl.ameslab.gov).

Volume 7 (1996)

Geoffrey M. Downs and **Peter Willett,** Similarity Searching in Databases of Chemical Structures.

Andrew C. Good* and **Jonathan S. Mason,** Three-Dimensional Structure Database Searches.

Jiali Gao,[†] Methods and Applications of Combined Quantum Mechanical and Molecular Mechanical Potentials.

Libero J. Bartolotti and **Ken Flurchick,** An Introduction to Density Functional Theory.

Alain St-Amant, Density Functional Methods in Biomolecular Modeling.

Danya Yang and **Arvi Rauk,** The A Priori Calculation of Vibrational Circular Dichroism Intensities.

Donald B. Boyd, Appendix: Compendium of Software for Molecular Modeling.

Volume 8 (1996)

Zdenek Slanina,[‡] **Shyi-Long Lee,** and **Chin-hui Yu,** Computations in Treating Fullerenes and Carbon Aggregates.

Gernot Frenking, Iris Antes, Marlis Böhme, Stefan Dapprich, Andreas W. Ehlers, Volker Jonas, Arndt Neuhaus, Michael Otto, Ralf Stegmann, Achim Veldkamp, and **Sergei F. Vyboishchikov,** Pseudopotential Calculations of Transition Metal Compounds: Scope and Limitations.

Thomas R. Cundari, Michael T. Benson, M. Leigh Lutz, and **Shaun O. Sommerer,** Effective Core Potential Approaches to the Chemistry of the Heavier Elements.

---

[*]Current address: Bristol-Myers Squibb, 5 Research Parkway, P.O. Box 5100, Wallingford, Connecticut 06492-7660 (Electronic mail: andrew.good@bms.com).

[†]Current address: Deperment Chemistry, University of Minnesota, 207 Pleasant St. SE, Minneapolis, Minnesota 55455-0431 (Electronic mail: gao@chem.umn.edu).

[‡]Current address: Institute of Chemistry, Academia Sinica, Nankang, Taipei 11529, Taiwan, Republic of China (Electronic mail: zdenek@chem.sinica.edu.tw).

Jan Almlöf and **Odd Gropen,*** Relativistic Effects in Chemistry.

Donald B. Chesnut, The Ab Initio Computation of Nuclear Magnetic Resonance Chemical Shielding.

Volume 9   (1996)

James R. Damewood Jr., Peptide Mimetic Design with the Aid of Computational Chemistry.

T. P. Straatsma, Free Energy by Molecular Simulation.

Robert J. Woods, The Application of Molecular Modeling Techniques to the Determination of Oligosaccharide Solution Conformations.

Ingrid Pettersson and **Tommy Liljefors,** Molecular Mechanics Calculated Conformational Energies of Organic Molecules: A Comparison of Force Fields.

Gustavo A. Arteca, Molecular Shape Descriptors.

Volume 10   (1997)

Richard Judson,[†] Genetic Algorithms and Their Use in Chemistry.

Eric C. Martin, David C. Spellmeyer, Roger E. Critchlow Jr., and **Jeffrey M. Blaney,** Does Combinatorial Chemistry Obviate Computer-Aided Drug Design?

Robert Q. Topper, Visualizing Molecular Phase Space: Nonstatistical Effects in Reaction Dynamics.

Raima Larter and **Kenneth Showalter,** Computational Studies in Nonlinear Dynamics.

Stephen J. Smith and **Brian T. Sutcliffe,** The Development of Computational Chemistry in the United Kingdom.

---

[*]Address: Institute of Mathematical and Physical Sciences, University of Tromsø, N-9037 Tromsø, Norway (Electronic mail: oddg@chem.uit.no).

[†]Current address: Genaissance Pharmaceuticals, Five Science Park, New Haven, Connecticut 06511 (Electronic mail: r.judson@genaissance.com).

Volume 11   (1997)

Mark A. Murcko, Recent Advances in Ligand Design Methods.

David E. Clark,* **Christopher W. Murray,** and **Jin Li,** Current Issues in De Novo Molecular Design.

Tudor I. Oprea† and **Chris L. Waller,** Theoretical and Practical Aspects of Three-Dimensional Quantitative Structure–Activity Relationships.

Giovanni Greco, Ettore Novellino, and **Yvonne Connolly Martin,** Approaches to Three-Dimensional Quantitative Structure–Activity Relationships.

Pierre-Alain Carrupt, Bernard Testa, and **Patrick Gaillard,** Computational Approaches to Lipophilicity: Methods and Applications.

Ganesan Ravishanker, Pascal Auffinger, David R. Langley, Bhyravabhotla Jayaram, Matthew A. Young, and **David L. Beveridge,** Treatment of Counterions in Computer Simulations of DNA.

Donald B. Boyd, Appendix: Compendium of Software and Internet Tools for Computational Chemistry.

Volume 12   (1998)

Hagai Meirovitch,‡ Calculation of the Free Energy and the Entropy of Macromolecular Systems by Computer Simulation.

Ramzi Kutteh and **T. P. Straatsma,** Molecular Dynamics with General Holonomic Constraints and Application to Internal Coordinate Constraints.

John C. Shelley¶ and **Daniel R. Bérard,** Computer Simulation of Water Physisorption at Metal–Water Interfaces.

---

*Current address: Computer-Aided Drug Design, Argenta Discovery Ltd., 8/9 Spire Green Centre, Flex Meadow, Harlow, Essex, CM19 5TR, United Kingdom (Electronic mail: david.clark@argentadiscovery.com).

†Current address: Office of Biocomputing, University of New Mexico School of Medicine, 915 Camino de Salud NE, Albuquerque, New Mexico 87131 (Electronic mail: toprea@salud.unm.edu).

‡Current address: Department of Molecular Genetics & Biochemistry, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania 15213 (Electronic mail: hagaim@pitt.edu).

¶Current address: Schrödinger, Inc., 1500 S.W. First Avenue, Suite 1180, Portland, Oregon 97201 (Electronic mail: jshelley@schrodinger.com).

Donald W. Brenner, Olga A. Shenderova, and **Denis A. Areshkin,** Quantum-Based Analytic Interatomic Forces and Materials Simulation.

Henry A. Kurtz and **Douglas S. Dudis,** Quantum Mechanical Methods for Predicting Nonlinear Optical Properties.

Chung F. Wong,* **Tom Thacher,** and **Herschel Rabitz,** Sensitivity Analysis in Biomolecular Simulation.

Paul Verwer and **Frank J. J. Leusen,** Computer Simulation to Predict Possible Crystal Polymorphs.

Jean-Louis Rivail and **Bernard Maigret,** Computational Chemistry in France: A Historical Survey.

Volume 13 (1999)

Thomas Bally and **Weston Thatcher Borden,** Calculations on Open-Shell Molecules: A Beginner's Guide.

Neil R. Kestner and **Jaime E. Combariza,** Basis Set Superposition Errors: Theory and Practice.

James B. Anderson, Quantum Monte Carlo: Atoms, Molecules, Clusters, Liquids, and Solids.

Anders Wallqvist[†] and **Raymond D. Mountain,** Molecular Models of Water: Derivation and Description.

James M. Briggs and **Jan Antosiewicz,** Simulation of pH-dependent Properties of Proteins Using Mesoscopic Models.

Harold E. Helson, Structure Diagram Generation.

Volume 14 (2000)

Michelle Miller Francl and **Lisa Emily Chirlian,** The Pluses and Minuses of Mapping Atomic Charges to Electrostatic Potentials.

---

*Current addrress: Howard Hughes Medical Institute, School of Medicine, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093-0365 (Electronic mail: c4wong@ucsd.edu).

[†]Current address: National Cancer Institute, P.O. Box B, Frederick, Maryland 21702 (Electronic mail: wallqvist@ncifcrt.gov).

T. Daniel Crawford* and **Henry F. Schaefer III,** An Introduction to Coupled Cluster Theory for Computational Chemists.

Bastiaan van de Graaf, Swie Lan Njo, and **Konstantin S. Smirnov,** Introduction to Zeolite Modeling.

Sarah L. Price, Toward More Accurate Model Intermolecular Potentials for Organic Molecules.

Christopher J. Mundy,† **Sundaram Balasubramanian, Ken Bagchi, Mark E. Tuckerman, Glenn J. Martyna,** and **Michael L. Klein,** Nonequilibrium Molecular Dynamics.

Donald B. Boyd and **Kenny B. Lipkowitz,** History of the Gordon Research Conferences on Computational Chemistry.

Mehran Jalaie and **Kenny B. Lipkowitz,** Appendix: Published Force Field Parameters for Molecular Mechanics, Molecular Dynamics, and Monte Carlo Simulations.

Volume 15 (2000)

F. Matthias Bickelhaupt and **Evert Jan Baerends,** Kohn–Sham Density Functional Theory: Predicting and Understanding Chemistry.

Michael A. Robb, Marco Garavelli, Massimo Olivucci, and **Fernando Bernardi,** A Computational Strategy for Organic Photochemistry.

Larry A. Curtiss, Paul C. Redfern, and **David J. Frurip,** Theoretical Methods for Computing Enthalpies of Formation of Gaseous Compounds.

Russell J. Boyd, The Development of Computational Chemistry in Canada.

Volume 16 (2000)

Richard A. Lewis, Stephen D. Pickett, and **David E. Clark,** Computer-Aided Molecular Diversity Analysis and Combinatorial Library Design.

---

*Current address: Department of Chemistry, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061-0212 (Electronic mail: crawdad@vt.edu).

†Current address: Computational Materials Science, L-371, Lawrence Livermore National Laboratory, Livermore, California 94550 (Electronic mail: mundy2@llnl.gov).

Keith L. Peterson, Artificial Neural Networks and Their Use in Chemistry.

Jörg-Rüdiger Hill, Clive M. Freeman, and **Lalitha Subramanian,** Use of Force Fields in Materials Modeling.

M. Rami Reddy, Mark D. Erion, and **Atul Agarwal,** Free Energy Calculations: Use and Limitations in Predicting Ligand Binding Affinities.

Volume 17 (2001)

Ingo Muegge and **Matthias Rarey,** Small Molecule Docking and Scoring.

Lutz P. Ehrlich and **Rebecca C. Wade,** Protein–Protein Docking.

Christel M. Marian, Spin-Orbit Coupling in Molecules.

Lemont B. Kier, Chao-Kun Cheng, and **Paul G. Seybold,** Cellular Automata Models of Aqueous Solution Systems.

Kenny B. Lipkowitz and **Donald B. Boyd,** Appendix: Books Published on the Topics of Computational Chemistry.

Topics Covered in Volumes 1–18*

Ab initio calculations, **1, 2, 3, 8, 13, 15**

Basis set superposition errors, **13**
Basis sets, **1**
Brownian dynamics, **5**

Cellular automata modeling, **17**
Charge distributions, **2, 5, 14**
Charge transfer, **18**
Cheminformatics, **1, 7, 18**
Chemometrics, **1**
Clustering methods, **18**
Computational chemistry literature, **2, 17**
Conformational analysis, **2, 9**
Coupled cluster theory, **14**
Crystal polymorphs, **12**

Databases of structures, **1, 7**
De novo design, **11**
Density functional theory, **7, 13, 15**
Descriptors, **1, 9, 18**
Distance geometry, **5**
Diversity analysis, **10, 16**
DNA, **11**
Docking, **16, 18**
Drug discovery, **1, 5, 9, 10, 11, 16, 18**

Effective core potentials, **8**
Electrostatics, **2, 5, 14**
Enthalpies of formation, **14**
Entropy, **12**

Force fields, **2, 6, 9, 14, 16**
Free energy calculations, **1, 9, 12, 16**
Fullerenes, **8**

Genetic algorithms, **10**

High performance computing, **6**

History of development of the field, **1, 10, 12, 14, 15, 18**
Hydrogen bonding, **2**

Inorganics, **6**
Intermolecular interactions, **1, 14**

Job market, **18**

Kohn-Sham orbitals, **15**

Library design, **10, 16**
Ligand design, **1, 5, 9, 10, 11, 16, 18**
Linear free energy relationships, **18**
Lipids, **6**
Lipophilicity, **11**

Materials modeling, **6, 8, 12, 14, 16**
Mesoscopic models, **13**
Metal–water interfaces, **12**
Molecular connectivity, **2**
Molecular design, **1, 5**
Molecular diversity, **16**
Molecular dynamics, **1, 4, 5, 9, 11, 12, 14, 18**
Molecular mechanics, **2, 6, 9, 14**
Molecular modeling, **1, 3, 9**
Molecular orbital calculations, **1, 2, 3, 6, 8, 13**
Molecular properties, **1**
Molecular shape, **9**
Molecular similarity, **7, 16**
Molecular simulations, **1, 4, 5, 9, 11, 12, 14, 18**
Molecular surfaces, **1**

Neural networks, **16**
Nonequilibrium molecular dynamics, **14**
Nonlinear dynamics, **10**

*This brief index lists topics and volume numbers.

Reviews in
Computational
Chemistry
Volume 18

CHAPTER 1

Clustering Methods and Their Uses in Computational Chemistry

Geoff M. Downs and John M. Barnard

Barnard Chemical Information Ltd., 46 Uppergate Road, Stannington, Sheffield S6 6BX, United Kingdom

INTRODUCTION

Clustering is a data analysis technique that, when applied to a set of heterogeneous items, identifies homogeneous subgroups as defined by a given model or measure of similarity. Of the many uses of clustering, a prime motivation for the increasing interest in clustering methods is their use in the selection and design of combinatorial libraries of chemical structures pertinent to pharmaceutical discovery.

One feature of clustering is that the process is unsupervised, that is, there is no predefined grouping that the clustering seeks to reproduce. In contrast to supervised learning, where the task is to establish relationships between given inputs and outputs to enable prediction of the output from new inputs, in unsupervised learning only the inputs are available and the task is to reveal aspects of the underlying distribution of the input data. Clustering is thus complemented by the related supervised process of classification, in which items are assigned labels applied to predefined groups: examples include recursive partitioning, naïve Bayesian analysis, and K nearest-neighbor selection. Clustering is a technique for exploratory data analysis and is used increasingly in preliminary analyses of large data sets of medium and high dimensionality as a method of selection, diversity analysis, and data reduction. This chapter reviews the main clustering methods that are used for analyzing chemical

data sets and gives examples of their application in pharmaceutical companies. Compared to the other costs of drug discovery, clustering can add significant value at minimal cost. First, we provide an outline of clustering as a discipline and define some of the terminology. Then, we give a brief tutorial on clustering algorithms, review progress in developing the methods, and offer some example applications.

Clustering methodology has been developed and used in a variety of areas including archaeology, astronomy, biology, computer science, electronics, engineering, information science, and medicine. Good, general introductory texts on the topic of clustering include those by Sneath and Sokal,[1] Kaufmann and Rousseeuw,[2] Everitt,[3] and Gordon.[4] The main text that is devoted to clustering of chemical data sets is by Willett,[5] with review articles by Bratchell,[6] Barnard and Downs,[7] and Downs and Willett.[8] The present chapter is a complement and update to the latter article. In a previous volume of this series, Lewis, Pickett, and Clark[9] reviewed the use of diversity analysis techniques in combinatorial library design.

As will be shown in the section on Chemical Applications, the current main uses of clustering for chemical data sets are to find representative subsets from high throughput screening (HTS) and combinatorial chemistry, and to increase the diversity of in-house data sets through selection of additional compounds from other data sets. Methods suitable for compound selection are the main focus of this chapter. The methods must be able to handle large data sets of high-dimensional data. For small, low-dimensional data sets, most clustering methods are applicable, and descriptions in the standard texts and implementations available in standard statistical software packages[10,11] suffice. Implementations designed for use on chemical data sets are available from most of the specialist software vendors,[12–17] the majority of which were reviewed by Warr.[18]

The overall process of clustering involves the following steps:

1. Generate appropriate descriptors for each compound in the data set.
2. Select an appropriate similarity measure.
3. Use an appropriate clustering method to cluster the data set.
4. Analyze the results.

This chapter focuses on step 3. For step 1, descriptors may include property values, biological properties, topological indexes, and structural fragments. The performance of these descriptors and forms of representation have been analyzed by Brown[19] and Brown and Martin.[20,21] Similarity searching for step 2 has been discussed by Downs and Willett;[22] characteristics of various similarity measures have been discussed by Barnard, Downs, and Willett.[23,24] For step 4, little has been published specifically about visualization and analysis of results for chemical data sets. However, most publications that focus on implementing systems that utilize clustering do provide details of how the results were displayed or analyzed.

The terminology associated with clustering is extensive, with many terms used to describe the same thing (reflecting the separate development of clustering methods within a multitude of disciplines). Clusters can be *overlapping* or *nonoverlapping*; if a compound occurs in more than one cluster, the clusters are overlapping. At one extreme, each compound is a member of all clusters to a certain degree. An example of this is *fuzzy* clustering in which the degree of membership of an individual compound is in the range 0 to 1, and the total membership summed across all clusters is normally required to be 1. This scheme contrasts with *crisp* clustering in which each compound's degree of membership in any cluster is either 0 or 1. At the other extreme, is the situation wherein each compound is a member of exactly one cluster, in which case the clusters are said to be nonoverlapping. Intermediate situations sometimes occur, where compounds can be members of several, though not of all, clusters. The majority of clustering methods used on chemical data sets generate crisp, nonoverlapping clusters, because analysis of such clusters is relatively simple.

If a data set is analyzed in an iterative way, such that at each step a pair of clusters is merged or a single cluster is divided, the result is *hierarchical*, with a parent–child relationship being established between clusters at each successive level of the iteration. The successive levels can be visualized using a dendrogram, as shown in Figure 1. Each level of the hierarchy represents a partitioning of the data set into a set of clusters. In contrast, if the data set is analyzed to produce a single partition of the compounds resulting in a set of clusters, the result is then *nonhierarchical*. Note that the term *partitioning*
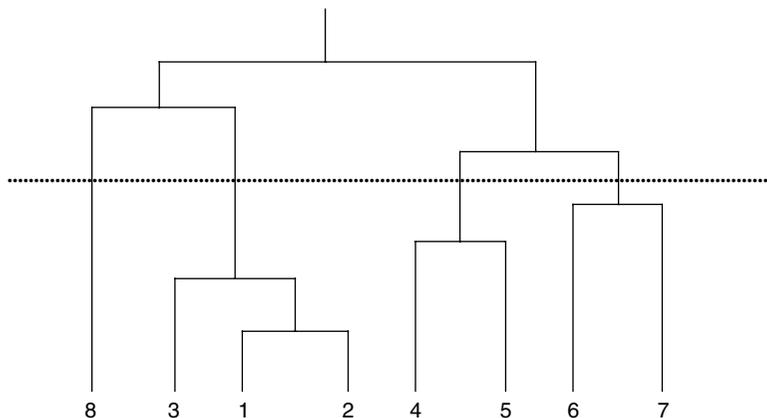
Figure 1 An example of a hierarchy (dendrogram) generated from the clustering of eight items (shown numbered 1–8 across the bottom). The top (root) is a single cluster containing all eight items. The vertical positions of the horizontal lines joining pairs of items or cluster indicate the relative similarities of those pairs. Items 1 and 2 are the most similar and clusters [8,3,1,2] and [4,5,6,7] are the least similar. The dotted horizontal line represents a single partition containing the four clusters [8], [3,1,2], [4,5], and [6,7].

in this context is different from the technique of partitioning (otherwise known as cell-based partitioning). The latter technique is a method of classification rather than of clustering, and a useful review of it, as applied to chemical data sets, is given by Mason and Pickett.[25] A broad classification of the most common clustering methods is shown in Figure 2. Note that, with the wide range of clustering methods devised, some can be placed in more than one of the given categories.

If a hierarchical method starts with all compounds as *singletons* (in clusters by themselves) and the latter are merged iteratively until all compounds are in a single cluster, the method is said to be *agglomerative*. With respect to the dendrogram in Figure 1, it is a bottom-up approach. If the hierarchical method starts with all compounds in a single cluster and iteratively splits one cluster into two until all compounds are singletons, the method is *divisive*, that
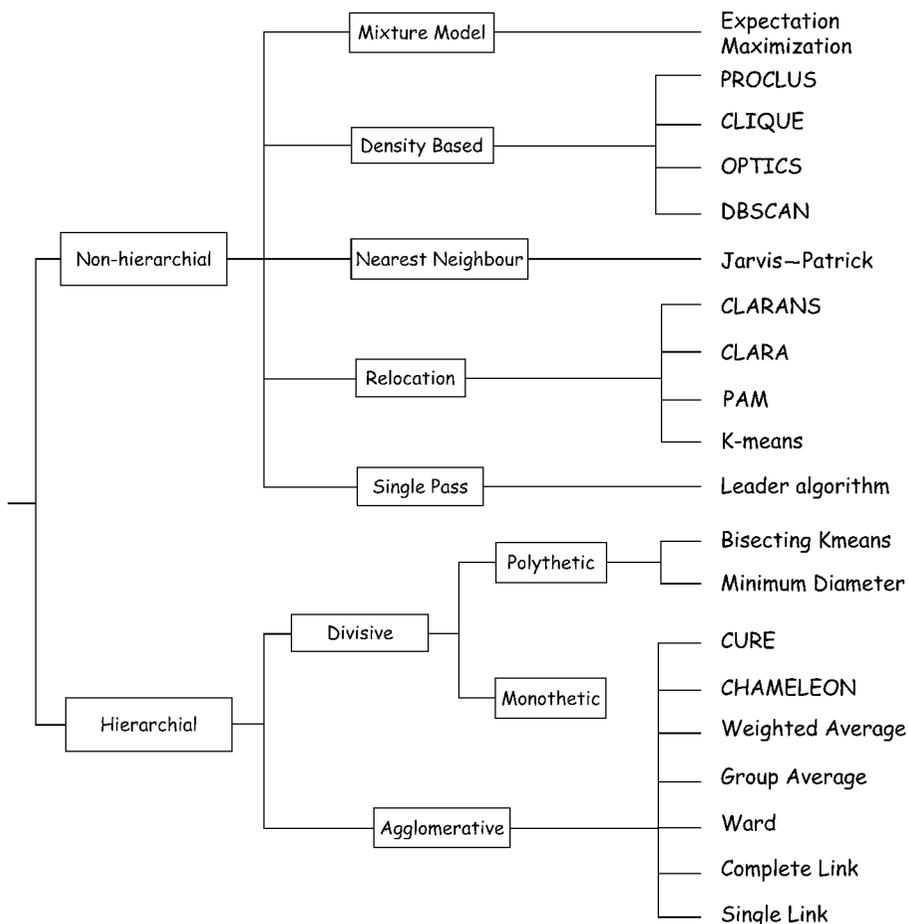
Figure 2 A broad classification of the most common clustering methods.

is, it is a top-down approach. If, at each split, only one descriptor is used to determine how the cluster is split, the method is *monothetic*; otherwise, more descriptors (typically all available) are used, and the method is *polythetic*.

Nonhierarchical methods encompass a wide range of different techniques to build clusters. A *single-pass* method is one in which the partition is created by a single pass through the data set or, if randomly accessed, in which each compound is examined only once to decide which cluster it should be assigned to. A *relocation* method is one in which compounds are moved from one cluster to another to try to improve on the initial estimation of the clusters. The relocating is typically accomplished based on improving a cost function describing the "goodness" of each resultant cluster. The *nearest-neighbor* approach is more compound centered than are the other nonhierarchical methods. In it, the environment around each compound is examined in terms of its most similar neighboring compounds, with commonality between nearest neighbors being used as a criterion for cluster formation. In *mixture model* clustering the data are assumed to exist as a mixture of densities that are usually assumed to be Gaussian (normal) distributions, since their densities are not known in advance. Solutions to the mixture model are derived iteratively in a manner similar to the relocation methods. *Topographic* methods, such as use of Kohonen maps, typically apply a variable cost function with the added restriction that topographic relationships are preserved so that neighboring clusters are close in descriptor space. Other nonhierarchical methods include *density-based* and *probabilistic* methods. Density-based, or mode-seeking, methods regard the distribution of descriptors across the data set as generating patterns of high and low density that, when identified, can be used to separate the compounds into clusters. Probabilistic clustering generates nonoverlapping clusters in which a compound is assigned a probability, in the range 0 to 1, that it belongs to the chosen cluster (in contrast to fuzzy clustering in which the clusters are overlapping and the degree of membership is not a probability).

Having now provided a broad overview of clustering methodology, we next focus on the "classical" methods, which include hierarchical and single-pass, relocation, and nearest-neighbor nonhierarchical techniques. The classification we have described in Figure 2 is one that is commonly used by many scientists; however, it is just one of many possible classifications. Another way to differentiate between clustering techniques is to consider *parametric* and *nonparametric* methods. Parametric methods require distance-based comparisons be made. Here access to the descriptors is required (typically given as Euclidean vectors), rather than just a proximity matrix derived from the descriptors. Parametric methods can be further organized into *generative* and *reconstructive* methods. Generative methods, including mixture model, density-based, and probabilistic techniques, try to match parameters (e.g., cluster centers, variances within and between clusters, and mixing coefficients for the descriptor distributions) to the distribution of descriptors within the data set. Reconstructive methods, such as relocation and topographic, are

based upon improving a given cost function. Nonparametric methods make fewer assumptions about the underlying data; they do not adapt given parameters iteratively and, in general, need only a matrix of pairwise proximities (i.e., a distance matrix).

The term proximity is used here to include similarity and dissimilarity coefficients in addition to distance measures. Individual proximity measures are not defined in this review; full definitions can be found in standard texts and in the articles by Barnard, Downs, and Willett.[23,24] We now define the terms *centroid* and *square-error*, because they will be used throughout this chapter. For a cluster of s compounds each represented by a vector, let $\mathbf{x}(r)$ be the rth vector. The vector of the cluster centroid, $\mathbf{x}(c)$, is then defined as

$$\mathbf{x}(c) = \left(\frac{1}{s}\right) \sum_{r=1}^{s} \mathbf{x}(r) \qquad [1]$$

Note that the centroid is the simple arithmetic mean of the vectors of the cluster members, and this mean is frequently used to represent the cluster as a whole. In situations where a mean is not applicable or appropriate, the median can be used to define the cluster *medoid* (see Kaufman and Rousseeuw[2] for details). The square-error (also called the *within-cluster variance*), e^2, for a cluster is the sum of squared Euclidean distances to the centroid or medoid for all s items in that cluster:

$$e^2 = \sum_{r=1}^{s} [\mathbf{x}(r) - \mathbf{x}(c)]^2 \qquad [2]$$

The square-error across all K clusters in a partition is the sum of the square-errors for each of the K clusters. (Note also that the standard deviation would be the square root of the square-error.)

CLUSTERING ALGORITHMS

This chapter concentrates on the "classical" clustering methods, because they are the methods that have been applied most often in the chemical community. Standard reference works devoted to clustering algorithms include those by Hartigan,[26] Murtagh,[27] and Jain and Dubes.[28]

Hierarchical Methods

Hierarchical Agglomerative
The most commonly implemented hierarchical clustering methods are those belonging to the family of *sequential agglomerative hierarchical nonoverlapping* (SAHN) methods. These are traditionally implemented using

what is known as the *stored-matrix* algorithm, so named because the starting point is a matrix of all pairwise proximities between items in the data set to be clustered. Each cluster initially corresponds to an individual item (singleton). As clustering proceeds, each cluster may contain one or more items. Eventually, there evolves one cluster that contains all items. At each iteration, a pair of clusters is merged (agglomerated) and the number of clusters decreases by 1. The stored-matrix algorithm proceeds as follows:

1. Calculate the initial proximity matrix containing the pairwise proximities between all pairs of clusters (singletons) in the data set.
2. Scan the matrix to find the most similar pair of clusters, and merge them into a new cluster (thus replacing the original pair).
3. Update the proximity matrix by inactivating one set of entries of the original pair and updating the other set (now representing the new cluster) with the proximities between the new cluster and all other clusters.
4. Repeat steps 2 and 3 until just one cluster remains.

The various SAHN methods differ in the way in which the proximity between clusters is defined in step 1 and how the merged pair is represented as a single cluster in step 3. The proximity calculation in step 3 typically uses the Lance–Williams matrix-update formula:[29]

$$d[k, (i, j)] = \alpha_i\, d[k, i] + \alpha_j\, d[k, j] + \beta\, d[i, j] + \gamma\, |d[k, i] - d[k, j]| \qquad [3]$$

where $d[k, (i, j)]$ is the proximity between cluster k and the cluster (i, j) formed from merging clusters i and j. Different values for α_i, α_j, β, and γ define various SAHN methods, some of which are shown in Table 1 and described below.

   In *single-link* clustering, the proximity between two clusters is the minimum distance between any pair of items (one from each cluster), that is, the closest pair of points between each cluster. In contrast, in *complete-link* clustering, the proximity between two clusters is the maximum distance between any pair of items, that is, the farthest pair of points between each cluster. Single-link and complete-link represent the extremes of SAHN clustering. In

Table 1 Parameter Values for Some Common SAHN Methods Defined by the Lance–Williams Matrix Update Formula[a]

SAHN Method	α_i	α_j	β	γ
Single-link	0.5	0.5	0	-0.5
Complete-link	0.5	0.5	0	0.5
Group-average	$\dfrac{N_i}{N_i + N_j}$	$\dfrac{N_j}{N_i + N_j}$	$\dfrac{-N_i \times N_j}{(N_i + N_j)^2}$	0
Ward	$\dfrac{N_i + N_k}{N_i + N_j + N_k}$	$\dfrac{N_j + N_k}{N_i + N_j + N_k}$	$\dfrac{-N_k}{N_i + N_j + N_k}$	0

[a] The parameters N_i, N_j, and N_k = number of compounds in clusters i, j, and k, respectively.

the middle is *average-link* clustering in which the proximity between two clusters is the arithmetic average of distances between all pairs of items. Also in the middle is *Ward's* method[30] in which the proximity is the variance between the clusters (where variance is defined as the sum of square-errors of the clusters; see Eq. [2]). At each iteration, the pair of clusters chosen is that whose merger produces the minimum change in square-error (or within-cluster variance; hence the method is also known as the *minimum-variance* method). As the number of clusters decreases, the square-error across all clusters increases. Ward's method minimizes the square-error increase and minimizes the intracluster variance while maximizing the intercluster variance. Because a cluster is represented by its centroid, Ward's method is classified as a *geometric* or *cluster-center* method. Other methods such as the single-link, complete-link, and group-average methods are classified as *graph-theoretic* or *linkage methods*. Murtagh[27] introduced the concept of a *reducibility property* that is applicable to geometric methods. The reducibility property states that for the merger of two clusters, a and b, to form cluster c, there cannot be another cluster, d, that is closer to c than to a or b. If the method satisfies the reducibility property, agglomerations can be performed in localized areas of the proximity space and then amalgamated to produce the full hierarchy. Ward's method, implemented using the Euclidean distance as the proximity measure, is one of the few geometric methods satisfying the reducibility property. Voorhees[31] subsequently showed that if the cosine coefficient of similarity is used as the proximity measure, the group-average method can be implemented as a geometric method, and it satisfies the reducibility property.

For a data set of N compounds, the stored-matrix algorithm for SAHN methods requires $O(N^2)$ time and $O(N^2)$ space for creation and storage of the proximity matrix while requiring $O(N^3)$ time for the clustering. This algorithm is thus very demanding of resources for anything other than small data sets. The importance of the reducibility property is that it enables the stored-matrix algorithm to be replaced by the more efficient *reciprocal nearest-neighbor* (RNN) algorithm that requires only $O(N^2)$ time and $O(N)$ space. Because agglomerations can be performed in localized areas of the proximity space, the RNN algorithm works by tracing paths through proximity space from one point to its nearest neighbor until a point is reached whose nearest neighbor is the previous point in the path, that is, a pair of points that are reciprocal nearest neighbors. These points represent a pair that should be merged into a single point as one of the agglomerations of the full hierarchy. The RNN algorithm is carried out using the following steps:

1. Mark all points as "unused."
2. Begin at an unused point and trace a path of unused nearest neighbors until a reciprocal nearest neighbor pair is found.
3. Add the pair of points to the list of RNNs along with the proximity between them; mark one of the pair of points as "used" (to inactivate it and

its centroid) and replace the centroid of the other point by the centroid of the merged pair.

4. Continue the path tracing from the penultimate point in the path if one exists; otherwise start path tracing from a new, unused starting point.

5. Repeat steps 2–4 until only one unused point remains.

6. Sort the list of RNNs by decreasing proximity values; the sorted list represents the agglomerations needed to construct the hierarchy.

Because path tracing moves from one nearest neighbor to the next, random access to each point is required.

Hierarchical Divisive

Most hierarchical divisive methods are monothetic, meaning that each split is determined on the basis of a single descriptor. The methods differ in how the descriptor is chosen with one possibility being to select the descriptor that maximizes the distance between the resultant clusters. Monothetic divisive methods are usually faster than the SAHN methods described above and have found utility in biological classification. However, for chemical applications, monothetic division often gives poor results when compared to polythetic division or SAHN methods, even though the closely related classification method of recursive partitioning can be very effective in chemical applications (e.g., see the article by Chen, Rusinko, and Young[32]). Unfortunately, most polythetic divisive methods are very resource demanding (more so than for SAHN methods), and accordingly they have not been used much for chemical applications. One exception is the *minimum-diameter* method published by Guenoche, Hansen, and Jaumard;[33] it requires $O(N^2 \log N)$ time and $O(N^2)$ space. This method is based on dividing clusters at each iteration in such a way as to minimize the cluster diameter. The cluster diameter is defined as the largest dissimilarity between any pair of its members, with singleton clusters having a diameter of zero. The minimum-diameter algorithm accomplishes its task by carrying out the following steps:

1. Generate a sorted list of all $N(N-1)/2$ dissimilarities, with the most dissimilar pairs listed first.

2. Perform an initial division by selecting the first pair from the sorted list (i.e., the most dissimilar points in the data set); assign every other point to the closest of the pair.

3. Choose the cluster with the largest diameter and divide it into two clusters so that the larger cluster has the smallest possible diameter.

4. Repeat step 3 for a maximum of $N-1$ divisions.

Nonhierarchical Methods

Single-Pass

Methods that cluster data on the basis of a single scan of the data set are referred to as single-pass. A proximity threshold is typically used to decide

whether a compound is assigned to an existing cluster (represented as a centroid) or if it should be used to start a new cluster. The first compound selected becomes the first cluster; a single sequential scan of the data set then assigns the remaining compounds, and cluster centroids are updated as each compound is assigned to a particular cluster. The most common single-pass algorithm is called the *leader algorithm*. The leader algorithm carries out the following steps to provide a set of nonhierarchical clusters:

1. Set the number of existing clusters to zero.
2. Use the first compound in the data set to start the first cluster.
3. Calculate the similarity, using some appropriate measure, between the next compound and all the existing clusters. If its similarity to the most similar existing cluster exceeds some threshold, assign it to that cluster; otherwise use it to start a new cluster.
4. Repeat step 2 until all compounds have been assigned.

This method is simple to implement and very fast. The major drawback is that it is order dependent; if the compounds are rearranged and scanned in a different order, then the resulting clusters can be different.

Nearest Neighbor

A simple way to isolate dense regions of proximity space is to examine the nearest neighbors of each compound to determine groups with a given number of mutual nearest neighbors. Although several nearest-neighbor methods have been devised, the *Jarvis–Patrick method*[34] is almost exclusively used for chemical applications. The method proceeds in two stages.

The first stage generates a list of the top K nearest neighbors for each of the N compounds, with proximity usually measured by the Euclidean distance or the Tanimoto coefficient;[23] K is typically 16 or 20. The Tanimoto coefficient has been found to perform well for chemical applications where the compounds are represented by fragment screens (bit strings denoting presence/absence of structural features). For finding nearest neighbors with Tanimoto coefficients as a proximity measure, one can use an efficient inverted file approach described by Perry and Willett[35] to speed up the creation of nearest-neighbor lists.

The second stage scans the nearest-neighbor lists to create clusters that fulfill the three following *neighborhood conditions*:

1. Compound i is in the top K nearest-neighbor list of compound j.
2. Compound j is in the top K nearest-neighbor list of compound i.
3. Compounds i and j have at least $Kmin$ of their top K nearest neighbors in common, where $Kmin$ is user-defined and lies in the range 1 to K.

Pairs of compounds that fail any of the above conditions are not put into the same cluster.

To scan the nearest-neighbor lists and create the clusters in this stage of nonhierarchical clustering, the following three steps are carried out:

1. Tag each compound with a sequential cluster label so that each is a singleton.
2. For each pair of compounds, i and j (i < j), compare the nearest-neighbor lists on the basis of the three neighborhood conditions. If the three conditions are passed, replace the cluster label for compound j with the cluster label for compound i. Then, scan all previously processed compounds and replace any occurrences of the cluster label for compound j by the cluster label for compound i.
3. Scan to extract clusters by retrieving all compounds assigned the same cluster label.

The Jarvis–Patrick method requires $O(N^2)$ time and $O(N)$ space.

Relocation
Relocation methods start with an initial guess as to where the centers of clusters are located. The centers are then iteratively refined by shifting compounds between clusters until stability is achieved. The resultant clustering is reliant upon the initial selection of seed compounds that serve as cluster centers. Hence, relocation methods can be adversely affected by outlier compounds. [An *outlier* is a cluster of one item (a singleton or noise). It is on its own, and the clustering method has not put it with anything else because it is not similar enough to anything else.] The iterative refinement seeks an optimal partitioning of the compounds but would likely find a suboptimal solution because it would require the analysis of all possible solutions to guarantee finding the global optimum. Nevertheless, the computational efficiency and mathematical foundation of these methods have made them very popular, especially with statisticians.

The best-known relocation method is the *k-means* method, for which there exist many variants and different algorithms for its implementation. The k-means algorithm minimizes the sum of the squared Euclidean distances between each item in a cluster and the cluster centroid. The basic method used most frequently in chemical applications proceeds as follows:

1. Choose an initial set of *k* seed compounds to act as initial cluster centers.
2. Assign each compound to its nearest cluster centroid (classification step).
3. Recalculate each cluster centroid (minimization step).
4. Repeat steps 2 and 3 for a given number of iterations or until no compounds are moved from one cluster to another.

In step 1, the initial compounds are usually selected at random from the data set. Random selection is quick and, for large heterogeneous data sets, likely to provide a reasonable initial set. Steps 2 and 3 can be performed separately or in combination. If done separately, the classification (step 2) is performed on

all compounds before recalculation of each cluster centroid (step 3). This approach is referred to as *noncombinatorial* (or *batch update*) classification. If steps 2 and 3 are done in combination, moving a compound from its current cluster to a new cluster (step 2) immediately necessitates recalculation of the affected cluster centroids (step 3). This latter approach is called *combinatorial* or *online update* classification. Most implementations for chemical applications use noncombinatorial classification. In step 4, convergence to a point where no further compounds move between clusters, is usually rapid, but, for safety, a maximum number of iterations can be specified. k-Means clustering requires $O(Nmk)$ time and $O(k)$ space. Here, m is the number of iterations to convergence, and k is the number of clusters. Because m is typically much smaller than N and the effect of k can be reduced substantially through efficient implementation, k-means algorithms essentially require $O(N)$ time.

Mixture Model

Clustering can be viewed as a density estimation problem. The basic premise used in such an estimation is that in addition to the observed variables (i.e., descriptors) for each compound there exists an unobserved variable indicating the cluster membership. The observed variables are assumed to arrive from a mixture model, and the mixture labels (cluster identifiers) are hidden. The task is to find parameters associated with the mixture model that maximize the likelihood of the observed variables given the model. The probability distribution specified by each cluster can take any form. Although mixture model methods have found little use in chemical applications to date, they are mentioned here for completeness and because they are obvious candidates for use in the future.

The most widely used and most effective general technique for estimating the mixture model parameters is the expectation maximization (EM) algorithm.[36] It finds (possibly suboptimally) values of the parameters using an iterative refinement approach similar to that given above for the k-means relocation method. The basic EM method proceeds as follows:

1. Select a model and initialize the model parameters.
2. Assign each compound to the cluster(s) determined by the current model (expectation step).
3. Reestimate the parameters for the current model, given the cluster assignments made in step 2, and generate a new model (maximization step).
4. Repeat steps 2 and 3 for n iterations or until the nth and $(n-1)$th model are sufficiently close.

This method requires prior specification of a model and typically takes a large number of iterations to converge.

Note that the k-means relocation method is really a special case of EM that assumes: (1) each cluster is modeled by a spherical Gaussian distribution, (2) each data item is assigned to a single cluster, and (3) the mixture weights

are equal. Assignment of each compound to the closest-cluster centroid is the expectation step; recalculation of the cluster centroids (model parameters) after assignment is the maximization step.

Topographic

Topographic clustering methods attempt to preserve the proximities between clusters, thus facilitating visualization of the clustering results. For k-means clustering, the cost function is invariant, whereas in topographic clustering it is not, and a predefined neighborhood is imposed on the clusters to preserve the proximities between them. The Kohonen, or self-organizing, map,[37,38] apart from being one of the most commonly used types of neural network, is also a topographic clustering method. A *Kohonen network* uses an unsupervised learning technique to map higher dimensional spaces of a data set down to, typically, two or three dimensions (2D or 3D), so that clusters can be identified from the neurons' coordinates (topological position); the values of the output are ignored. Initially, the neurons are assigned weight vectors with random values (weights). During the self-organization process, the data vectors of the neuron having the most similar weight vector to each data vector and its immediately adjacent neurons are updated iteratively to place them closer to the data vector. The *Kohonen mapping* thus proceeds as follows:

1. Initialize each neuron's weight vector with random values.
2. Assign the next data vector to the neuron having the most similar weight vector.
3. Update the weight vector of the neuron of step 2 to bring it closer to the data vector.
4. Update neighboring weight vectors using a given updating function.
5. Repeat steps 2–4 until all data vectors have been processed.
6. Start again with the first data vector, and repeat steps 2–5 for a given number of cycles.

The iterative adjustment of weight vectors is similar to the iterative refinement of k-means clustering to derive cluster centroids. The main difference is that adjustment affects neighboring weight vectors at the same time. Kohonen mapping requires $O(Nmn)$ time and $O(N)$ space, where m is the number of cycles and n the number of neurons.

Other Nonhierarchical Methods

We have delineated the main categories of clustering methods applicable to chemical applications above. We have also provided one basic algorithm as an example of each. Researchers in other disciplines sometimes use variants of these main categories. The main categories that have been used by those researchers but omitted here include density-based clustering and graph-based clustering techniques. These will be mentioned briefly in the next section.

PROGRESS IN CLUSTERING METHODOLOGY

The representations used for chemical compounds are typically "*dataprints*" (tens or hundreds of real number descriptors, such as topological indexes and physicochemical properties) or *fingerprints* (thousands of binary descriptors indicating the presence or absence of 2D structural fragments or 3D pharmacophores). These numbers can be compared to the tens or hundreds of descriptors typically encountered in data mining and the thousands of descriptors encountered in information retrieval. We now outline the development of clustering methods that are suited to handling these representations and that have been, or in the near future are likely to be, used for chemical applications. Specific examples of chemical applications are given later in the section entitled Chemical Applications.

Algorithm Developments

Having briefly outlined the basic algorithms that are implemented in many of the standard clustering methods, we now set the algorithms in context by reviewing their historical development, discuss the characteristics of each method, and then highlight some of the variants that have been developed for overcoming certain limitations. Clustering is now such a large area of research and everyday use that this chapter must be selective rather than comprehensive in scope. The interested reader can access further details from the references cited throughout this chapter and from the recent review by Murtagh.[39]

Most of the development of hierarchical clustering methods occurred from the 1960s through the mid-1980s, after which there was a period of consolidation, with little new development until recently. From this developmental period, two key publications were those of Lance and Williams[29] in 1967 and the review of hierarchical clustering methods by Murtagh[27] in 1985. Following this developmental period, a few variations have been proposed. Matula[40] developed algorithms that implemented both divisive and agglomerative average-linkage methods, but with high computational costs for processing large data sets. That same year, Jain, Indrayan, and Goel[41] compared single and complete linkage, group and weighted average, centroid, and median agglomerative methods and concluded that complete linkage performed best in failing to find clusters from random data. Podani[42] produced a useful classification of agglomerative methods, in which the standard Lance–Williams update recurrence formula[29] is split into two formulas. He also introduced three new parameter variations, that is, three new agglomerative methods were defined, but these seem to represent more of an inclusion for the sake of completeness than a significant alternative to previously defined parameter variations. Roux[43] recognized the complexity problems in Matula's implementations[40] and mentioned restrictions that could be applied to

overcome them for a polythetic divisive implementation. Unfortunately, no algorithmic details were given.

Overall, the Lance–Williams recurrence formula, and its subsequent extensions, provides a consolidating basis for the implementation of hierarchic agglomerative methods. However, the standard ways of implementation, that is, by generating, storing, and updating the full distance matrix, or by generating distances as required, tend to be very demanding of computational resources. The review by Murtagh[39] explained how substantial reductions in the computational requirements for some of these methods could be achieved by using the reciprocal nearest neighbor approach. El-Hamdouchi and Willett[44] described the use of this approach for the implementation of the Ward method for document clustering. That same year (1989) Rasmussen and Willett[45] discussed parallel implementation of single-link and Ward methods for both document and chemical structure clustering. The RNN approach and single-link clustering have the additional advantage of only requiring a list of descriptor vectors and a function to return the nearest neighbor of any input vector, rather than a full proximity matrix. Downs, Willett, and Fisanick[46] used RNN implementations of the Ward and group-average methods in a comparison of methods for clustering compounds on the basis of property data (see section below on Comparative Studies on Chemical Data Sets). These two agglomerative methods have been used successfully in comparative studies covering a wide range of nonchemical applications, and they have been shown to provide consistently reasonable results. However, centroid- and medoid-based methods, such as Ward (and k-means nonhierarchical), and the group-average and complete-link methods tend to favor similarly sized hyperspherical clusters (i.e., clusters that are shaped like spheres in a space of more than three dimensions), and they can fail to separate clusters of different shapes, densities, or sizes. Single-link is not a centroid method; it uses just the pairwise similarities and is more analogous to density-based methods. Accordingly, it can find clusters of different shapes and sizes, but it is very susceptible to noise, such as outliers, and artifacts, and it has a tendency to produce straggly clusters (an effect known as *chaining*). The development of traditional hierarchical methods largely ignored the issues of noise, and, although the abilities of different methods to separate clusters were noted, little was done about this problem other than to advise users to adopt more than one method so that different types of clusters could be revealed.

Recent developments in the data mining community have produced methods better suited to finding clusters of different shapes, densities, and sizes. For example, Guha, Rastogi, and Shim[47,48] developed an algorithm called ROCK (RObust Clustering using linKs) that is a sort of hybrid between nearest-neighbor, relocation, and hierarchical agglomerative methods. Although more expensive computationally than RNN implementations of the Ward method, the algorithm is particularly well suited to nonnumerical data (of which the Boolean fingerprints used for chemical data sets are a

special case, although they can also be represented as binary, a special case of numeric). The same authors developed another algorithm called CURE (Clustering Using REpresentatives).[49] Here centroid and single-link-type approaches are combined by choosing more than one representative point from each cluster. With CURE, a user-specified number of diverse points is selected from a cluster, so that it is not represented by just the centroid (which tends to lead to hyperspherical clusters). To avoid the problem of influence from selected points that might be outliers, which can result in a chaining effect, the selected points are shrunk toward the cluster centroid by a specified proportion. This results in a computationally more expensive procedure, but the separation of differently shaped and sized clusters is better. Karypis, Han, and Kumar[50] also addressed the problems of cluster shapes and sizes in their Chameleon algorithm. These authors provide a useful overview of the problems of other clustering methods in their summary. Chameleon measures similarity on the basis of a dynamic model, which is to be contrasted with the fixed model of traditional hierarchical methods. Two clusters are merged only if their interconnectivity and closeness is high relative to the internal interconnectivity and closeness within the two clusters. The characteristics of each cluster are thus taken into account during the merging process rather than assuming a fixed model that, if the clusters do not conform to it, can result in inappropriate merging decisions that cannot be undone subsequently. In a different study, Karypis, Han, and Kumar[51] evaluated the use of multi-level refinement methods to detect and correct inappropriate merging decisions in a hierarchy. Fasulo[52] reviewed some of the other recent developments in the area of data mining with World Wide Web search engines. The developments cited in that review describe work that reassesses the manner in which clustering is performed; a range of methods, which are more flexible in their separation of clusters, were evaluated. It is further pointed out that problems still remain when scaling-up hierarchical clustering methods to the very high dimensional spaces characteristic of many chemical data sets. Other fundamental issues remain, such as the problem of tied proximities in hierarchical clustering.[53] This problem was mentioned many years earlier by Jain and Dubes,[28] among others. Tied proximities occur when the proximities between two different pairs of data items are equal, and result in ambiguous decision points when building the hierarchy, effectively leading to many possible hierarchies of which only one is chosen. MacCuish, Nicolaou, and MacCuish[53] show tied proximities to be surprisingly common with the types of fingerprints commonly used in chemical applications, and the problem increases with data set size. What is not clear is whether such ties have a major deleterious effect on the overall clustering and whether the chosen hierarchy is significantly different from any of the others that might have been chosen.

There has been little development of polythetic divisive methods since the publication of the minimum-diameter method[33] in 1991. Garcia et al.[54] developed a path-based approach with similarities to single-link. The method

has time requirements of $O(MN^2)$ for M clusters and N compounds, making the method particularly suitable for finding a small number of clusters. Wang, Yan, and Sriskandarajah[55] updated the single criterion minimum-diameter method with a multiple criteria algorithm that considers both maximum split (intercluster separation) and minimum diameter in deciding the best bipartition. Their algorithm reduces the *dissection* effect (similar items forced into different clusters because doing so reduces the diameter) associated with the minimum-diameter criterion and the chaining effect associated with the maximum-split criterion. More recently, Steinbach, Karypis, and Kumar[56] reported an interesting variant of k-means that is actually a hierarchical polythetic divisive method. At each point where a cluster is to be split into two clusters, the split is determined by using k-means, hence the name "bisecting k-means." The results for document clustering, using keywords as descriptors, are shown to be better than standard k-means, with cluster sizes being more uniform, and better than the agglomerative group-average method.

Monothetic divisive clustering has largely been ignored, although there have been applications and development of a classification method closely related to monothetic divisive clustering. This classification is recursive partitioning, a type of decision tree method.[57–60]

Nonhierarchical algorithms that cluster the data set in a single pass, such as the leader algorithm, have had little development, except to identify appropriate ways of preordering the data set so as to get around the problem of dependency on processing order (work on this is discussed in the Chemical Applications section). For multipass algorithms, however, efforts have been made to minimize the number of passes required, in some cases reducing them to single-pass algorithms. In the area of data mining, this work has resulted in a method that does not fit neatly into the categorization used in this review. Zhang, Ramakrishnan, and Livny[61] developed a program called BIRCH (Balanced Iterative Reducing and Clustering using Heuristics), an $O(N^2)$ method that performs a single scan of the data set to sort items into a cluster features (CF) tree. This operation has some similarity with the leader algorithm; the nodes of the tree store summary information about clusters of dense points in the data so that the original data need not be accessed again during the clustering process. Clustering then proceeds on the in-memory summaries of the data. However, the initial CF tree building requires the maximum cluster diameter to be specified beforehand, and the subsequent tree building is thus sensitive to the value chosen. Overall, the idea of BIRCH is to bring together items that should always be grouped together, with the maximum cluster diameter ensuring that the cluster summaries will all fit into available memory. Ganti et al.[62] outlined a variant of BIRCH called BUBBLE. It does not rely on vector operations but builds up the cluster summaries on the basis of a distance function that obeys the triangle inequality, an operation that is more CPU demanding than operations in coordinate space.

Nearest-neighbor nonhierarchical methods have received much attention in the chemical community because of their fast processing speeds and ease of implementation. The comparative studies outlined in the next section (Comparative Studies on Chemical Data Sets) led to the widespread adoption of the Jarvis–Patrick nearest-neighbor method for clustering large chemical data sets. To improve results obtained by the standard Jarvis–Patrick implementation, several extensions have been developed. The standard implementation tends to produce a few large heterogeneous clusters and an abundance of singletons, which is hardly surprising because the method was originally designed to be space distorting,[34] that is, contraction of sparsely populated areas clusters and splitting of densely populated areas. Attempts to overcome these tendencies include the use of variable-length nearest-neighbor lists,[12,20] reclustering of singletons,[63] and the use of fuzzy clustering.[64] For variable-length nearest-neighbor lists, the user specifies a proximity threshold so that the lists will contain all neighbors that pass the threshold test rather than a fixed number of nearest neighbors. During clustering, the comparison between nearest-neighbor lists is made on the basis of a specified minimum percentage of the neighbors in the shorter list being in common. These modifications help prevent true outliers from being forced to join a cluster while preventing the arbitrary splitting of large clusters arising from the limitations imposed by fixed length lists. When using fingerprints for clustering chemical data sets, Brown and Martin[20] showed improved results compared with the standard implementation, whereas Taraviras, Ivanciuc, and Cabrol-Bass[65] show contrary results when clustering descriptors.

The reclustering of singletons is used in the "cascaded clustering" method of Menard, Lewis, and Mason.[63] This method applies the standard Jarvis–Patrick clustering iteratively, removes the singletons, and reclusters them using less strict parameters until fewer than a specified percentage of singletons remain. The fuzzy Jarvis–Patrick method outlined by Doman et al.[64] is the most radical Jarvis–Patrick variant. In the fuzzy method, clusters in dense regions are extracted using a similarity threshold and the standard crisp method. The compounds are then assigned probabilities of belonging to each of the crisp clusters. Any previously unclustered compounds not exceeding a specified threshold probability of belonging to any of the crisp clusters are regarded as outliers and remain as singletons.

Other nearest-neighbor methods include the agglomerative hierarchical method of Gowda and Krishna,[66] which uses the position of nearest neighbors, rather than just the number, in a measure called the *mutual neighborhood value* (MNV). Given points i and j, if i is the pth neighbor of j and j is the qth neighbor of i, then the MNV is $(p + q)$. Smaller values of MNV indicate greater similarity, and a specified threshold MNV is used to determine whether points should be merged. Dugad and Ahuja[67] extended the MNV concept to include the density of two clusters that are being considered for merger. In addition to the threshold MNV, if there exists a point k with

MNV (i,k) less than MNV (i,j) but distance (i,k) greater than or equal to distance (i,j), then i is not a valid neighbor of j, and j is not a valid neighbor of i. The neighbor validity check can result in many small clusters, but these clusters can be merged afterward by relaxing the reciprocal nature of the check.

Relocation algorithms are widely used outside of chemical applications, largely because of their simplicity and speed. The original k-means noncombinatorial methods, such as that by Forgy,[68] and the combinatorial methods, such as that by MacQueen,[69] have been modified into different versions for use in many disciplines, a few of which are mentioned here. Efficient implementations of k-means include those by Hartigan and Wong[70] and Spaeth.[71] A variation of the k-means algorithm, referred to as the *moving method*, looks ahead to see whether moving an item from one cluster to another will result in an overall decrease in the square error (Eq. [2]); if it does, then the moving is carried out. Duda and Hart[72] and Ismail and Kamel[73] originally outlined this variant, while Zhang, Wang, and Boyle[74] further developed the idea and obtained better results than a standard noncombinatorial implementation of k-means. Because the method relies on the concept of a centroid, it is usually used with numerical data. However, Huang[75] reported variants that use k-modes and k-prototypes that are suitable for use with categorical and mixed-numerical and categorical data, respectively.

The main problems with k-means are (1) the tendency to find hyperspherical clusters, (2) the danger of falling into local minima, (3) the sensitivity to noise, and (4) the variability in results that depends on the choice of the initial seed points. Because k-means (and its fuzzy equivalent, c-means) is a centroid-based method, nothing much can be done about the tendency to produce hyperspherical clusters, although the CURE methodology mentioned above might alleviate this tendency somewhat. Falling into local minima cannot be avoided, but rerunning k-means with different seeds is a standard way of producing alternative solutions. After a given number of reruns, the solution is chosen that has produced the lowest square-error across the partition. An alternative to this is to perturb an existing solution, rather than starting again. Zhang and Boyle[76] examined the effects of four types of perturbation on the moving method and found little difference between them. Estivell-Castro and Yang[77] suggested that the problem of sensitivity to noise is due to the use of means (and centroids) rather than medians (and medoids). These authors proposed a variant of k-means based on the use of medoids to represent each cluster. However, calculation of a point to represent the medoid is more CPU-expensive [$O(n \log n)$ for each cluster of size n] than that required for the centroid, resulting in a method that is slightly slower than k-means (but faster than EM algorithms[36]). A similar variant based on medoids is the PAM (Partitioning Around Medoids) method developed by Kaufman and Rouseeuw.[2] This method is very time consuming, and so the authors developed CLARA (Clustering LARge Applications), which takes a sample of a data

set and applies PAM to it. An alternative to sampling the compounds has been developed by Ng and Han.[78] Their CLARANS (Clustering Large Applications based on RANdomized Search) method samples the neighbors, rather than the compounds, to make PAM more efficient.

The most common way of choosing seeds for k-means is by random selection, which is statistically reasonable given a large heterogeneous data set. Alternatively, a set of k diverse seeds could be selected using, for example, the MaxMin subset selection method.[79,80] Diverse seeds have been shown to give better clustering results by Fisher, Xu, and Zard.[81] One of the early suggestions, by Milligan,[82] was that a partition resulting from hierarchical agglomerative clustering should be used as the initial partition for k-means to refine. It may seem counterproductive to initialize an $O(N)$ method by first running an $O(N^2)$ method, because it means that very large data sets cannot be processed, but k-means is then effectively being used to refine individual partitions and to correct inappropriate assignments made by the hierarchical method. An iterative method for refining an entire hierarchy has been discussed by Fisher.[83] The iterative method starts at the root (i.e., the top of the hierarchy, with all compounds in one cluster), recursively removes each cluster, resorts it into the hierarchy, and continues iterating until no clusters are moved, other than moving individual items from one cluster to another.

Of the mixture model methods, the expectation maximization (EM) algorithm[36] is the most popular because it is a general and effective method for estimating the model parameters and for fitting the model to the data. Though now quite old, the method was relatively unused until a surge of recent interest has propelled its further development and implementation for data mining applications.[84] As mentioned earlier, k-means is a special case of EM. However, because standard k-means uses the Euclidean metric, it is not appropriate for clustering discrete or categorical data. The EM algorithm does not have these limitations, and, since the mixture model is probabilistic, it can also effectively separate clusters of different sizes, shapes, and densities. A major contribution to the development of the EM algorithm came from Banfield and Raftery[85] who reparameterized the standard distributions to make them more flexible and include a Poisson distribution to account for noise. Various models were developed and compared using the approximate weight of evidence (AWE) statistic, which estimates the Bayesian posterior probability of the clustering solution. Fraley and Raftery[86] subsequently replaced AWE by the more reliable Bayesian information criterion (BIC), which enabled them to produce an EM algorithm that simultaneously yields the best model and determines the best number of clusters. One other interesting aspect of their work is that the EM algorithm is seeded with the clustering results from hierarchical agglomerative clustering. It is not clear whether, by using a less expensive seed selection, the EM algorithm will scale to the very large, high-dimensional data sets of chemical applications, or if the necessary parameterization will be acceptable in practice.

The use of a fixed model in a clustering method favors retrieval of clusters of certain shapes (as exemplified by the hyperspherical clusters retrieved by centroid-based methods). An alternative is to use a density-based approach, in which a cluster is formed from a region of higher density than its surrounding area. The clustering is then based on local criteria, and it can pick out clusters of any shape and internal distribution. Such approaches are typically not applicable directly to high dimensions, but progress is being made in that direction within the data mining community. An example is the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) method of Ester et al.[87] that was subsequently extended by Ankerst et al.[88] to give the OPTICS (Ordering Points To Identify the Clustering Structure) method. These two methods work on a principle that each point of a cluster must have at least a given number of other points within a specified radius. Points fulfilling these conditions are clustered; any remaining points are considered to be outliers, that is, noise. The OPTICS method has been enhanced by Breunig et al.[89] to identify outliers, and by Breunig, Kriegel, and Sander,[90] who combined it with BIRCH[61] to increase speed.

Other density-based approaches designed for high dimensions include CLIQUE (Clustering In QUEst) by Agrawal et al.,[91] and PROCLUS (PROjected CLUSters), by Aggarwal et al.[92] These two methods recognize that high dimensional spaces are typically sparse so that the similarity between two points is determined by a few dimensions, with the other dimensions being irrelevant. Clusters are thus formed by similarity with respect to subspaces rather than full dimensional space. In the CLIQUE algorithm, dense regions of data space are determined by using cell-based partitioning, which are then used as initial bases for forming the clusters. The algorithm works from lower to higher dimensional subspaces by starting from cells identified as dense in $(k-1)$-dimensional subspace and extending them into k-dimensional subspace. The result is a set of overlapping dense regions that are extracted as the clusters. Research into improving grid-based methods is continuing, as demonstrated by the variable grid method of Nagesh.[93] In contrast, the PROCLUS program generates nonoverlapping clusters by identifying potential cluster centers (medoids) using a MaxMin subset selection procedure. The best medoids are selected from the initial set by an iterative procedure in which data items within the locality of a medoid (i.e., within the minimum distance between any two medoids) are assigned to that cluster. Rather than using all dimensions, the dimensions associated with each cluster are used in the Manhattan segmental distance[92] to calculate the distance of an item from the cluster. The Manhattan segmental distance is a normalized form of the Manhattan distance that enables comparison of different clusters with varying numbers of dimensions. (The Manhattan, or city-block, or Hamming, distance is the sum of absolute differences between descriptor values; in contrast, the Euclidean distance is the square root of the sum of squares differences between descriptor values.) Once the best medoids have been

selected, a final single pass over the data set assigns each item to its nearest medoid.

Graph-theoretic algorithms have seen little use in chemical applications. The basis of these methods is some form of a graph in which the vertices are the items in the data set and the edges are the proximities between them. Early methods created clusters by removing edges from a minimum spanning tree or by constructing a Gabriel graph, a relative neighborhood graph, or a Delauney triangulation, but none of these graph-theoretic methods are suitable for high dimensions. Reviews of these methods are given by Jain and Dubes[28] and Matula.[94] Recent advances in computational biology have spurred development of novel graph-theoretic algorithms based on isolating areas called cliques or "almost cliques" (i.e., highly connected subgraphs) from the graph of all pairwise similarities. Examples include the algorithms by Ben-Dor, Shamir, and Yakhini,[95] Hartuv et al.,[96] and Sharan and Shamir[97] that find clusters in gene expression data. Jonyer, Holder, and Cook[98] developed a hierarchical graph-theoretic method that begins with the graph of all pairwise similarities and then iteratively finds subgraphs that maximally compress the graph. The time consumption of these graph-theoretic methods is currently too great to apply to very large data sets.

One way to speed up the clustering process is to implement algorithms on parallel hardware. In the 1980s Murtagh[27,99] outlined a parallel version of the RNN algorithm for hierarchical agglomerative clustering. Also in that decade, Rasmussen, Downs, and Willett[45,100] published research on parallel implementations of Jarvis–Patrick, single-link, and Ward clustering for both document and chemical data sets, and Li and Fang[101] developed parallel algorithms for k-means and single-link clustering. In 1990, Li[102] published a review of parallel algorithms for hierarchical clustering. This in turn elicited a classic riposte from Murtagh[103] to the effect that the parallel algorithms were no better than the more recent $O(N^2)$ serial algorithms. Olson[104] presented $O(N)$ and $O(N \log N)$ algorithms for hierarchical methods using N processors. For chemical applications, in-house parallel implementations include the leader algorithm at the National Cancer Institute[105] and k-means at Eli Lilly[79] (both discussed in the section on Chemical Applications), and commercially available parallel implementations include the highly optimized implementation of Jarvis–Patrick by Daylight[14] and the multiprocessor version of the Ward and group-average methods by Barnard Chemical Information.[12]

Another way of speeding up clustering calculations is to use a quick and rough calculation of distance to assess an initial separation of items and then to apply the more CPU-expensive, full-distance calculation on only those items that were found to cluster using the rough calculation. McCallum, Nigam, and Ungar[106] exploited this idea by using the rough calculation to divide the data into *canopies* (roughly overlapping clusters). Only items within the same canopy, or canopies, were used in the subsequent full-distance calculations to determine nonoverlapping clusters (using, e.g., a hierarchical agglomerative, EM,

or k-means method). The nature of the rough-distance measure used can guarantee that the canopies will be sufficiently broad to encompass all candidates for the ensuing full-distance measure. These ideas to speed up nearest-neighbor searches are similar to the earlier use of bounds on the distance measure, as discussed by Murtagh.[27]

Comparative Studies on Chemical Data Sets

Much of the use of clustering for chemical applications is based on the *similar property principle*.[107] This principle, which holds in many, but certainly not all, structure–property relationships, states that compounds with similar structure are likely to exhibit similar properties. Clustering on the basis of structural descriptors is thus likely to group compounds having similar properties. However, there exist many different clustering methods, each having its own particular characteristics that are likely to affect the composition of the resultant clusters. Consequently, there have been several comparative studies on the performance of different clustering methods when applied to chemical data sets. The first such studies were conducted by Willett and Rubin[5,108–110] in the early 1980s. These studies were highly influential in the subsequent implementation of clustering methods in commercial and in-house software systems used by the pharmaceutical industry. Over 30 hierarchical and nonhierarchical methods were tested on 10 small data sets for which certain properties were known. Clustering was conducted using 2D fingerprints as compound representations. The leave-one-out approach (based on the similar property principle) was used to compare the results of different clustering methods by predicting the property of each compound (as the average of the property of the other members of the cluster) and correlating it with the actual property. High correlations indicate that compounds with similar properties have been clustered together. The results indicated that the Ward hierarchical method gave the best overall performance. But, this method was not well suited to processing large data sets due to the requirement for random access to the fingerprints. The Jarvis–Patrick nonhierarchical method results were almost as good and, because it does not require the fingerprints to be in memory, it became the recommended method.

In the early 1990s, a subsequent study by Downs, Willett, and Fisanick[46] compared the performance of the Ward and group-average agglomerative methods, the minimum-diameter divisive hierarchical method, and the Jarvis–Patrick nonhierarchical method when using dataprints of calculated physicochemical properties. In this assessment, a data set was used that was considerably larger than those used in the original studies.[108–110] The results highlighted the poor performance of the Jarvis–Patrick method in comparison with the hierarchical methods. The hierarchical methods all had similar levels of performance with the minimum-diameter method being slightly better for small numbers of clusters. Brown and Martin[20] then investigated the same

clustering methods to compare their performance for compound selection, using various 2D and 3D fingerprints. Active/inactive data was available for the compounds in the data sets used, so assessment was based on the degree to which clustering separated active from inactive compounds (into nonsingleton clusters). Although the Jarvis–Patrick method was the fastest of the methods, it again gave the poorest results. The results were improved slightly by using a variant of the Jarvis–Patrick method that uses variable rather than fixed-length nearest-neighbor lists.[12] Overall, the Ward method gave the best and most consistent results. The group-average and minimum-diameter methods were broadly similar and only slightly worse in performance than the Ward method.

The influence of the studies summarized above can be seen in the methods subsequently implemented by many other researchers for their applications (see the section on Chemical Applications). One method that was included in the original assessment studies, but not in the later assessments, is k-means. This method did not perform particularly well on the small data sets of the original studies, and the resultant clusters were found to be very dependent on the choice of initial seeds; hence it was not included in the subsequent studies. However, k-means is computationally efficient enough to be of use for very large data sets. Indeed, over the last decade k-means and its variants have been studied extensively and developed for use in other disciplines. Because it is being increasingly used for chemical applications, any future comparisons of clustering methods should include k-means.

How Many Clusters?

A problem associated with the k-means, expectation maximization, and hierarchical methods involves deciding how many ''natural'' (intuitively obvious) clusters exist in a given data set. Determining the number of ''natural'' clusters is one of the most difficult problems in clustering and to date no general solution has been identified. An early contribution from Jain and Dubes[28] discussed the issue of *clustering tendency*, whereby the data set is analyzed first to determine whether it is distributed uniformly. Note that randomly distributed data is not generally uniform, and, because of this, most clustering methods will isolate clusters in random data. To avoid this problem, Lawson and Jurs[111] devised a variation of the Hopkins' statistic that indicates the degree to which a data set contains clusters. McFarland and Gans[112] proposed a method for evaluating the statistical significance of individual clusters by comparing the within-cluster variance with the within-group variance of every other possible subset of the data set with the same number of members. However, for large heterogeneous chemical data sets it can be assumed that the data is not uniformly or randomly distributed, and so the issue becomes one of identifying the most natural clusters.

Nonhierarchical methods such as k-means and EM need to be initialized with *k* seeds. This presupposes that *k* is a reasonable estimation of the number

of natural clusters and that the seeds chosen are reasonably close to the centers of these clusters. Epter, Krishnamoorthy, and Zaki[113] published one of the few papers addressing these issues for large data sets. Their solution is applicable to distance-based clustering and involves analysis of the histogram of pairwise distances between data items. For small data sets, all pairwise distances can be used, whereas for large data sets, random sampling (up to 10% of the data set) can be used to lessen the quadratic increase in time needed to generate the distances. For the distances calculated, the corresponding histogram is generated and then scanned to find the first spike (a large maximum followed by a large minimum). This point is used as the threshold for intracluster distance. The graph containing distances within this threshold contains connected components used to determine both the number of clusters present in the data set and the initial starting points from within these clusters. Assuming that a reasonable value for k is known, Fayyad, Reima, and Bradley[114,115] showed that one can minimize the problem of poor initial starting points by sampling the data set to derive a better set of starting points. A series of randomly selected subsets, larger than k, are extracted, clustered by k-means, amalgamated, and then clustered again using each solution from the subsets. The starting points from the subset giving the best clustering of the amalgamated subset are then chosen as the set of refined points for the main clustering, where "best" means the clustering that gives the minimal "distortion," that is, minimum error across the amalgamated subset. The method aims to avoid selecting outliers, which may occur with other selection methods such as MaxMin.

In hierarchical clustering, each level defines a partition of the data set into clusters. However, there is no associated information indicating which level is best in terms of splitting the data set into the "natural" number of clusters present and with each cluster containing the most appropriate compounds. Many methods and criteria have been proposed to try to derive such information from the hierarchy so that the "best" level is selected. Milligan and Cooper[116] published the first comprehensive comparison of hierarchy level selection methods, using psychology data. Thirty methods were tested for their ability to retrieve the correct number of clusters from several small data sets containing from 2 to 5 "natural" clusters. Fifteen years later, Wild and Blankley[117] published a major comparison of hierarchy level selection methods using chemical data sets. As part of that study, Ward clustering with 2D fingerprints was used to evaluate the performance of nine hierarchy level selection methods. The methods chosen were those that would be easy to implement and that did not require parameters. Eight of those methods were ones that Milligan and Cooper had previously examined; the ninth was a more recent method published by Kelley, Gardner, and Sutcliffe.[118] The study by Wild and Blankley concluded that the point biseral,[119] variance ratio criterion,[120] and Kelley methods gave the best overall results, with the Kelley method being more computationally efficient than the others [scaling at less
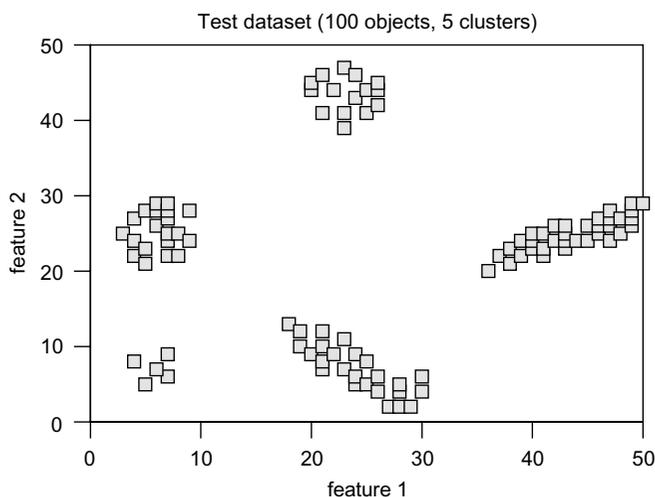
Figure 3 An example data set of 100 objects, represented by 2 features, that fall into 5 natural clusters.

than $O(N^2)$]. A test data set of 100 objects, represented by 2 features and grouped into 5 natural clusters, is shown in Figure 3. The corresponding plot of penalty values (calculated using the Kelley method) against the number of clusters (Figure 4) shows a clear minimum at 5 clusters.
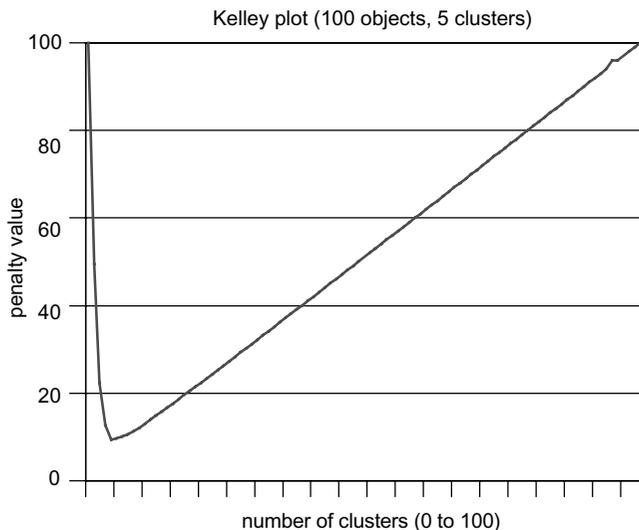
Figure 4 Kelley plot of the penalty value against number of clusters for the data set of 100 items in Figure 3, showing the minimum at 5 clusters.

Hierarchy level selection methods provide useful guidance in selecting reasonable partitions from hierarchies where the underlying structure of the data set is unknown. They are, however, a compromise in that they compare entire partitions with each other rather than individual clusters. In disciplines outside of chemistry, there is an increasing awareness that such global comparisons can mask comparative differences in local densities. For example, the situation in Figure 5 shows three clusters (below the dendrogram) that cannot be retrieved by using a conventional straight horizontal line across the dendrogram (such as that shown in Figure 1). Using a straight line can include either item 8 with cluster [3,1,2] but merge [4,5] with [6,7], or keep [4,5] and [6,7] separate but maintain 8 as a singleton. What may be required for the selection of the "best," nonoverlapping clusters from different partitions is a *stepped* (or *segmental*) horizontal line, which is illustrated by the dotted line across the dendrogram in Figure 5. No solution to deciding which is the best selection of nonoverlapping clusters appears to have been published to date, but there are examples of methods that are moving toward a solution. One such example
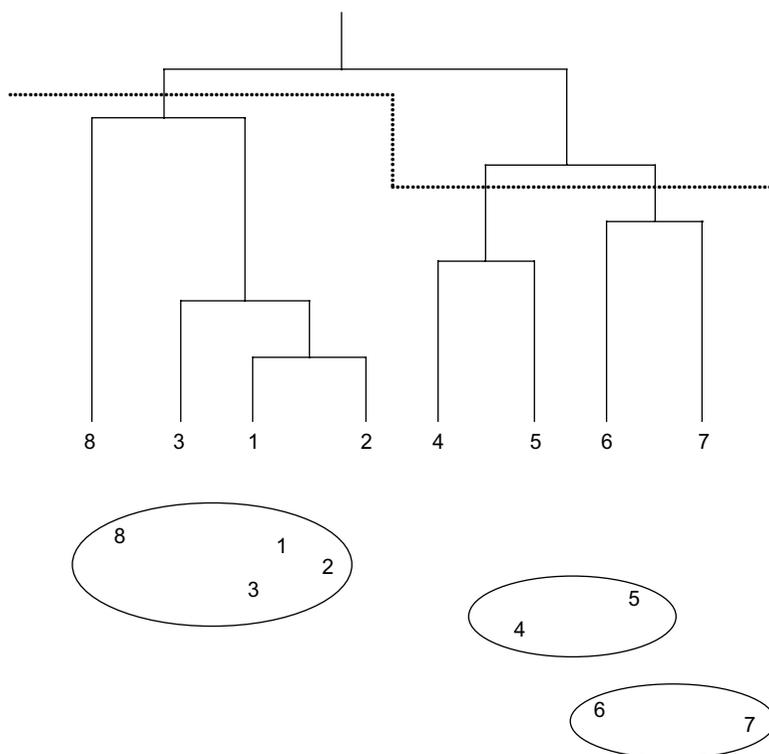
Figure 5 An illustration of how a stepped hierarchy partition can extract particular clusters (clusters [8,1,2,3], [4,5], and [6,7], as shown below the hierarchy).

is the OPTICS method that orders items in a data set in terms of local criteria, thus providing an equivalent to a density-based analysis.

A variety of different requirements exist for chemical applications. These requirements dictate whether it is important to address the issues of how many clusters exist, what the best partition is, and which the best clusters are. When using representative sampling, for example, for high-throughput screening in pharmaceutical research, the number of required clusters is usually set beforehand. Hence, it is necessary to generate only a reasonable partition from which to extract the required number of representative compounds. For analysis of an unknown data set in, say, a list of vendor compounds, the number of clusters is unknown. Hierarchical clustering with optimum level analysis should provide suitable results for this scenario since the actual composition of each cluster is not critical. For analysis of quantitative structure–activity relationships (QSAR), the number of clusters is unknown, and the quality of the clusters becomes an important issue since complete clusters are required for further analysis. It may be that recent developments[87–93] related to density-based clustering will help in this circumstance.

CHEMICAL APPLICATIONS

Having introduced and described the various kinds of clustering methods used in chemistry and other disciplines, we are in a position to present some illustrative examples of chemical applications. This section reviews a representative selection of publications that have reported or analyzed the use of clustering methods for processing chemical data sets, largely from groups of scientists working within pharmaceutical companies. The main applications for these scientists are high-throughput screening, combinatorial chemistry, compound acquisition, and QSAR. The emphasis is on pharmaceutical applications because these workers tend to process very large and high dimensional data sets. This section is presented according to method, starting with hierarchical and then moving to nonhierarchical methods.

Little has been reported on the use of hierarchical divisive methods for processing chemical data sets (other than the inclusion of the minimum-diameter method in some of the comparative studies mentioned above). Recursive partitioning, which is a supervised classification technique very closely related to monothetic divisive clustering, has, however, been used at the GlaxoSmithKline[57,58] and Organon[59] companies.

There is, however, widespread use of hierarchical agglomerative techniques, particularly the Ward method. At Organon, Bayada, Hamersma, and van Geerestein[121] compared Ward clustering with the MaxMin diversity selection method, Kohonen maps, and a simple partitioning method to help select diverse yet representative subsets of compounds for further testing. The data came from HTS or combinatorial library results. Ward clustering

was the only method that gave results consistently better than random selection of compounds. It was also found that the standard technique of selecting the compound closest to the centroid to serve as the representative for a cluster tends to result in the selection of the smallest compound or the one with the fewest features. This finding is not surprising because the centroid is the arithmetic average of items in a cluster and hence the representative will be the most common denominator. Users should be aware of this tendency toward biased selection of a representative compound, since such a compound could be less interesting as a drug-like molecule than others in the data set. This effect was not observed when the clustering was done using the first 10 principal components of the descriptor set rather than relying directly on the descriptors (such as fingerprints) themselves.

Van Geerestein, Hamersma, and van Helden[122] used Ward clustering to show that cluster representatives provide a significantly better sampling of activity space than random selection. This key paper shows how clustering can separate actives from inactives in a data set, so that a cluster containing at least one active will contain more than an average number of other actives. The introduction to their article also gives a succinct summary of why diversity analysis (such as clustering) is of use as a lead finding strategy.

At Parke-Davis (now Pfizer), Wild and Blankley[123] incorporated Ward clustering and level selection (by the Kelley function[118]) into a program called VisualiSAR, which supports structure browsing and the development of structure–activity relationships in HTS data sets. At the Janssen unit of the Johnson and Johnson company, Engels et al.[124] have similarly incorporated Ward clustering and the Kelley function into a system (called CerBeruS) that is used for analysis of their corporate compound database. The clustering was used to produce smaller, more homogeneous subsets from which one representative compound was selected as a screening candidate using the Kelley function to determine the optimal clustering level(s). Engels et al.[124] noted two further advantages of a cluster-based approach. First, if a hit was found, related compounds could be tested subsequently by extracting other possible candidates from the cluster containing the hit, and, second, analyses of structure–activity relationships (SAR) could be formulated by linking the results of all the screening runs so as to examine the cluster hierarchy at different levels. Engels and Venkatarangan[125] subsequently developed a two-stage sequential screening procedure supported by clustering to make HTS more efficient.

Stanton et al.[126] reported the use of complete-link clustering in the HTS system at the Proctor & Gamble company. In situations where the screening produces large numbers of hits, clustering was used to determine which compound classes were present so that representatives could be taken. The amount of follow-up analysis was reduced by an order of magnitude while still evaluating which classes of compounds were present in the hits, thus increasing the efficiency of selecting potential leads. The clusters also provided sets of compounds to build preliminary SAR models. Furthermore, the clustering was

found useful in the detection of false positives, especially from combinatorial libraries. In these cases, the structural similarity between the hits was low and their biological activity was subsequently attributed to a common side product. Clustering was performed by Stanton[127] using BCUT (Burden–CAS–University of Texas) descriptors,[128] with the optimum hierarchy level determined visually from the dendrogram. Visual selection was possible because the hit sets were typically a few hundred compounds.

The most significant application of a nonhierarchical single-pass method was for screening antitumor activity at the National Cancer Institute. A variant of the leader algorithm was developed[129] in which the descriptors were weighted by occurrence in each compound, size of the fragment, and frequency of occurrence in the data set. Because of the use of these weighted descriptors, an asymmetric coefficient[129] was used to determine similarity, rather than the more usual Tanimoto coefficient. The data set was then ordered by the increasing sum of fragment weights to remove the order dependency associated with the leader algorithm (or at least, to have a reasonable basis for choosing a particular order) and to enable the use of heuristics to reduce the number of similarity calculations. Compounds were then assigned to any existing cluster for which they exceeded the given similarity threshold, thus creating overlapping clusters. The algorithm was implemented on parallel hardware,[105] and the results from clustering several data sets were presented with a discussion on the large number of singleton clusters produced.[130] Another variant on the leader algorithm was proposed by Butina.[131] In his approach, the compounds are first sorted by decreasing number of near neighbors (within a specified threshold similarity), thus again removing the order dependence of the basic algorithm. Of course, identifying the number of near neighbors for each compound introduces an $O(N^2)$ step, which in turn obviates the single-pass algorithm's primary advantage of linear speed.

At Rohm and Haas Company, Reynolds, Drucker, and Pfahler[132] developed a two-pass method similar to the initial assignment stage of k-means. In the first pass, a similarity threshold is specified, and then the sphere exclusion diverse subset selection method[80] is used to select the cluster seeds (referred to as *probes*). In the second pass, all other compounds are assigned to the most similar probe (the published version unnecessarily performs this in two stages). Clark and Langton[133] adopted a similar methodology in the Tripos OptiSim fast clustering system for selecting diverse yet representative subsets. OptiSim works by selecting an initial seed at random, selecting a random sample of size K, analyzing the random sample by choosing the most dissimilar member of the sample from existing seeds, and, if the minimum similarity threshold, R, to all existing seeds is exceeded, adding it to the seed set. This operation continues until the specified number of seeds, M, has been selected or no more candidates remain. All other compounds are then assigned to their nearest seed (which is equivalent to the initial assignment stage of k-means, with no refinement). OptiSim is an obvious amalgam of the MaxMin and sphere

exclusion subset selection methods[80] and the Reynolds system mentioned above. It also bears similarities with other methods, particularly the clustering of merged multiple random samples reported by Bradley and Fayyad.[115]

The widespread application of the Jarvis–Patrick nonhierarchical method exists in part because of the influence of the publications by Willett et al.[5,108–110] but also because of the availability of the efficient commercial implementation from Daylight[14] for handling very large data sets. The first publication on the use of Jarvis–Patrick clustering for compound selection from large chemical data sets was from researchers who implemented it at Pfizer Central Research (UK).[134] Clustering was done using 2D fragment descriptors, with calculation of the list of 20 nearest neighbors using the efficient Perry–Willett inverted file approach.[35] After clustering the data set of about 240,000 compounds, singletons were moved to the most similar nonsingleton cluster, and representative compounds were then extracted by generating cluster centroids and selecting the compound closest to each centroid.

Earlier in this chapter, we mentioned the cascaded Jarvis–Patrick[63] and fuzzy Jarvis–Patrick[64] variants. The cascaded Jarvis–Patrick method was implemented at Rhone-Poulenc Rorer (RPR) based on using Daylight 2D structural fingerprints and Daylight's Jarvis–Patrick program. With this variant, singletons are reclustered using less strict parameters so that the singletons do not dominate the set of representative compounds selected. The applications reported by the RPR researchers[63] include selection of compounds from the corporate database for HTS and comparison of the corporate database with external databases, such as the Available Chemicals Directory, to assist in compound acquisition. The fuzzy Jarvis–Patrick variant was developed and implemented at G. D. Searle and Company for analysis of their compound database to help support their screening program. The Searle researchers[64] initially used the Daylight implementation but found the chaining and singleton characteristics of the standard method to be significant drawbacks. This in turn prompted them to develop a variant with different characteristics.

McGregor and Pallai[135] discussed an in-house implementation of the standard Jarvis–Patrick algorithm at Procept Inc. They used the MDL 2D structural descriptors to compare and analyze external databases for efficient compound acquisition. Shemetulskis et al.[136] also reported the use of Jarvis–Patrick clustering to assist in compound acquisition at Parke-Davis, giving results from analysis and comparison of the CAST3D and Maybridge compound databases with the corporate database. In a two-stage process, representatives, comprising about a quarter of the compounds, were selected from each data set by clustering on the basis of 2D fingerprints. Each data set was then merged with the corporate database, and the clustering run again on the basis of calculated physicochemical property descriptors. Clusters containing only CAST3D or Maybridge compounds were tagged as highest priority for acquisition. Dunbar[137] summarized the compound acquisition

report,[135] discussed the use of clustering methods to assist in HTS, and then outlined the use at Parke-Davis of Jarvis–Patrick clustering to assist traditional, low-throughput screening. The aim of the Parke-Davis group was to generate a representative subset of no more than 2000 compounds selected from about 126,000 compounds in the Parke-Davis corporate database so that they could be used in a particularly labor-intensive cell-based assay. Jarvis–Patrick clustering was run to generate an initial set of 25,000 non-singleton clusters. The compounds closest to the centroids were reclustered to give about 2,300 clusters. The compounds closest to these centroids were then analyzed manually providing a final selection of about 1,400 compounds. An interesting feature of this process was that singletons were rejected at each stage, rather than being assigned to the nearest nonsingleton cluster (as at Pfizer, UK) or being reclustered separately (as in the cascaded clustering method used at Rhone-Poulenc Rorer).

Jarvis–Patrick clustering has also been used to support QSAR analysis in a system developed at the European Communities Joint Research Center.[7,138–140] The EINECS (European Inventory of Existing Chemical Substances) database contains more than 100,000 compounds and has been clustered using 2D structural descriptors. That database also has associated physicochemical properties and activities, but the data is very sparse. Jarvis–Patrick clustering was used to extract clusters containing sufficient compounds with measured data for an attempt to be made to estimate the properties of members of the cluster lacking the data. For a few clusters, it was used to develop reasonable QSAR models.

An example of how use of k-means clustering can be used for QSAR analysis of small data sets is that by Lawson and Jurs[141] who clustered a set of 143 acrylates from the ToSCA (Toxic Substances Control Act) inventory. For large chemical data sets, the seminal paper is that published by Higgs et al.,[79] at Eli Lilly and Company. These authors examined three methods of subset selection to assist their HTS and development of combinatorial libraries. The three methods were k-means, MaxMin, and D-optimal design. Seed compounds were selected by the MaxMin method, and the k-means algorithm was implemented on parallel hardware. This research was part of the compound acquisition strategy to support HTS. The Lilly group used an extensive system of filters to ensure that selected compounds were pharmaceutically acceptable. No recommendations were offered in the paper as to the best method.

The use of a topographic clustering method for chemical data sets is exemplified by the work of Sadowski, Wagener, and Gasteiger.[142] The authors compared three combinatorial libraries using Kohonen mapping. Each compound within a library was represented by a 12-element autocorrelation vector (a sort of 3D-QSAR descriptor). The vectors were used as input to a 50×50 Kohonen network. Mapping the combinatorial libraries onto the same network placed each compound from the library at a particular node in the network. A 2D display of the positions of each compound revealed the degree of

overlap between the libraries. Two very dissimilar libraries formed two distinct clusters with little overlap, whereas two very similar libraries showed no distinction.

The use of mixture-model or density-based clustering has not yet been reported for processing chemical data sets. An interesting application of these techniques is their use to group the compound descriptors so as to obtain a set of orthogonal descriptors. Up to this point, the clustering that we have discussed has been applied to the patterns (fingerprints or dataprints) characterizing each compound; this is the "*Q-mode clustering*" referred to by Sneath and Sokal.[1] One can also cluster the features (the descriptors used in the fingerprints or dataprints) to highlight groups of similar descriptors. Sneath and Sokal call this "*R-mode clustering*." The similar property principle, upon which structure–property relationships depend, assumes that the compound descriptors are independent of each other. Reducing the number of descriptors can thus help in subsequent Q-mode clustering by reducing the dimensionality. Clustering the descriptors, so that a subset of orthogonal descriptors can be extracted, is an alternative to factor analysis and principal components analysis. Using an orthogonal subset of descriptors has the benefit that the result is a set of individual descriptors rather than composite descriptors. Taraviras, Ivanciuc, and Cabrol-Bass[65] applied the single-link, group-average, complete-link, and Ward hierarchical methods, along with Jarvis–Patrick, variable-length Jarvis–Patrick, and k-means nonhierarchical methods to a set of 240 topological indices in an attempt to reveal any "natural" clusters of the descriptors. Descriptors that were found to exist in the same clusters across all seven methods were regarded as being strongly clustered. Reducing the number of methods that needed to be in agreement revealed progressively weaker clusters. Overall, it was found that the strategy of using multiple clustering methods for R-mode clustering could be used to provide representative sets of orthogonal descriptors for use in QSAR analysis.

CONCLUSIONS

Clustering methodology has been developed over many decades. The application of clustering to chemical data sets began in the 1980s, coinciding with the increasing size of in-house compound collections having their information contained in structural databases and with advances made by the information retrieval community to analyze large document collections. In the 1990s the advent of high-throughput screening, combinatorial libraries, and commercially available external chemical inventories placed a greater emphasis on rational compound selection. The demands of clustering data sets of several million compounds with high-dimensional representations led to the widespread adoption of a few inherently efficient and optimally implemented methods, namely, the Jarvis–Patrick, Ward, and k-means methods.

Acceptance of these methods—and inclusion of them as routine operations within such applications as lead-finding strategies, QSAR analyses, and compound acquisition—has been a gradual process rather than an abrupt revolution. The current decade should see this process continue as the methodologies are refined. The push for such advancement appears to be coming again from the information retrieval community but also from the data mining community, which has made significant progress. The emphasis of current research is turning toward the quality of the resultant clusters. It has been shown that, using representatives selected from clusters for lead-finding can increase the active hit rate significantly and consistently.

The results so far in chemistry are promising, but research in other areas outside of chemistry suggests that clustering is still a blunt instrument that can be sharpened by refinements. An example of this refinement is to be able to handle mixed or nonnumerical data, and another example is to take more consideration of cluster sizes, shapes, and distribution. The existing methods and implementations used to analyze chemical data sets do an impressive job when compared with the situation a decade ago. What is exciting is the number of new ideas that are being generated that should result in significant advances in the next decade.

REFERENCES

1. P. H. A. Sneath and R. R. Sokal, *Numerical Taxonomy,* W. H. Freeman, San Francisco, CA, 1973.

2. L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis,* Wiley-Interscience, New York, 1990.

3. B. S. Everitt, *Cluster Analysis,* 3rd ed., Edward Arnold, London, 1993.

4. A. D. Gordon, *Classification,* 2nd ed., Chapman and Hall, London, 1999.

5. P. Willett, *Similarity and Clustering in Chemical Information Systems,* Research Studies Press, Letchworth, UK, 1987.

6. N. Bratchell, *Chemom. Intell. Lab. Systems,* **6**, 105 (1989). Cluster Analysis.

7. J. M. Barnard and G. M. Downs, *J. Chem. Inf. Comput. Sci.,* **32** (6), 644 (1992). Clustering of Chemical Structures on the Basis of Two-Dimensional Similarity Measures.

8. G. M. Downs and P. Willett, in *Advanced Computer-Assisted Techniques in Drug Discovery,* H. van de Waterbeemd, Ed., VCH Publishers, Weinheim, 1994, pp. 111–130. Clustering of Chemical Structure Databases for Compound Selection.

9. R. A. Lewis, S. D. Pickett, and D. E. Clark, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., Wiley-VCH, New York, 2000, Vol. 16, pp. 1–51. Computer-Aided Molecular Diversity Analysis and Combinatorial Library Design.

10. Clustan Ltd., 16 Kingsburgh Road, Edinburgh, UK. http://www.clustan.com.

11. SAS Institute Inc., SAS Campus Drive, Cary, NC 27513, USA. http://www.sas.com.

12. Barnard Chemical Information Ltd., 46 Uppergate Road, Stannington, Sheffield S6 6BX, UK. http://www.bci.gb.com.

13. Chemical Computing Group Inc., 1010 Sherbrooke Street West, Suite 910, Montreal, Quebec H3A 2R7, Canada. http://www.chemcomp.com.

14. Daylight Chemical Information Systems Inc., 441 Greg Avenue, Santa Fe, NM 87501, USA. http://www.daylight.com.

15. MDL Information Systems, Inc., 14600 Catalina Street, San Leandro, CA 94577, USA. http://www.mdl.com.

16. Accelrys (formerly Molecular Simulations Inc.), 9685 Scranton Road, San Diego, CA 92121-3752, USA. http://www.accelrys.com.

17. Tripos Inc., 1699 South Hanley Road, St. Louis, MO 63144, USA. http://www.tripos.com.

18. W. A. Warr, in *Computational Methods for the Analysis of Molecular Diversity*, P. Willett, Ed., *Perspectives in Drug Discovery and Design,* Vol. 7/8, Kluwer/ESCOM, Dordrecht, The Netherlands, 1997, pp. 115–130. Commercial Software Systems for Diversity Analysis.

19. R. D. Brown, in *Computational Methods for the Analysis of Molecular Diversity,* P. Willett, Ed., *Perspectives in Drug Discovery and Design,* Vol. 7/8, Kluwer/ESCOM, Dordrecht, The Netherlands, 1997, pp. 31–49. Descriptors for Diversity Analysis.

20. R. D. Brown and Y. C. Martin, *J. Chem. Inf. Comput. Sci.,* **36** (*3*), 572 (1996). Use of Structure–Activity Data to Compare Structure-Based Clustering Methods and Descriptors for Use in Compound Selection.

21. R. D. Brown and Y. C. Martin, *J. Chem. Inf. Comput. Sci.,* **37** (*1*), 1 (1997). The Information Content of 2D and 3D Structural Descriptors Relevant to Ligand-Receptor Binding.

22. G. M. Downs and P. Willett, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1995, Vol. 7, pp. 1–66. Similarity Searching in Databases of Chemical Structures.

23. P. Willett, J. M. Barnard, and G. M. Downs, *J. Chem. Inf. Comput. Sci.,* **38** (*6*), 983 (1998). Chemical Similarity Searching.

24. J. M. Barnard, G. M. Downs, and P. Willett, in *Virtual Screening for Bioactive Molecules,* H.-J. Böhm and G. Schneider., Eds., Wiley, New York, 2000, pp. 59–80. Descriptor-Based Similarity Measures for Screening Chemical Databases.

25. J. S. Mason and S. D. Pickett, in *Computational Methods for the Analysis of Molecular Diversity,* P. Willett, Ed., *Perspectives in Drug Discovery and Design,* Vol. 7/8, Kluwer/ESCOM, Dordrecht, The Netherlands, 1997, pp. 85–114. Partition-Based Selection.

26. J. A. Hartigan, *Clustering Algorithms,* Wiley, New York, 1975.

27. F. Murtagh, *COMPSTAT Lectures,* **4,** (1985). Multidimensional Clustering Algorithms.

28. A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data,* Prentice Hall, Englewood Cliff, NJ, 1988.

29. G. N. Lance and W. T. Williams, *Computer J.,* **9,** 373 (1967). A General Theory of Classificatory Sorting Strategies. 1. Hierarchical Systems.

30. J. H. Ward, *J. Am. Stat. Assoc.,* **58,** 236 (1963). Hierarchical Grouping to Optimize an Objective Function.

31. E. M. Voorhees, *Inf. Processing Management,* **22** (*6*), 465 (1986). Implementing Agglomerative Hierarchic Clustering Algorithms for Use in Document Retrieval.

32. X. Chen, A. Rusinko III, and S. S. Young, *J. Chem. Inf. Comput. Sci.,* **38** (*6*), 1054 (1998). Recursive Partitioning Analysis of a Large Structure-Activity Data Set Using Three-Dimensional Descriptors.

33. A. Guenoche, P. Hansen, and B. Jaumard, *J. Classification,* **8,** 5 (1991). Efficient Algorithms for Divisive Hierarchical Clustering with the Diameter Criterion.

34. R. A. Jarvis and E. A. Patrick, *IEEE Trans. Computers,* **C-22** (*11*), 1025 (1973). Clustering Using a Similarity Measure Based on Shared Near Neighbors.

35. S. A. Perry and P. Willett, *J. Inf. Sci.,* **6,** 59 (1983). A Review of the Use of Inverted Files for Best Match Searching in Information Rretrieval Systems.

36. A. P. Dempster, N. M. Laird, and D. B. Rubin, *J. Royal Stat. Soc.,* **B39,** 1 (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm.

37. T. Kohonen, *Self-organizing Maps,* Springer-Verlag, Berlin, 1995.

38. J. Zupan and J. Gasteiger, *Neural Networks for Chemists, An Introduction,* VCH, Weinheim, 1993. See also, K. L. Peterson, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., Wiley-VCH, New York, 2000, Vol. 16, pp. 53–140. Artificial Neural Networks and Their Use in Chemistry.

39. F. Murtagh, in *Handbook of Massive Data Sets,* J. Abello, P. M. Pardalos, and M. G. C. Reisende, Eds., Kluwer, Dordrecht, The Netherlands, 2001, pp. 401–545. Clustering in Massive Data Sets.

40. D. W. Matula, in *Classification as a Tool of Research,* W. Gaul and M. Schader, Eds., Elsevier Science (North-Holland), Amsterdam, 1986, pp. 289–301. Divisive vs. Agglomerative Average Linkage Hierarchical Clustering.

41. N. C. Jain, A. Indrayan, and L. R. Goel, *Pattern Recognition,* **19** (*1*), 95 (1986). Monte Carlo Comparison of Six Hierarchical Clustering Methods on Random Data.

42. J. Podani, *Vegetatio,* **81,** 61 (1989). New Combinatorial Clustering Methods.

43. M. Roux, in *Applied Multivariate Analysis in SAR and Environmental Studies,* J. Devillers and W. Karcher, Eds., Kluwer, Dordrecht, The Netherlands, 1991, pp. 115–135. Basic Procedures in Hierarchical Cluster Analysis.

44. A. El-Hamdouchi and P. Willett, *Computer J.,* **32,** 220 (1989). Hierarchic Document Clustering using Ward's Method.

45. E. M. Rasmussen and P. Willett, *J. Doc.,* **45** (*1*), 1 (1989). Efficiency of Hierarchical Agglomerative Clustering Using the ICL Distributed Array Processor.

46. G. M. Downs, P. Willett, and W. Fisanick, *J. Chem. Inf. Comput. Sci.,* **34** (*5*), 1094 (1994). Similarity Searching and Clustering of Chemical-Structure Databases Using Molecular Property Data.

47. S. Guha, R. Rastogi, and K. Shim, Technical Report, Bell Laboratories, Murray Hill, NJ, 1997. A Clustering Algorithm for Categorical Attributes.

48. S. Guha, R. Rastogi, and K. Shim, *Inf. Systems,* **25** (*5*), 345 (2000). ROCK: A Robust Clustering Algorithm for Categorical Attributes.

49. S. Guha, R. Rastogi, and K. Shim, in *Proceedings of the ACM SIGMOD International Conference on Management of Data,* Seattle, WA, 1998, pp. 73–84. CURE: An Efficient Clustering Algorithm for Large Datasets.

50. G. Karypis, E.-H. Han, and V. Kumar, *IEEE Computer: Special Issue on Data Analysis and Mining,* **32** (*8*), 68 (1999). Chameleon: A Hierarchical Clustering Algorithm Using Dynamic Modeling.

51. G. Karypis, E.-H. Han, and V. Kumar, *Technical Report No. 99-020,* Department of Computer Science & Engineering, University of Minnesota, Minneapolis, MN, 1999. Multilevel Refinement for Hierarchical Clustering.

52. D. Fasulo, *Technical Report No. 01-03-02,* Department of Computer Science & Engineering, University of Washington, Seattle, WA, 1999. An Analysis of Recent Work on Clustering Algorithms.

53. J. MacCuish, C. Nicolaou, and N. E. MacCuish, *J. Chem. Inf. Comput. Sci.,* **41** (*1*), 134 (2001). Ties in Proximity and Clustering Compounds.

54. J. A. Garcia, J. Fdez-Valdivia, J. F. Cortijo, and R. Molina, *Signal Processing,* **44** (*2*), 181 (1994). A Dynamic Approach for Clustering Data.

55. Y. Wang, H. Yan, and C. Sriskandarajah, *J. Classification,* **13,** 231 (1996). The Weighted Sum of Split and Diameter Clustering.

56. M. Steinbach, G. Karypis, and V. Kumar, *Technical Report 00-034,* Department of Computer Science & Engineering, University of Minnesota, Minneapolis, MN, 2000. A Comparison of Document Clustering Techniques.

57. D. M. Hawkins, S. S. Young, and A. Rusinko, *Quant. Struct.–Act. Relat.,* **16,** 396 (1997). Analysis of a Large Structure–Activity Data Set Using Recursive Partitioning.

58. X. Chen, A. Rusinko, and S. S. Young, *J. Chem. Inf. Comput. Sci.,* **38** (*6*), 1054 (1998). Recursive Partitioning Analysis of a Large Structure–Activity Data Set Using Three-Dimensional Descriptors.

59. M. Wagener and V. J. van Geerestein, *J. Chem. Inf. Comput. Sci.,* **40** (2), 280 (2000). Potential Drugs and Nondrugs: Prediction and Identification of Important Structural Features.

60. D. W. Miller, *J. Chem. Inf. Comput. Sci.,* **41** (1), 168 (2001). Results of a New Classification Algorithm Combining K Nearest Neighbors and Recursive Partitioning.

61. T. Zhang, R. Ramakrishnan, and M. Livny, *ACM SIGMOD Record,* **25** (2), 103 (1996). BIRCH: An Efficient Data Clustering Method for Very Large Databases.

62. V. Ganti, R. Ramakrishnan, J. Gehrke, A. Powell, and J. French, *Proceedings of the 15th International Conference on Data Engineering,* Sydney, Australia, 1999, pp. 502–511. Clustering Large Datasets in Arbitrary Metric Spaces.

63. P. R. Menard, R. A. Lewis, and J. S. Mason, *J. Chem. Inf. Comput. Sci.,* **38** (3), 379 (1998). Rational Screening Set Design and Compound Selection: Cascaded Clustering.

64. T. N. Doman, J. M. Cibulskis, M. J. Cibulskis, P. D. McCray, and D. P. Spangler, *J. Chem. Inf. Comput. Sci.,* **36** (6), 1195 (1996). Algorithm5: A Technique for Fuzzy Clustering of Chemical Inventories.

65. S. L. Taraviras, O. Ivanciuc, and D. Cabrol-Bass, *J. Chem. Inf. Comput. Sci.,* **40** (5), 1128 (2000). Identification of Groupings of Graph Theoretical Descriptors Using a Hybrid Cluster Analysis Approach.

66. K. C. Gowda and G. Krishna, *Pattern Recognition,* **10** (2), 105 (1978). Agglomerative Clustering Using the Concept of Mutual Nearest Neighborhood.

67. R. Dugad and N. Ahuja, *IEEE International Conference on Acoustics Speech and Signal Processing,* **5**, 2761 (1998). Unsupervised Multidimensional Hierarchical Clustering.

68. E. Forgy, *Biometrics,* **21**, 768 (1965). Cluster Analysis of Multivariate Data: Efficiency vs. Interpretability of Classifications.

69. J. MacQueen, in *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability,* University of California Press, Berkeley, CA, 1967, Vol. 1, pp. 281–297. Some Methods for Classification and Analysis of Multivariate Observations.

70. J. A. Hartigan and M. A. Wong, *Appl. Stat.,* **28**, 100 (1979). A K-Means Clustering Algorithm.

71. H. Spaeth, *Eur. J. Operat. Res.,* **1**, 23 (1977). Computational Experiences with the Exchange Method.

72. R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis,* Wiley, New York, 1973.

73. M. A. Ismail and M. S. Kamel, *Pattern Recognition,* **22**, 75 (1989). Multidimensional Data Clustering Utilizing Hybrid Search Strategies.

74. Q. Zhang, Q. R. Wang, and R. D. Boyle, *Pattern Recognition,* **24** (4), 331 (1991). A Clustering Algorithm for Datasets with a Large Number of Classes.

75. Z. Huang, *Int. J. Data Mining Knowledge Disc.,* **2** (3), 283 (1998). Extensions to the K-Means Algorithm for Clustering Large Data Sets with Categorical Values.

76. Q. Zhang and R. D. Boyle, *Pattern Recognition,* **24** (9), 835 (1991). A New Clustering Algorithm with Multiple Runs of Iterative Procedures.

77. V. Estivell-Castro and J. Yang, *Technical Report No. 99-03,* Department of Computer Science & Software Engineering, University of Newcastle, Callaghan, NSW 2308, Australia. A Fast and Robust General Purpose Clustering Algorithm.

78. R. T. Ng and J. Han, in *Proceedings of the 20th International Conference on Very Large Data Bases,* Santiago, Chile, 1994, pp. 144–155. Efficient and Effective Clustering Methods for Spatial Data Mining.

79. R. E. Higgs, K. G. Bemis, I. A. Watson, and J. H. Wikel, *J. Chem. Inf. Comput. Sci.,* **37** (5), 861 (1997). Experimental Designs for Selecting Molecules from Large Chemical Databases.

80. M. Snarey, N. K. Terrett, P. Willett, and D. J. Wilton, *J. Mol. Graphics Modell.,* **15**, 372 (1997). Comparison of Algorithms for Dissimilarity-Based Compound Selection.

81. D. Fisher, L. Xu, and N. Zard, in *Proceedings of the 6th International Workshop on Machine Learning,* Morgan Kaufmann, Aberdeen, UK, 1992, pp. 163–168. Ordering Effects in Clustering.

82. G. W. Milligan, *Psychometrika,* **45** (*3*), 325 (1980). An Examination of the Effect of Six Types of Error Perturbation on Fifteen Clustering Algorithms.

83. D. Fisher, *J. Artif. Intell. Res.,* **4**, 147 (1996). Iterative Optimization and Simplification of Hierarchical Clusterings.

84. G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions,* Wiley, New York, 1997.

85. J. D. Banfield and A. E. Raftery, *Biometrics,* **49,** 803 (1993). Model-Based Gaussian and Non-Gaussian Clustering.

86. C. Fraley and A. E. Raftery, *Computer J.,* **41** (*8*), 578 (1988). How Many Clusters? Which Clustering Method? Answers via Model-Based Cluster Analysis.

87. M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining,* Portland, OR, 1996, pp. 226–231. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.

88. M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, in *Proceedings of the ACM SIGMOD International Conference on Management of Data,* Philadelphia, PA, 1999, pp. 49–60. OPTICS: Ordering Points to Identify Clustering Structure.

89. M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, *Proceedings of the Conference on Data Mining and Knowledge Discovery,* Prague, Czech Repub., 1999, in *Lecture Notes in Computer* Science, Springer, **1704,** 262–270 (1999). OPTICS-OF: Identifying Local Outliers.

90. M. M. Breunig, H.-P. Kriegel, and J. Sander, in *Proceedings of the European Conference on Principles and Practice of Knowledge Discovery in Databases,* Lyon, France, 2000. Fast Hierarchical Clustering Based on Compressed Data and OPTICS.

91. R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan, in *Proceedings of the ACM SIGMOD International Conference on Management of Data,* Seattle, WA, 1998, pp. 94–105. Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications.

92. C. C. Aggarwal, C. Procopiuc, J. L. Wolf, P. S. Yu, and J. S. Park, in *Proceedings ACM SIGMOD International Conference on Management of Data,* Philadelphia, PA, 1999, pp. 61–72. Fast Algorithms for Projected Clustering.

93. H. S. Nagesh, M.Sc. Thesis, Northwestern University of Illinois, Evanston, IL, 1999. High Performance Subspace Clustering for Massive Data Sets.

94. D. W. Matula, in *Classification and Clustering,* J. van Ryzin, Ed., Academic Press, 1977, pp. 95–129. Graph Theoretic Techniques for Cluster Analysis Algorithms.

95. A. Ben-Dor, R. Shamir, and Z. Yakhini, *J. Comput. Biol.,* **6** (*3/4*), 281 (1999). Clustering Gene Expression Patterns.

96. E. Hartuv, A. Schmitt, J. Lange, S. Meier-Ewart, H. Lehrach, and R. Shamir, in *Proceedings 3rd International Conference on Computational Molecular Biology (RECOMB 99),* Lyon, France, 1999. An Algorithm for Clustering cDNAs for Gene Expression Analysis.

97. R. Sharan and R. Shamir, in *Proceedings of the 8th International Conference on Intelligent Systems for Molecular Biology,* AAAI Press, Menlo Park, CA, 2000, pp. 307–316. CLICK: A Clustering Algorithm with Application to Gene Expression Analysis.

98. I. Jonyer, L. B. Holder, and D. J. Cook, in *Proceedings of the 13th Annual Florida AI Research Symposium,* pp. 91–95, 2000 (http://www-cse.uta.edu/~cook/pubs). Graph-Based Hierarchical Conceptual Clustering.

99. F. Murtagh, *Computer J.,* **26** (*4*), 354 (1983). A Survey of Recent Advances in Hierarchical Clustering Algorithms.

100. E. M. Rasmussen, G. M. Downs, and P. Willett, *J. Comput. Chem.,* **9** (*4*), 378 (1988). Automatic Classification of Chemical Structure Databases Using a Highly Parallel Array Processor.

101. X. Li and Z. Fang, *Parallel Computing,* **11**, 275 (1989). Parallel Clustering Algorithms.

102. X. Li, *IEEE Trans. Pattern Anal. Machine Intelligence,* **12** (*11*), 1088 (1990). Parallel Algorithms for Hierarchical Clustering and Cluster Validity.

103. F. Murtagh, *IEEE Trans. Pattern Anal. Machine Intelligence,* **14** (*10*), 1056 (1992). Comments on "Parallel Algorithms for Hierarchical Clustering and Cluster Validity".

104. C. F. Olson, *Technical Report CSD-94-786,* University of California, Berkeley, CA, 1994. Parallel Algorithms for Hierarchical Clustering.

105. R. Whaley and L. Hodes, *J. Chem. Inf. Comput. Sci.,* **31** (2), 345 (1991). Clustering a Large Number of Compounds. 2. Using a Connection Machine.

106. A. McCallum, K. Nigam, and L. H. Ungar, in *Advances in Knowledge Discovery and Data Mining,* MIT Press, Cambridge, MA, 2000. Efficient Clustering of High Dimensional Data Sets with Application to Reference Matching.

107. M. A. Johnson and G. M. Maggiora, *Concepts and Applications of Molecular Similarity,* Wiley, New York, 1990.

108. P. Willett, *Anal. Chim. Acta,* **136,** 29 (1982). A Comparison of Some Hierarchical Agglomerative Clustering Algorithms for Structure–Property Correlation.

109. V. Rubin and P. Willett, *Anal. Chim. Acta,* **151,** 161 (1983). A Comparison of Some Hierarchical Monothetic Divisive Clustering Algorithms for Structure–Property Correlation.

110. P. Willett, *J. Chem. Inf. Comput. Sci.,* **24** (1), 29 (1984). Evaluation of Relocation Clustering Algorithms for the Automatic Classification of Chemical Structures.

111. R. G. Lawson and P. C. Jurs, *J. Chem. Inf. Comput. Sci.,* **30** (1), 36 (1990). New Index for Clustering Tendency and Its Application to Chemical Problems.

112. J. W. McFarland and D. J. Gans, *J. Med. Chem.,* **29,** 505–514 (1986). On the Significance of Clusters in Graphical Display of Structure–Activity Data.

113. S. Epter, M. Krishnamoorthy, and M. Zaki, *Technical Report No. 99-6,* Computer Science Department, Rensselaer Polytechnic Institute, Troy, NY, 1999. Clusterability Detection and Initial Seed Selection in Large Data Sets.

114. U. M. Fayyad, C. A. Reima, and P. S. Bradley, in *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining,* R. Agrawal, P. Stolorz, and G. Piatetsky-Shapiro, Eds., AAAI Press, Menlo Park, CA, 1998, pp. 194–198. Initialization of Iterative Refinement Clustering Algorithms.

115. P. S. Bradley and U. M. Fayyad, in *Proceedings of the 15th International Conference on Machine Learning,* Morgan Kaufmann, San Francisco, 1998, pp. 91–99. Refining Initial Points for K-Means Clustering.

116. G. W. Milligan and M. C. Cooper, *Psychometrika,* **50** (2), 159 (1985). An Examination of Procedures for Determining the Number of Clusters in a Data Set.

117. D. J. Wild and C. J. Blankley, *J. Chem. Inf. Comput. Sci.,* **40** (1), 155 (2000). Comparison of 2D Fingerprint Types and Hierarchy Level Selection Methods for Structural Grouping Using Ward's Clustering.

118. L. A. Kelley, S. P. Gardner, and M. J. Sutcliffe, *Protein Eng.,* **9,** 1063 (1996). An Automated Approach for Clustering an Ensemble of NMR-Derived Protein Structures into Conformationally-Related Subfamilies.

119. G. W. Milligan, *Psychometrika,* **46** (2), 187 (1981). A Monte Carlo Study of Thirty Internal Criterion Measures for Cluster Analysis.

120. T. Calinski and J. Harabasz, *Commun. Stat.,* **3** (1), 1 (1974). A Dendrite Method for Cluster Analysis.

121. D. M. Bayada, H. Hamersma, and V. J. van Geerestein, *J. Chem. Inf. Comput. Sci.,* **39** (1), 1 (1999). Molecular Diversity and Representativity in Chemical Databases.

122. V. J. van Geerestein, H. Hamersma, and S. P. van Helden, in *Computer-Assisted Lead Finding and Optimization* (Proceedings 9th European QSAR Meeting, Lausanne, Switzerland, 1996), H. van de Waterbeemd, B. Testa, and G. Folkers, Eds., Wiley-VCH, Basel, Switzerland, 1997, pp. 159–178. Exploiting Molecular Diversity: Pharmacophore Searching and Compound Clustering.

123. D. J. Wild and C. J. Blankley, *J. Mol. Graphics Modell.,* **17** (2), 85 (1999). VisualiSAR: A Web-Based Application for Clustering, Structure Browsing, and Structure–Activity Relationship Study.

124. M. F. M. Engels, T. Thielmans, D. Verbinnen, J. P. Tollenaere,, and R. Verbeeck, *J. Chem. Inf. Comput. Sci.,* **40** (*2*), 241 (2000). CerBeruS: A System Supporting the Sequential Screening Process.

125. M. F. M. Engels and P. Venkatarangan, *Curr. Opin. Drug Discovery Dev.*, **4** (*3*), 275 (2001). Smart Screening: Approaches to Efficient HTS.

126. D. T. Stanton, T. W. Morris, S. Roychoudhury, and C. N. Parker, *J. Chem. Inf. Comput. Sci.,* **39** (*1*), 21 (1999). Application of Nearest-Neighbor and Cluster Analyses in Pharmaceutical Lead Discovery.

127. D. T. Stanton, *J. Chem. Inf. Comput. Sci.,* **39** (*1*), 11 (1999). Evaluation and Use of BCUT Descriptors in QSAR and QSPR Studies.

128. R. S. Pearlman and K. M. Smith, *J. Chem. Inf. Comput. Sci.,* **39** (*1*) 28–35 (1999). Metric Validation and the Receptor-Relevant Subspace Concept.

129. L. Hodes, *J. Chem. Inf. Comput. Sci.,* **29** (*2*), 66 (1989). Clustering a Large Number of Compounds. 1. Establishing the Method on an Initial Sample.

130. R. Whaley and L. Hodes, *J. Chem. Inf. Comput. Sci.,* **31** (*2*), 347 (1991). Clustering a Large Number of Compounds. 3. The Limits of Classification.

131. D. Butina, *J. Chem. Inf. Comput. Sci.,* **39** (*4*), 747 (1999). Unsupervised Data Base Clustering Based on Daylight's Fingerprint and Tanimoto Similarity a Fast and Automated Way to Cluster Small and Large Data Sets.

132. C. H. Reynolds, R. Druker, and L. B. Pfahler, *J. Chem. Inf. Comput. Sci.,* **38** (*2*), 305 (1998). Lead Discovery Using Stochastic Cluster Analysis (SCA): A New Method for Clustering Structurally Similar Compounds.

133. R. D. Clark and W. J. Langton, *J. Chem. Inf. Comput. Sci.,* **38** (*6*), 1079 (1998). Balancing Representativeness Against Diversity Using Optimizable K-dissimilarity and Hierarchical Clustering.

134. P. Willett, V. Winterman, and D. Bawden, *J. Chem. Inf. Comput. Sci.,* **26** (*3*), 109 (1986). Implementation of Nonhierarchic Cluster-Analysis Methods in Chemical Information Systems; Selection of Compounds for Biological Testing and Clustering of Substructure Search Output.

135. M. J. McGregor and P. V. Pallai, *J. Chem. Inf. Comput. Sci.,* **37** (*3*), 443 (1997). Clustering Large Databases of Compounds: Using the MDL "Keys" as Structural Descriptors.

136. N. E. Shemetulskis, J. B. Dunbar, B. W. Dunbar, D. W. Moreland, and C. Humblet, *J. Comput.-Aided Mol. Design,* **9,** 407 (1995). Enhancing the Diversity of a Corporate Database using Chemical Database Clustering and Analysis.

137. J. B. Dunbar, in *Computational Methods for the Analysis of Molecular Diversity,* P. Willett, Ed., *Perspectives in Drug Discovery and Design,* Vol. 7/8, Kluwer/ESCOM, Dordrecht, The Netherlands, 1997, pp. 51–63. Cluster-Based Selection.

138. G. M. Downs and P. Willett, in *Applied Multivariate Analysis in SAR and Environmental Studies,* J. Devillers and W. Karcher, Eds., Kluwer, Dordrecht, The Netherlands, 1991, pp. 247–279. The Use of Similarity and Clustering Techniques for the Prediction of Molecular Properties.

139. J. Nouwen and B. Hansen, *SAR and QSAR in Environmental Research,* **4,** 1 (1995). An Investigation of Clustering as a Tool in Quantitative Structure–Activity Relationships (QSARs).

140. J. Nouwen, F. Lindgren, B. Hansen, W. Karcher, H. J. M. Verhaar, and J. L. M. Hermens, *J. Chemometrics,* **10,** 385 (1996). Fast Screening of Large Databases Using Clustering and PCA based on Structure Fragments.

141. R. G. Lawson and P. C. Jurs, *J. Chem. Inf. Comput. Sci.,* **30** (*2*), 137 (1990). Cluster Analysis of Acrylates to Guide Sampling for Toxicity Testing.

142. J. Sadowski, M. Wagener, and J. Gasteiger, *Angew. Chem., Int. Ed. Engl.,* **34** (*23/24*), 2674 (1995/1996). Assessing Similarity and Diversity of Combinatorial Libraries by Spatial Autocorrelation Functions and Neural Networks.

CHAPTER 2

The Use of Scoring Functions in Drug Discovery Applications

Hans-Joachim Böhm and Martin Stahl

F. Hoffmann-La Roche AG, Pharmaceuticals Division, Chemical Technologies, CH-4070 Basel, Switzerland

INTRODUCTION

Structure-based design has become a mature and integral part of medicinal chemistry. It has been convincingly demonstrated for a large number of targets that the three-dimensional (3D) structure of a protein can be used to design small molecule ligands binding tightly at this target. Indeed, several marketed compounds can be attributed to a successful structure-based design.[1–4] Several reviews summarize these results.[5–9]

Since the introduction of molecular modeling and structure-based design into the drug discovery process in the 1980s, there has been a significant change in the role these computational techniques are playing. Early molecular modeling work concentrated on the manual design of protein ligands using the 3D structure of a target. Usually, the creativity of the designer was used to build a novel putative ligand using computer graphics followed by a molecular mechanics calculation of the resulting protein–ligand complex. A geometric and energetic analysis of the energy-minimized complex was then used to assess the putative ligand. A good complementarity of the shape and surface properties between the protein and ligand was used as an indication that the ligand might indeed bind to the protein with high affinity.

However, designing a single, active, synthetically accessible compound turned out to be a greater challenge than expected. It is still difficult to computationally predict induced-fit phenomena and binding affinities for new ligand candidates. But while existing modeling tools are certainly not suitable to design the one perfect drug molecule, they can help to *enrich* sets of molecules with greater numbers of biologically active candidates, even though the rates of false positives (and false negatives) are still high. Thus, an important current goal of molecular design is to increase the hit rate in biological assays compared to random compound selections, which means that structure-based design approaches now focus on the processing of large numbers of molecules. These "virtual libraries" of molecules can consist of either existing molecules (e.g., the compound collection of a pharmaceutical company) or of putative novel structures that could be synthesized via combinatorial chemistry. The computational goal is to rapidly assess millions of possible molecules by filtering out the majority that are predicted to be extremely unlikely to bind, and then to prioritize the remaining ones. This approach is, in fact, a successful strategy, and several recent publications have demonstrated impressive enrichment of active compounds.[10–15] The change of focus from individual compounds to compound libraries has been supported by three major developments that have taken place since the early days of molecular design:

1. An exponentially growing number of 3D protein structures is available in the public domain. Consequently, the number of projects relying on structural information has increased, and structure-based ligand design is nowadays routinely carried out at all major pharmaceutical companies. The amount of structural knowledge is so large that automated methods are needed to make full use of it.
2. High throughput screening (HTS) has become a well-established process. Large libraries of several hundred thousand compounds are routinely tested against new targets. This biological testing can, in many cases, be carried out in less than one month.
3. Synthetic chemistry has undergone a major change with the introduction of combinatorial and parallel chemistry techniques. There is a continuous trend to move away from the synthesis of individual compounds toward the synthesis of compound libraries, whose members are accessible through the same chemical reaction using different chemical building blocks.

To offer a competitive advantage, structure-based design tools must now be fast enough to prioritize thousands of compounds per day. Several algorithms have been developed that allow for de novo design[16,17] or for flexible docking[18] of hundreds to thousands of small molecules into a protein binding site per day on a single CPU computer. Essential components of all these structure-based design software tools are scoring functions that translate computationally determined protein–ligand interactions into approximate

estimations of binding affinity. These scoring functions guide the conformational and orientational search of a ligand within the binding site and ultimately provide a relative ranking of putative ligands with respect to a target. The purpose of this chapter is to describe some of these functions, discuss their strengths and weaknesses, explain how they are used in practical applications, and present selected results to highlight the current status of the field.

The Process of Virtual Screening

In this section, we discuss a general strategy of virtual screening based on the 3D structure of a target. Typically, the following steps are typically taken.

1. Analysis of the 3D protein structure.
2. Selection of one or more key interactions that need to be satisfied by all candidate molecules.
3. Computational search (by docking and/or pharmacophore queries) in chemical databases for compounds that fit into the binding site and satisfy key interactions.
4. Analysis of the retrieved hits and removal of undesirable compounds.
5. Synthesis or purchase of the selected compounds.
6. Biological testing.

The first step is usually a careful analysis of available 3D protein structures. If possible, highly homologous structures will also be analyzed, either to generate additional ideas about possible ligand structural motifs or to gain some insight on how to achieve selectivity relative to other proteins of the same class. A superposition of different protein–ligand complexes can provide some indication about key interactions that are repeatedly found in tight binding protein–ligand complexes. Such an overlay will also highlight flexible parts of the protein. Programs like GRID[19] or LUDI[20,21] are frequently used to visualize potential interaction sites (hot spots) in the binding site of the protein. If there are conserved water molecules in the binding site mediating hydrogen bonds between the protein and the ligand, and if these water molecules cannot be replaced, then including them in the docking process can dramatically improve the hit rate.[13–15]

An important result from the aforementioned 3D structure analysis is usually the identification of one or more key interactions that all ligands should satisfy. An example of such a binding hypothesis is that aspartic protease inhibitors should form at least one hydrogen bond to the catalytic Asp side chains. Although it could be left to the computational algorithm using a good scoring function to pick molecules, experience indicates that the percentage of active compounds in a designed library can be significantly increased if a good binding hypothesis is used as filter. In addition, part of a known ligand may be used as a starting scaffold, and virtual screening techniques can then be used to select side chains.

Once a reasonable binding hypothesis has been generated, the next step is the actual virtual screening. Whether one uses databases of commercially available compounds or ''virtual'' libraries of hypothetical chemical structures, it makes sense to dock not just any compound, but only those that pass a number of simple property filters. Such filters remove

1. Compounds with reactive functional groups. Reactive groups such as $-SO_2Cl$ and $-CHO$ cause problems in some biological assays due to nonspecific covalent binding to the protein.
2. Compounds with a molecular weight below 150 or above 500. Very small molecules like benzene are known to bind to proteins in a rather nonspecific manner and at several sites. Very large molecules (like polypeptides) are difficult to optimize subsequently because bioavailability is usually low for compounds with a molecular weight above 500.
3. Compounds that are not ''drug-like'' according to criteria that have been derived from sets of known drugs.[22,23]

Each remaining compound is then docked into the binding site and scored. The docking process is the most demanding step computationally and is usually carried out on multiprocessor computers. Depending on the docking algorithm and the scoring function, this step may easily take several days of CPU time. The result is a list of several hundred to a few thousand docked small molecule structures each with a computed score, which is further analyzed to weed out undesirable structures. Selection criteria could be

1. Lipophilicity, if not addressed before. Highly lipophilic molecules are difficult to test because of their low solubility in water.
2. Structural class. If 50% of the docked structures belong to a single chemical class, it is probably unnecessary to test all of them.
3. Improbability of docked binding mode. Fast docking tools cannot produce reasonable solutions for all compounds. Often even some high-scoring compounds are found to be docked to the outer surface of the protein. Computational filters help to detect such situations.

Finally, the selected compounds are purchased or synthesized and then tested. If the goal is to identify weakly binding small molecules, it is important to ensure that the biological assay is sensitive and robust enough to pick up these molecules. Measurements using 100–1000 μM concentration of the ligand frequently cause problems due to the limited solubility of the ligands in water. To compensate for this, the assay is often carried out in the presence of 1–5% dimethyl sulfoxide (DMSO) (see, e.g., Ref. 14).

Note that the process of virtual screening still involves manual interventions at various stages. In principle, the whole process can be carried out in a fully automated manner, but in practice visual inspection and manual selection are still very useful.

Major Contributions to Protein–Ligand Interactions

The selective binding of a low molecular weight ligand to a specific protein is determined by the structural and energetic recognition of those two molecules. For ligands of pharmaceutical interest, the protein–ligand interactions are usually noncovalent in nature. The binding affinity can be determined from the experimentally measured binding constant K_i

$$\Delta G = -RT \ln K_i = \Delta H - T\Delta S \qquad [1]$$

The experimentally determined binding constants K_i are typically in the range of 10^{-2} to 10^{-12} mol/L, corresponding to a Gibbs free energy of binding ΔG between -10 and -70 kJ/mol in aqueous solution.[6,24]

There exists a growing body of experimental data on 3D structures of protein–ligand complexes and binding affinities.[25] These data indicate that several features can be found in almost all complexes of tightly bound ligands. These features include

1. A high steric complementarity between the protein and the ligand. This observation is consistent with the long established lock-and-key paradigm.
2. A high complementarity of the surface properties. Lipophilic parts of the ligands are most frequently found to be in contact with lipophilic parts of the protein. Polar groups are usually paired with suitable polar protein groups to form hydrogen bonds or ionic interactions.
3. The ligand usually adopts an energetically favorable conformation.

Generally speaking, direct interactions between the protein and the ligand are essential for binding. The most important types of direct interactions are depicted in Figure 1.

Structural data on unfavorable protein–ligand interactions are sparse. The scarcity of such complexes is due, in part, to the fact that structures of weakly binding ligands are more difficult to obtain and they are usually considered less interesting by many drug discovery chemists and structural biologists. However, weak binding data are vital for the development of scoring functions. What data are available indicate that unpaired buried polar groups at the protein–ligand interface are strongly adverse to binding. For example, few buried CO and NH groups in folded proteins fail to form hydrogen bonds.[26] Therefore, in the ligand design process, one has to ensure that polar functional groups, either of the protein or the ligand, will find suitable counterparts if they become buried upon ligand binding. Another situation that can lead to diminished binding affinity is imperfect steric fitting, which leads to holes at the protein–ligand interface.

The enthalpic and the entropic component of the binding affinity can be determined experimentally, for example, by isothermal titration calorimetry
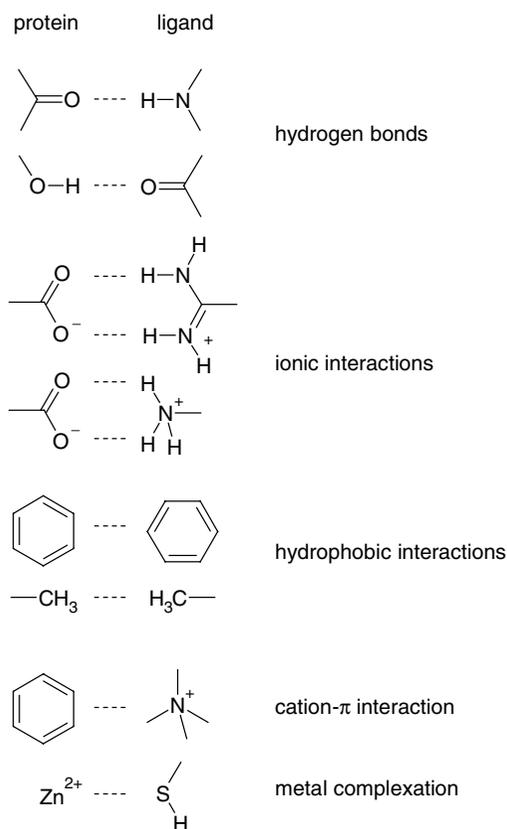
Figure 1 Typical interactions found in protein–ligand complexes. Usually, the lipophilic part of the ligand is in contact with the lipophilic parts of the protein (side chains of Ile, Val, Leu, Phe, Trp, perpendicular contact to amide bonds). In addition, hydrogen bonds are often involved. Some interactions can be charge assisted. Cation–π interactions and metal complexation can also play a significant role in individual cases.

(ITC). Unfortunately, these data are still sparse and are difficult to interpret.[27] Existing thermodynamic data indicate that there is always a substantial compensation between enthalpic and entropic contributions.[28–30] The data also show that the binding may be enthalpy-driven (e.g., streptatividin–biotin, $\Delta G = -76.5 \, \text{kJ/mol}, \Delta H = -134 \, \text{kJ/mol}$) or entropy-driven (e.g., streptavidin–2-(4′-hydroxy-azobenzene)benzoic acid (HABA), $\Delta G = -22.0 \, \text{kJ/mol}, \Delta H = +7.1 \, \text{kJ/mol}$).[31] Data from protein mutants yield estimates of $5 \pm 2.5 \, \text{kJ/mol}$ for the contribution from individual hydrogen bonds to the binding affinity.[32–34] Similar values have been obtained for the contribution of an intramolecular hydrogen bond to protein stability.[35–37] The consistency of experimental values derived from different proteins suggests some degree of additivity in the hydrogen-bonding interactions.

The contribution of hydrogen bonds to the binding affinity strongly depends on solvation and desolvation effects. Here lies the biggest challenge
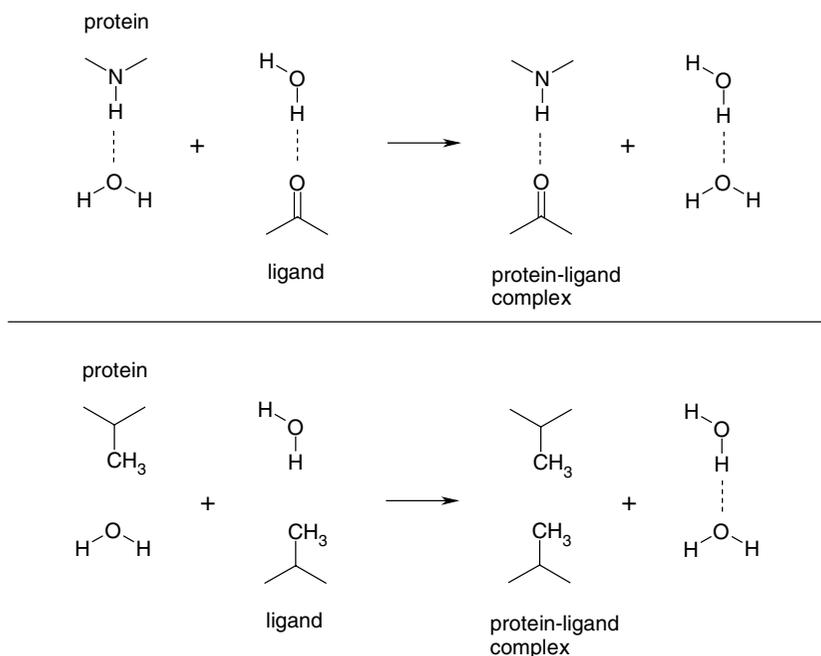
Figure 2 Role of water molecules in hydrogen bonds (upper part) and lipophilic interactions (lower part). In the unbound state (left side), the polar groups of the ligand and the protein form hydrogen bonds to water molecules. These water molecules are replaced upon complex formation. The hydrogen-bond inventory (total number of hydrogen bonds) does not change. In contrast, the formation of lipophilic contact increases the total number of hydrogen bonds due to the release of water molecules from the unfavorable lipophilic environment.

in the quantitative treatment of protein–ligand interactions: providing an accurate description of the role of water molecules (Figure 2). It has been shown, by comparing the binding affinities of ligand pairs differing by just one hydrogen bond, that the existence of an individual hydrogen bond can even be adverse to binding.[38] Charge-assisted hydrogen bonds are stronger than neutral ones, but this enhancement in binding is paid for by higher desolvation penalties. The electrostatic interaction of an exposed salt bridge is worth as much as a neutral hydrogen bond (5 ± 1 kJ/mol according to Ref. 39), and the same interaction in the interior of a protein can be significantly larger.[40]

The experimental determination of ΔH and ΔS sometimes yields surprising results, as, for example, in the thermodynamics of hydrogen-bond formation in the complex of FK506 or rapamycin with FK506-binding protein (FKBP).[34] Binding to the wild-type and to the mutant Tyr 82 $\rightarrow$ Phe 82 was

studied. From X-ray studies, it was known that the side chain hydroxyl of Tyr 82 forms a hydrogen bond with the ligand. If Tyr 82 is replaced by Phe, then one hydrogen bond is lost. As expected, the ligand-binding affinity was slightly reduced. The free enthalpy difference is 4 ± 1.5 kJ/mol. Somewhat unexpectedly, however, this destabilization is due to an entropy loss. In other words, the formation of this particular hydrogen bond is enthalpically unfavorable but entropically favorable. The entropy gain appears to be mainly due to the replacement of two water molecules by the ligand.[41]

Lipophilic interactions are essentially contacts between apolar parts of the protein and the ligand. The generally accepted view is that lipophilic interactions are mainly the result of the replacement and release of ordered water molecules and thus are entropy-driven processes.[42,43] The entropy gain is due to the fact that the water molecules are no longer positionally confined. There are also enthalpic contributions to lipophilic interactions. Water molecules occupying lipophilic binding sites are unable to form hydrogen bonds with the protein. If they are released, they can form strong hydrogen bonds with bulk water. It has been shown in many cases that the lipophilic contribution to the binding affinity is proportional to the lipophilic surface area buried from the solvent and typically has values in the range of 80–200 J/(mol Å^2).[44–46]

Conformational flexibility is another factor influencing the binding affinity. Usually, a ligand binds in a single conformation and therefore loses much of its conformational flexibility upon binding. Greater binding affinities have been observed for cyclic derivatives of ligands that otherwise adopt the same binding mode.[47,48] The entropic cost of freezing a single rotatable bond has been estimated to be 1.6–3.6 kJ/mol at 300 K.[49,50] Recent estimates derived from nuclear magnetic resonance (NMR) shift titrations of open-chain dications and dianions are much lower (0.5 kJ/mol),[51] but in those systems the conformational restriction may not have been as high as in a protein-binding site. The entropic cost of the external (translational and orientational) degrees of freedom has been estimated to be around 10 kJ/mol.[52,53]

In spite of many inconsistencies and difficulties in interpretation, most of the experimental data suggests that simple additive models for the protein–ligand interactions might be a good starting point for the development of empirical scoring functions. Indeed, the first published scoring functions were actually built based on experimental work that was published by about 1992, including studies on thermolysin[54] and vancomycin.[50,55]

Figure 3 summarizes some of the interactions that play a role in receptor–ligand binding. Binding involves a complex equilibrium between ensembles of solvated species. In the next section, we will discuss various approaches that are used to capture essential elements of this equilibrium in computationally efficient scoring functions. The discussion focuses on general approaches rather than individual functions. The reader is referred to Table 1 for original references to the most important scoring functions.[56–114]
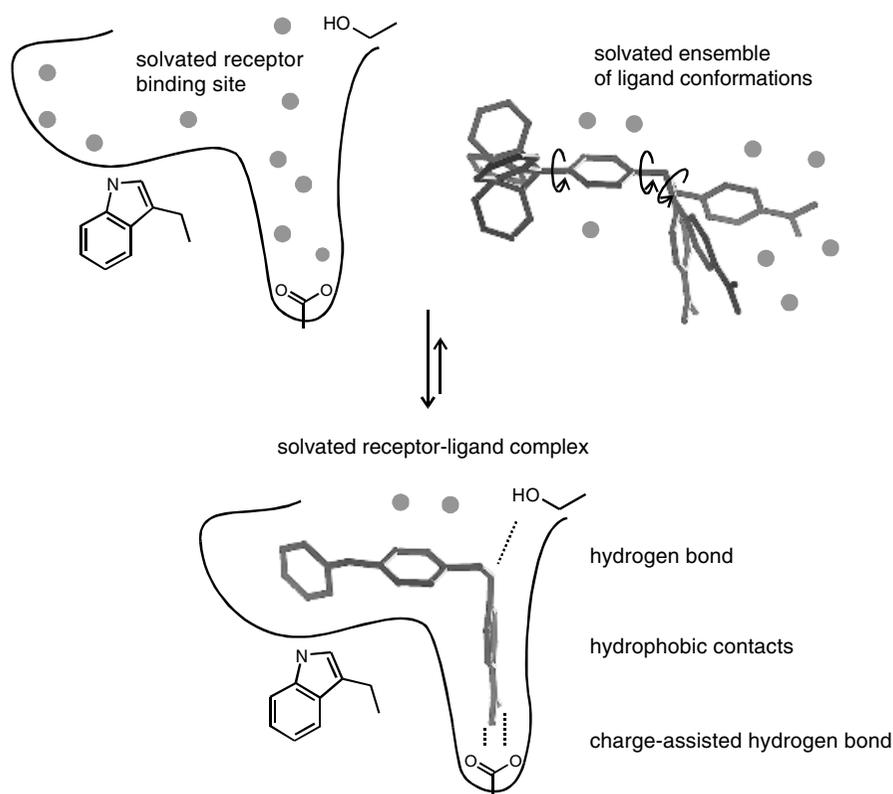
Figure 3 Overview of the receptor–ligand binding process. All species involved are solvated by water (symbolized by gray spheres). The binding free energy difference between the bound and unbound state is a sum of enthalpic components (breaking and formation of hydrogen bonds, formation of specific hydrophobic contacts), and entropic components (release of water from hydrophobic surfaces to solvent, loss of conformational mobility of receptor and ligand).

DESCRIPTION OF SCORING FUNCTIONS FOR RECEPTOR–LIGAND INTERACTIONS

A rigorous theoretical treatment of reversible receptor–ligand binding requires full consideration of all species involved in the binding equilibrium. In the unbound state, both the ligand and the receptor are separately solvated and do not interact. In the bound state, both partners are partially desolvated and form interactions with each other. Since it is the free energy of binding one is interested in determining, the energies of the solvated receptor, the solvated ligand, and the solvated complex should be calculated as ensemble averages.

Table 1 Reference List for the Most Important Published Scoring Functions[a]

Type of Function	Name	Year Published	Original References	Selected References to Applications
Empirical	SCORE1 (the LUDI scoring function)	1989–1993	115–117	GRID[19]
		1994	56	LUDI[20,21], e.g. 57,58
	GOLD score	1994	59	FLOG[59]
	PLP	1995	60, 61	GOLD[60,61]
	FlexX score	1995, 2000	62, 63	
	VALIDATE	1996	64	FlexX[64–67]
		1996	68	
	ChemScore	1996	69	Hammerhead[70]
	SCORE2	1997	71	72
		1998	73	
	SCORE	1998	74	
	Fresno	1998	75	
	ScreenScore	1999	76	77
		2001	78	
Desolvation terms	HINT	1991	79	80
		1997	81	
Knowledge-based	SMOG	1996	82	SMOG[83]
	BLEEP	1999	84, 85	
	PMF	1999	86	87–91
	DrugScore	2000	92	93
Force field	AMBER	1992	94	DOCK[94–99]
	CHARMM	1998	100	
Force field + desolvation		1998	101	
		1999	102	
	AMBER + desolvation	1999	103	
	CHARMM + PB	1999	104	
	AMBER + desolvation	1999	105	
	MM PB/SA	1999	106	107, 108
Linear response simplified free energy perturbation	OWFEG[113] Grid	1994	109	110–112
		2001	114	

[a]Where force fields are used, the entry in "Original Reference" column refers to the use as a scoring function, not to the force field itself.

The appropriate statistical mechanics treatment has been reviewed else-where[118] and is not the topic of this chapter. Large-scale Monte Carlo or molecular dynamics (MD) simulations are necessary to derive reasonably accurate values of binding free energies. These computational methods are only suitable for small sets of compounds, since they require large amounts of computational resources. Moreover, even the most advanced techniques are reliable only for calculating binding free energy differences between closely related ligands.[119–122] However, a number of less rigorous but faster scoring schemes have been developed that should be amenable to larger numbers of ligands. For example, recent experience has shown that continuum solvation models can replace explicit solvent molecules, at least in the final energy evaluation of the simulation trajectory.[123] Another less expensive alternative for computing binding free energies is the use of linear reponse theory[109,110] in conjunction with a surface term.[112]

Scoring functions that can be evaluated quickly enough to be practical in docking and virtual screening applications are very crude approximations to the free energy of binding. They usually take into account only one receptor–ligand complex structure and disregard ensemble averaging and properties of the unbound state of the binding partners. Furthermore, all scoring methods have in common the fact that the free energy is obtained from a sum of terms. In a strict physical sense, this is not possible, since the free energy of binding is a state function, but its components are not.[124] Furthermore, simple additive models cannot describe subtle cooperativity effects.[125] Nevertheless, it is often useful to interpret receptor–ligand binding in an additive fashion,[126–128] and estimates of binding free energy are available in this way at very low computational cost. Fast scoring functions can be categorized into three main classes: (1) force field-based methods, (2) empirical scoring functions, and (3) knowledge-based methods. Each of these is now discussed.

Force Field-Based Methods

An obvious idea to circumvent parameterization efforts for scoring is to use nonbonded energies of existing, well-established molecular mechanics force fields for the estimation of binding affinity. In doing so, one substitutes estimates of the free energy of binding in solution by an estimate of the gas-phase enthalpy of binding. Even this crude approximation can lead to satisfying results. A good correlation was obtained between nonbonded interaction energies calculated with a modified MM2 force field and IC_{50} values of 33 inhibitors of human immunodeficiency virus (HIV)-1 protease.[129] Similar results were reported in a study of 32 thrombin–inhibitor complexes with the CHARMM force field.[130] In both studies, however, experimental data represented rather narrow activity ranges and little structural variation.

The AMBER[131,132] and CHARMM[133] nonbonded terms are used as a scoring function in several docking programs. Protein terms are usually

precalculated on a cubic grid, such that for each ligand atom only the interactions with the closest grid points have to be evaluated.[94] This leads to an increase in speed of about two orders of magnitude compared to traditional atom-by-atom evaluation. Distance-dependent dielectric constants are usually employed to approximate the long-range shielding of electrostatic interactions by water.[100] However, compounds with high formal charges still obtain unreasonably high scores due to overestimated ionic interactions. For this reason, it has been common practice in virtual screening to separate databases of compounds into subgroups according to their total charge and then to rank these groups separately.

When electrostatic interactions are complemented by a solvation term calculated by the Poisson–Boltzmann equation[134] or faster continuum solvation models (as in Ref. 135), the deleterious effects of high formal charges are diminished. In a validation study on three protein targets, Shoichet et al.[103] observed a significantly improved ranking of known inhibitors after correction for ligand solvation. The current version of the docking program DOCK calculates solvation corrections based on the generalized Born[136] solvation model.[105] The method has been validated in a study where several peptide libraries were docked into various serine protease active sites.[137]

In the context of scoring, the van der Waals term of force fields is mainly responsible for penalizing docking solutions with steric overlap between receptor and ligand atoms. The term is often omitted when only the binding of experimentally determined complex structures is analyzed.[102,138,139]

A recent addition to the list of force field-based scoring methods has been developed by Charifson and Pearlman. Their so-called OWFEG (one window free energy grid) method[114] is an approximation to the expensive first-principles method of free energy perturbation (FEP).[140] For the purpose of scoring, an MD simulation is carried out with the ligand-free, solvated receptor site. The energetic effects of probe atoms on a regular grid are collected and averaged during the simulation. Three simulations are run with three different probes: a neutral methyl-like atom, a negatively charged atom, and a positively charged atom. The resulting three grids contain information on the score contributions of neutral, positively, and negatively charged ligand atoms located in various positions of the receptor site and can thus be used in a straightforward manner for scoring. The OWFEG approach seems to be successful for K_i prediction as well as for virtual screening applications.[113] Its conceptual advantage is the implicit consideration of entropic and solvent effects and the inclusion of some protein flexibility in the simulations.

The calculation of ligand strain energy traditionally lies in the realm of molecular mechanics force fields. Although effects of strain energy have rarely been determined experimentally,[141] it is generally accepted that high-affinity ligands bind in low-energy conformations.[142,143] If a compound must adopt a strained conformation to fit into a receptor pocket, a less negative binding free energy should result. Strain energy can be estimated by calculating the

difference in conformational energy between the global minimum structure and the current conformation of the ligand in the complex. However, force field estimates of energy differences between individual conformations are not reliable for all systems. In practice, better correlations with experimental binding data are obtained when strain energy is used as a filter to weed out unlikely binding geometries rather than when strain energy is added to the final score. Estimation of ligand strain energy based on force fields can be time consuming, and so alternatives such as empirical rules derived from small-molecule crystal structure data are often employed.[144] Conformations generated by such programs are, however, often not strain free, because only one torsional angle is treated at a time. Some strained conformations can be excluded when two consecutive dihedral angles are simultaneously taken into account, however.[78]

Empirical Scoring Functions

The underlying idea of empirical scoring functions is that the binding free energy of a noncovalent receptor–ligand complex can be interpreted as a sum of localized, chemically intuitive interactions. Such energy decompositions can be a useful tool to understand binding phenomena even without analyzing 3D structures of receptor–ligand complexes. Andrews, Craik, and Martin[126] calculated average functional group contributions to binding free energy from a set of 200 compounds whose affinity to a receptor had been experimentally determined. These average functional group contributions can then be used to estimate a receptor-independent binding energy for a compound that can be compared to experimental values. If the experimental value is approximately the same as or higher than the calculated value, one can infer a good fit between receptor and ligand and essentially all functional groups of the ligand are involved in protein interactions. If the experimental energy is significantly lower, one can infer that the compound can not fully form its potential interactions with the protein. Experimental binding affinities have also been analyzed on a per atom basis in quest of the maximal binding affinity of noncovalent ligands.[145] It was concluded that for the strongest binding ligands, each nonhydrogen atom on average contributes 1.5 kcal/mol to the total binding energy.

With 3D structures of receptor–ligand complexes at hand, the analysis of binding phenomena can of course be much more detailed. The binding affinity $\Delta G_{\text{binding}}$ can be estimated as a sum of interactions multiplied by weighting coefficients ΔG_i

$$\Delta G_{\text{binding}} \approx \sum_i \Delta G_i f_i(r_{\text{l}}, r_{\text{p}}) \qquad [2]$$

where each f_i is a function of the ligand coordinates r_{l} and the protein receptor coordinates r_{p}, and the sum is over all atoms in the complex. Scoring schemes

that use this concept are called "empirical scoring functions." Several reviews summarize details of individual parameterizations.[17,146–151] The individual terms in empirical scoring functions are usually chosen so as to intuitively cover important contributions of the total binding free energy. Most empirical scoring functions are derived by evaluating the functions f_i for a set of protein–ligand complexes and fitting the coefficients ΔG_i to experimental binding affinities of these complexes by multiple linear regression or by supervised learning techniques. The relative weight of the individual contributions depends on the training set. Usually, between 50 and 100 complexes are used to derive the weighting factors, but in a recent study it was shown that many more than 100 complexes were needed to achieve convergence.[75] The reason for this large number is probably due to the fact that the publicly available protein–ligand complexes fall in a few heavily populated classes of proteins, such that in small sets of complexes few interaction types dominate.

Empirical scoring functions usually contain individual terms for hydrogen bonds, ionic interactions, hydrophobic interactions, and for binding entropy. Hydrogen bonds are typically scored by simply counting the number of donor–acceptor pairs falling within a given distance and angle range considered to be favorable for hydrogen bonding, weighted by penalty functions for deviations from preset ideal values.[56,71,73] The amount of error-tolerance in these penalty functions is critical to the success of scoring methodology. When large deviations from ideality are tolerated, the scoring function may be unable to discriminate between different orientations of a ligand. Contrarily, small tolerances lead to situations where many structurally similar complex structures result in very similar scores. Attempts have been made to reduce the localized nature of such interaction terms by using continuous modulating functions on an atom-pair basis.[69] Other workers have avoided the use of penalty functions altogether and introduced separate regression coefficients for strong, medium, and weak hydrogen bonds.[75] For example, at Agouron a simple four-parameter potential, which is called the piecewise linear potential (PLP), was developed that is an approximation of a potential well without angular terms.[62] Most empirical scoring functions treat all types of hydrogen-bond interactions equally, but some attempts have been made to distinguish between different donor–acceptor functional group pairs. Hydrogen-bond scoring in the docking program GOLD,[60,61] for example, is based on a list of hydrogen-bond energies for all combinations of 12 defined donor and 6 acceptor atom types derived from ab initio calculations of model systems incorporating those atom types. A similar differentiation of donor and acceptor groups is made in the hydrogen-bond functions in the program GRID,[152] a program commonly used for the characterization of binding sites.[115–117] The inclusion of such lookup tables in scoring functions is presumed to avoid errors originating from the oversimplification of individual interactions.

Reducing the weight of hydrogen bonds formed at the outer surface of the binding site is a useful measure for reducing the number of false positive

hits in virtual screening applications. Reducing the weight can be done by reducing charges of surface residues when explicit electrostatic terms are used[100] or by multiplying the hydrogen-bond by a factor that depends on the accessibility of the protein-bonding partner[91] in empirical scoring functions.

Ionic interactions are treated in a similar manner as hydrogen bonds. Long-distance charge–charge interactions are usually disregarded, and so it is more appropriate to refer to salt bridges or charged hydrogen bonds here. The SCORE1 function by Boehm implemented in LUDI[56] gives greater weight to salt bridges than to neutral hydrogen bonds. This function was found to be useful in scoring several series of thrombin inhibitors.[57,72] But just as with force field scoring functions, this weighting introduces the danger of giving unreasonably high scores to highly charged molecules. Our experience with the docking program FlexX,[64–67] which contains a variant of SCORE1 (the LUDI scoring function), has been that better results are generally obtained when charged and uncharged hydrogen are treated equally in virtual screening applications. This observation is also the case for the ChemScore function by Protherics.[71]

Hydrophobic interaction energies are usually estimated by the size of the contact surface at the receptor–ligand interface. A reasonable correlation between experimental binding energies can often be achieved with a surface term alone (see, e.g., Refs. 24,153,154 and the discussion in the earlier section on Major Contributions to Protein–Ligand Interactions). Various approximations for surface terms have been used, such as grid-based methods[56] and volume-based methods (see especially the discussion in Ref. 101). Many functions employ distance-scaled sums over neighboring receptor–ligand atom pairs. Distance cutoffs for these functions have been chosen to be short[64] or to be longer to include atom pairs that do not form direct van der Waals contacts.[62,71] The assignment of the weighting factor ΔG_i for the hydrophobic term depends strongly on the training set. Its value might have been underestimated in most derivations of empirical scoring functions,[155] because most training sets contain an overly large proportion of ligands containing an excessive number of donor and acceptor groups (many peptide and carbohydrate fragments).

In most existing empirical scoring functions, a number of atom types are defined as being hydrophobic, and all their contributions are treated in the same manner. Alternatively, the propensity of specific atom types to be located in the solvent or in the interior of a protein can be assessed by so-called "atomic solvation parameters" that can be derived from experimental data such as octanol–water partition coefficients[156,157] or from structural data.[81,158] Atomic solvation parameters are used in the VALIDATE scoring function,[68] and they have been tested in DOCK.[80] Entropy terms in empirical scoring functions account for the restriction of conformational degrees of freedom of the ligand upon complex formation. A crude but useful estimate of this entropy contribution is the number of freely rotatable bonds of a ligand.

This simple measure has the advantage of being a function of the ligand only.[56,73] Since it is argued that purely hydrophobic contacts allow more residual motion in the ligand fragments, more elaborate estimates try to take into account the nature of each ligand fragment on either side of a flexible bond and the interactions they form with the receptor.[68,71] Such penalty terms are also robust with respect to the distribution of rotatable bonds in the ligands of the training set, but they offer little or no advantage in the virtual screening of compound databases. The group at Agouron has further used an entropy penalty term that is proportional to the score[159] to account for entropy–enthalpy compensation.[28–30,160]

Knowledge-Based Methods

Empirical scoring functions "see" only those interactions that are part of the model. Many less common interactions are usually disregarded, even though they can be strong and specific, as exemplified, for example, by NH–π hydrogen bonds. It would become a difficult task to generate a comprehensive and consistent description of all these interactions within the framework of empirical scoring functions. But there exists a growing body of structural data on receptor–ligand complexes that can be used to detect favorable binding geometries. "Knowledge-based" scoring functions try to capture the knowledge about receptor–ligand binding hidden in the Protein Data Bank[161] (PDB) by means of statistical analysis of structural data alone—and they do so without referring to inconsistent experimentally determined binding affinities.[162] They have their foundation in the inverse formulation of the Boltzmann law:

$$E_{ij} = -kT \ln(p_{ijk}) + kT \ln(Z) \qquad [3]$$

The energy function E_{ij} is called a potential of mean force for a state defined by three variables i, j, and k; p_{ijk} is the corresponding probability density, and Z is the partition function. The second term of Eq. [3] is constant at constant temperature T and does not need to be considered, because $Z = 1$ can be chosen by definition of a suitable reference state leading to normalized probability densities p_{ijk}. The inverse Boltzmann technique has been applied to derive potentials for protein folding from databases of protein structures.[163] For the purpose of deriving scoring functions, the variables i, j, and k are chosen to be protein atom types, ligand atom types, and their interatomic distance. The frequency of occurrence of individual contacts is presumed to be a measure of their energetic contribution to binding. When a specific contact occurs more frequently than that from a random or average distribution, this indicates an attractive interaction. When it occurs less frequently, it is interpreted as being a repulsive interaction between those two atom types. The frequency

distributions for a data set of interacting molecules can thus be converted to sets of atom-pair potentials that are straightforward to evaluate.

The first applications of knowledge-based scoring functions in drug research[164–166] were restricted to small data sets of HIV protease–inhibitor complexes and did not result in generally applicable scoring functions. Recent publications[82–86,92,93] have shown that useful general scoring functions can be derived with this method. The de novo design program SMoG[82,83] contained the first general-purpose implementation of such a potential.

The "PMF" function by Muegge[86] consists of a set of distance-dependent atom-pair potentials $E_{ij}(r)$ that are written as

$$E_{ij}(r) = -kT \ln[f_j(r)\rho^{ij}(r)/\rho^{ij}] \qquad [4]$$

Here, r is the atom pair distance, and $\rho^{ij}(r)$ is the number density of pairs ij that occur in a given radius range around r. The term ρ^{ij} in the denominator is the average density of receptor atoms j in the whole reference volume. The number density is calculated in the following manner. A maximum search radius is defined. This radius describes a reference sphere around each ligand atom j, in which receptor atoms of type i are searched, and which is divided into shells of a specified thickness. The number of receptor atoms i found in each spherical shell is divided by the volume of the shell and averaged over all occurrences of ligand atoms i in the database of protein–ligand complexes. Muegge argues that the spherical reference volume around each ligand atom needs to be corrected by eliminating the volume of the ligand itself, because ligand–ligand interactions are not regarded. This correction is done by the volume correction factor $f_j(r)$ that is a function of the ligand atom only and gives a rough estimate of the preference of atom j to be solvent exposed rather than buried within the binding pocket. Muegge could show that the volume correction factor contributes significantly to the predictive power of the PMF function.[90] Also, a relatively large reference radius of at least 7–8 Å must be applied to implicitly include solvation effects, particularly the propensity of individual atom types to be located inside a protein cavity or in contact with solvent.[89] For docking calculations, the PMF scoring function is evaluated in a grid-based manner and combined with a repulsive van der Waals potential at short distances and minima extended slightly toward shorter distances.

The DrugScore function created by Gohlke, Hendlich, and Klebe[92] is based on roughly the same formalism, albeit with several differences in the derivation leading to different potential forms. Most notably, the statistical distance distributions $\rho^{ij}(r)/\rho^{ij}$ for the individual atom pairs ij are divided by a common reference state that is simply the average of the distance distributions of all atom pairs $\rho(r) = \sum_i \sum_j \rho^{ij}(r)/i_{max} j_{max}$, where the product in the denominator yields the total number of pair functions. Furthermore, no

volume correction term is used, and the sampling cutoff (the radius of the reference sphere) is set to only 6 Å. The individual potentials have the form

$$E_{ij}(r) = -kT(\ln[\rho^{ij}(r)/\rho^{ij}] - \ln[\rho(r)]) \qquad [5]$$

The pair potentials in Eq. [5] are used in combination with other potentials, depending on one (protein or ligand) atom type only, that express the propensity of an atom type to be buried within a lipophilic protein environment upon complex formation. Contributions of these surface potentials and the pair potentials are weighted equally in the final scoring function. DrugScore was developed with the aim of differentiating between correctly docked ligand structures versus decoy (arbitrarily placed) structures for the same protein–ligand pair.

A different type of reference state was chosen by Mitchell et al.[85] The pair interaction energy is written as

$$E_{ij}(r) = kT \ln[1 + m^{ij}\sigma] - kT \ln[1 + m^{ij}\sigma\rho^{ij}(r)/\rho(r)]$$

Here, the number density $\rho^{ij}(r)$ is defined as in Eq. [4], but it is normalized by the number density of all atom pairs at this same distance instead of by the number of pairs ij in the whole reference volume. The variable m^{ij} is the number of atom pairs ij found in the data set of protein–ligand complexes, and σ is an empirical factor that defines the weight of each observation. This potential is combined with a van der Waals potential as a reference state to compensate for the lack of sampling at short distances and for certain underrepresented atom pairs. Apart from data on 90 protein–ligand complexes used in the original validation, no further application has been published.

CRITICAL ASSESSMENT OF CURRENT SCORING FUNCTIONS

Influence of the Training Data

All fast scoring functions share a number of deficiencies that one should be aware of for any application. First, most scoring functions are in some way fitted to or derived from experimental data. The resulting functions necessarily reflect the accuracy of the data that were used in their derivation. A general problem with empirical scoring functions is the fact that the experimental binding energies are compiled from many different sources and therefore form inconsistent data sets containing systematic experimental errors. Scoring functions not only reflect the quality, but also the type of experimental data on which they are based. Most scoring functions are still derived from data on mostly high-affinity receptor–ligand complexes. Moreover, many of these

structures are peptidic in nature, whereas interesting lead molecules in pharmaceutical research are usually nonpeptidic. This influence of peptides is reflected in the relatively high contributions of hydrogen bonds in the total score. The balance between hydrogen bonding and hydrophobic interactions is a critical issue in scoring, and its consequences are especially obvious in virtual screening applications, as will be illustrated in the later section on Hydrogen Bonding versus Hydrophobic Interactions.

Molecular Size

The simple additive nature of most fast scoring functions often leads to large molecules being assigned high scores. Although it is true that small molecules with a molecular weight below 200–250 are rarely of very high affinity, there is no guarantee that larger compounds are more active. When it comes to comparing scores of two compounds of different size, it therefore makes sense to include a penalty term that diminishes the dependence of the score on molecular size. In some applications, a constant penalty value has been added to the score for each heavy atom.[167] Alternatively, a penalty term proportional to the molecular weight has been used.[168] The scoring function of the docking program FLOG, which contains force field and empirical terms, has been normalized to remove the linear dependence of the crude score on the number of ligand atoms that was found in a docking study of a 7500 compound database.[59] Entropy terms designed to estimate the restriction of conformational mobility upon ligand binding also help to eliminate overly large and flexible molecules, although they were originally introduced to improve the correlation between experimental and calculated affinities.[56,71] The size of the solvent-accessible surface of the ligand within the protein-binding pocket is also a useful penalty term because it helps avoid excessively large ligands that cannot fit completely into the binding site. Note, however, that all these approaches are very pragmatic in nature and do not solve the problem of size dependence, which is closely linked to the understanding of cooperativity effects.[125]

Other Penalty Terms

Scoring functions in general reward certain favorable interactions such as hydrogen bonds, but rarely penalize unfavorable interactions. Since scoring functions are derived from experimentally determined crystal structures, "unnatural" and energetically unfavorable orientations of a ligand within the receptor cavity are rarely observed and therefore cannot be accounted for by the scoring function. Knowledge-based scoring functions try to capture such effects indirectly by making those interactions repulsive that are not observed in crystal structures. It seems, however, that the statistical difference

between what is not observed and what is to be expected on average is often not significant enough to form reliable repulsive interactions. Furthermore, the neglect of angular terms in the derivation of knowledge-based scoring functions leads to average pair potentials that cannot discriminate well enough between different binding geometries.

In the derivation of regression-based empirical scoring schemes, on the other hand, penalty terms have traditionally not been included. However, some situations like obvious electrostatic and steric clashes can be avoided by guessing reasonable penalty terms or by importing them from molecular mechanics force fields. An example of this is the ''chemical scoring'' function available in the docking program DOCK.[94–99] This function is a modified van der Waals potential made to be attractive or repulsive between particular groups of donor, acceptor, and lipophilic receptor atoms and ligand atoms.[169,170] Other unacceptable binding orientations cannot be avoided by simple clash terms, but instead require a more refined analysis of binding geometry. Among the causes for poor results are an imperfect steric fit of the ligand within the cavity, an unnaturally high degree of solvent-accessible ligand surface in the complex or the formation of voids at the receptor–ligand interface. Possible remedies are empirical filters that measure such fit parameters and remove docking solutions above a user-specified threshold.[171] A promising approach along these lines is the inclusion of artificially generated, erroneous, decoy solutions in the optimization of scoring functions. In the process of deriving weights for individual terms of the scoring function, the decoy solutions should always obtain lower ranks than the correct solutions, and thus suitable penalty terms could be derived automatically. Such a procedure was first reported for the scoring function of a flexible ligand superposition algorithm.[172,173]

Specific Attractive Interactions

Another general deficiency of scoring functions stems from the simplified description of attractive interactions. Molecular recognition is not based only on classical hydrogen bonds and hydrophobic contacts. Many researchers, especially those active in host–guest chemistry, are making use of other specific types of interactions. For example, hydrogen bonds that are formed between acidic protons and π systems.[174] These bonds can substitute for conventional hydrogen bonds in both strength and specificity, as has been noted, for example, in protein–DNA recognition[175] and as can be observed in serine protease complexes deposited in the PDB.[161] Another class of ''unconventional'' interactions is the cation–π interaction, which is especially important at the surface of proteins.[176,177] Current empirical scoring functions do not model these interactions and mostly disregard the directionality of, for example, interactions between aromatic rings.[178,179] In the derivation of empirical scoring functions, one thus implicitly attributes some of the binding energy arising

from these interactions to conventional interaction terms, which may be one more reason why conventional hydrogen-bond contributions have traditionally been overestimated. One could imagine adding terms to empirical scoring functions that are omitted in the calibration of the functions, but adjusted empirically to reward especially good fits, in a way analogous to penalty terms. Knowledge-based methods would also allow one to incorporate these interactions in a scoring function, again provided that directionality is taken into account, which is not the case in current approaches.

Water Structure and Protonation State

Uncertainties about protonation states and water structure at the receptor–ligand interface also make scoring difficult. These effects play a role in the derivation as well as in the application of scoring functions. The entropic and energetic contributions of water reorganization upon ligand binding are very difficult to predict (see, e.g., Ref. 180). The only reasonable approach for addressing this problem is to concentrate on conserved water molecules and make them part of the receptor. For example, the docking program FLOG has been applied to the search of inhibitors for a metallo-β-lactamase[13] within the Merck in-house database. Docking was performed with three different configurations of bound water in the active site. The top-scoring compounds showed an enrichment in biphenyl tetrazoles, several of which were found to be active at a concentration below 20 μM. A crystal structure of one tetrazole ($IC_{50} = 1.9$ μM) not only confirmed the predicted binding mode of one of the inhibitors, but also displayed the water configuration that had—retrospectively—been the most predictive one of the three models.

Scoring functions rely on a fixed assignment of a general atom type to each protein and ligand atom. This also implies a fixed assignment of protonation state for each acidic and basic functional group. Even though these assignments can be reliable enough for conditions in aqueous solution, significant pK_a shifts can be witnessed upon ligand binding.[181] This phenomenon can arise from local changes of dielectric conditions inside the binding pocket. The change of a donor to an acceptor functionality due to modified protonation states has important consequences for scoring.[137] Accordingly, improved docking and scoring algorithms will eventually need to have a more detailed and flexible description of protonation states.

Performance in Structure Prediction

The multitude of different solutions that have been used for receptor–ligand scoring calls for an objective assessment that could help future users to decide which function to use under a given set of circumstances. To do this, one must differentiate between predicting protein–ligand complex structures

(i.e., the scoring function is used as the objective function in docking), rank ordering a set of ligands with respect to the same protein (K_i prediction), and the use of scoring functions to discover weakly binding compounds from a large database of mostly nonbinders (virtual screening). Note that the latter two tasks are indeed very different. In virtual screening, the focus is on elimination of nonbinders, whereas the correct rank order of weakly, medium, and strongly binding molecules is of secondary interest.

Even when the criteria are clear, a comprehensive assessment of scoring functions is difficult because very few functions have been tested on the same data sets. For example, studies where each scoring function is used in conjunction with two different docking algorithms (e.g., Ref. 170) are not meaningful in this context, because each docking algorithm produces different sets of solution structures. For structure prediction, several studies have shown that knowledge-based scoring functions are at least as good as empirical functions. They are somewhat "softer" than empirical functions,[162] meaning that small root-mean-square deviations from the crystal structure usually do not lead to huge changes in score, a fact that can mainly be attributed to the neglect of directionality. The PMF function has been successfully applied to structure prediction of inhibitors of neuraminidase[88] and stromelysin 1 (matrix metalloprotease-3; MMP-3)[182] in the program DOCK, yielding superior results to the DOCK force field and chemical scoring options. The DrugScore function was tested on a large set of PDB complexes and gave significantly better results than the standard FlexX scoring function with FlexX as the docking engine. DrugScore performed as well as the force field score in DOCK, but outperformed chemical scoring. Grueneberg, Wendt, and Klebe[15] used the Drug-Score function in a virtual screening study to find novel carbonic anhydrase inhibitors (see the section on Application of Scoring Functions in Virtual Screening later in this chapter). Two of the virtual hits that turned out to be highly active compounds were then examined crystallographically. The docking solution predicted by DrugScore was closer to the experimental structure than that predicted by the FlexX score.

Although the objective function (the function whose global minimum is searched during docking) is used for both structure generation and energy evaluation in many docking programs, better results can often be obtained if different functions are used. More specifically, the docking objective function can be adapted to the docking algorithm used. In a parameter study, Vieth et al.[100] found that by using a soft-core van der Waals potential their MD-based docking algorithm became more efficient. Using FlexX as the docking engine, we observed that when directed interactions (mostly hydrogen bonds) are emphasized in the docking phase, library ranking can be done successfully with the more simple, undirected PLP potential (see the prior section on Empirical Scoring Functions) that emphasizes the general steric fit of receptor and ligand. Results are significantly worse when PLP is used for both docking and energy evaluation.

Rank Ordering Sets of Related Ligands

For structure prediction, structures of protein–ligand complexes from the PDB can serve as a common pool to test scoring functions. It is more difficult to draw valid conclusions about the relative performance of scoring functions to rank order sets of ligands with respect to their binding to the same target. First, there are few published studies in which different functions have been applied to the same data sets. Second, experimental data are often not measured under the same conditions but collected from various literature references. The latter practice can have especially dramatic effects when inhibitory concentrations for 50% reduction of a biological effect (IC_{50} data) are used instead of K_i values.

On average, empirical scoring functions seem to lead to better correlations between experimental and calculated binding energies than do force field based approaches because the nonbonded interactions in the latter are usually not optimized to reproduce individual intermolecular binding phenomena. However, the only available calculated data for most published functions are those for the complexes used in the derivation of the functions themselves. Very promising results of rank ordering have also been obtained with the knowledge-based functions DrugScore[93] and PMF.[86,88,182]

The task of rank ordering small (ca. 10–100) sets of related ligands with respect to a target can also be accomplished with methods that are computationally more demanding than simple scoring functions. The most generally applicable methods are probably force field scores augmented with electrostatic desolvation and surface area terms. An example is the MM–PBSA method that combines Poisson–Boltzmann electrostatics with AMBER MD calculations.[183] This method has been applied to an increasing number of studies, and it has led to promising results.[106–108,184] Poisson–Boltzmann calculations have been performed on a variety of targets with many related computational protocols.[102,138,139,185–188] Alternatively, extended linear response protocols[112] can be used. The OWFEG grid method by Pearlman has also shown promising results.[114]

APPLICATION OF SCORING FUNCTIONS IN VIRTUAL SCREENING

In recent years, virtual screening of large databases has emerged as the central application of scoring functions. In the following sections, we describe special requirements that scoring functions must fulfill for successful virtual screening, and we indicate the level of accuracy that can nowadays be expected from virtual screening.

As discussed in the introductory sections, the goal of virtual screening is to use computational tools together with the known 3D structure of the target to select a subset of compounds from chemical libraries for synthesis and

biological testing. This subset typically consists of ca. 100–2000 compounds selected from libraries containing 100,000–500,000 compounds. Therefore, it is essential that the computational process including the scoring function is fast enough to handle several thousand compounds in a short period of time. Consequently, only the fastest scoring functions are currently used for this purpose. Speed is especially important for those scoring functions used as objective functions during the docking calculations, since they are evaluated several hundred to a thousand or so times during the docking process of a single compound.[18]

Following a successful virtual screening run, the selected subset of compounds contains a significantly enhanced number of active compounds as compared to a random selection. A key parameter to measure the performance of docking and scoring methods is the so-called "enrichment factor." It is simply the ratio of active compounds in the subset selected by docking divided by the number of active compounds in a randomly chosen subset of equal size. In practice, enrichment factors are far from the ideal case, where all active compounds are placed on the top ranks of a prioritized list. Insufficiencies of current scoring functions, as discussed in the previous section, are partly responsible for moderate enrichment rates. Another major reason is the fact that the receptor is still treated as a rigid object in the computational protocols being used. To generate correct binding modes of different molecules, it is necessary to predict induced fit phenomena. Unfortunately, predicting protein flexibility remains extremely difficult and computationally expensive.[189–196]

Seeding Experiments

Enrichment factors can be calculated only when experimental data are available for the full library. But only a few libraries containing experimental data that have been measured under uniform conditions for all members are available to the public. Several authors have therefore tested the predictive ability of docking and scoring tools by compiling an arbitrarily selected set of diverse, drug-like compounds and then adding to it a number of known active compounds. This "seeded" library is then subjected to the virtual screen, and, for the purpose of evaluation, it is assumed that the added active compounds are the only true actives in the library. Several such experiments have been published. An example is a study performed at Merck with the docking program FLOG.[59] A library consisting of 10,000 compounds including inhibitors of various types of proteases and HIV protease was docked into the active site of HIV protease. This resulted in excellent enrichment of the HIV protease inhibitors: all inhibitors but one were among the top 500 library members. However, inhibitors of other proteases were also considerably enriched.[197]

Seeding experiments allow for comparisons of different scoring functions with respect to their performance for different targets. Seeding experiments

also teach how to recognize typical failure cases. Recent examples of library ranking experiments include those by Charifson et al.,[198] Bissantz, Folkers, and Rognan,[77] and Stahl and Rarey.[78] Charifson and co-workers compiled sets of several hundred active molecules for three different targets: p38 MAP kinase, inosine monophosphate dehydrogenase, and HIV protease. The members of these sets were then docked into the corresponding active sites together with 10,000 randomly chosen, but drug-like, commercial compounds using DOCK[98] and the Vertex in-house docking tool Gambler. Three scoring functions performed consistently well in enriching active compounds, namely, ChemScore,[71,199] the DOCK AMBER force field score, and PLP.[62] The finding that these three scoring functions performed so well was partially attributed to the fact that a rigid-body optimization could be carried out with these functions, because the functions include repulsive terms in contrast to many of the other tested functions. The study by Stahl and Rarey[78] compared the performance of DrugScore[92] and PMF[86] to that of PLP[62] and FlexX score using the docking program FlexX.[64–66] Interestingly, the two knowledge-based scoring functions showed significantly different behavior for extreme cases of active sites. DrugScore coped well with situations where ligands are tightly bound in narrow lipophilic cavities (e.g., COX-2 and the thrombin S1 pocket), whereas PMF did not lead to good enrichment in such cases. Conversely, for the very polar binding site of neuraminidase, PMF gave better enrichment than any other scoring function, whereas DrugScore failed. The description of complexes in which many hydrogen bonds play a role seems to be a general strength of PMF. This has also been noted by Bissantz, Folkers and Rognan,[77] who found PMF to perform well for the polar target thymidine kinase and less well for the estrogen receptor.

Hydrogen Bonding versus Hydrophobic Interactions

It is of central importance in virtual screening to achieve a balanced description of hydrogen bonding and hydrophobic contributions to the score in order to avoid a bias toward either highly polar or completely hydrophobic molecules. Empirical scoring functions have the advantage that they can be quickly reparameterized to achieve such a balance, whereas such an adjustment is impossible with knowledge-based functions. Because this is such an important topic, we will illuminate it with a number of examples.

Consider the following database ranking experiment. A database of about 7600 compounds was flexibly docked into the ATP binding site of p38 MAP kinase. The database consisted of ca. 7500 random compounds from the World Drug Index (WDI)[200] and 72 inhibitors of p38 MAP kinase, which in turn consisted of 30 inhibitors forming two hydrogen bonds with the receptor and 20 inhibitors forming only one. Both groups covered the same activity range from low micromolar (μM) to about 10 nM. For each of the docked compounds, up to 800 alternative docking solutions were
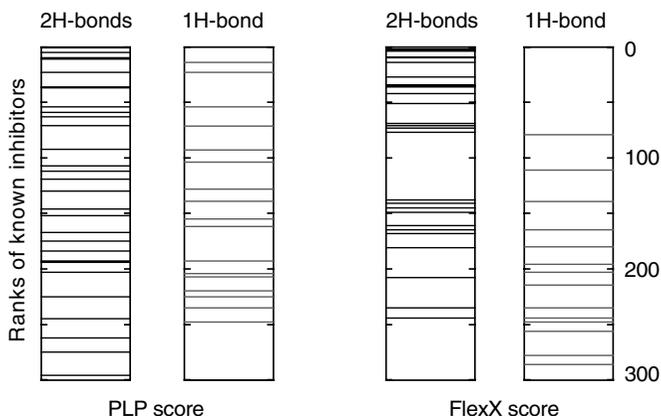
Figure 4 Results of a seeding experiment. The ranks of known p38 MAP kinase inhibitors are shown as horizontal lines in the four diagrams. Inhibitors have been divided into two classes: those forming one or two hydrogen bonds to the p38 MAP kinase ATP binding site. The FlexX scoring function preferentially enriches those inhibitors that form two hydrogen bonds. This tendency is less pronounced for the PLP scoring function. The inhibitors with the best predicted affinities are at the top. Data is shown for the top 300 compounds in terms of docking scores.

generated by FlexX[64–66] using the FlexX scoring function. These alternative solutions were rescored separately by the FlexX and PLP[62] scoring functions to select the lowest energy docking solution per compound. The compounds in the database were then ranked according to these scores. Figure 4 shows the ranks of the known inhibitors among the top 350 compounds as calculated by both scoring functions. Although the overall performance of both scoring functions in enriching inhibitors is comparable, it is obvious that the FlexX score "specializes" on the doubly hydrogen-bonded inhibitors. On the other hand, if one were to select screening candidates from the PLP list, one would most likely select both types of inhibitors.

The PLP function generally emphasizes steric complementarity and hydrophobic interactions with its more far-reaching pair potential, whereas the FlexX score emphasizes hydrogen-bond complementarity. A combination of PLP and FlexX scoring functions called ScreenScore was published recently.[78] It was derived by performing a systematic optimization of library ranking results over seven targets, whose receptor sites cover a wide range of form, size, and polarity. ScreenScore was designed to be a robust and general scoring function that combines the virtues of both PLP and FlexX. Figure 5 shows that this is indeed the case. ScreenScore gives good enrichment values for cyclooxygenase-2 (COX-2 has a highly lipophilic binding site), and neuraminidase (which has a highly polar site), whereas the individual functions fail in one of the two cases. The authors of PLP have recently enhanced their
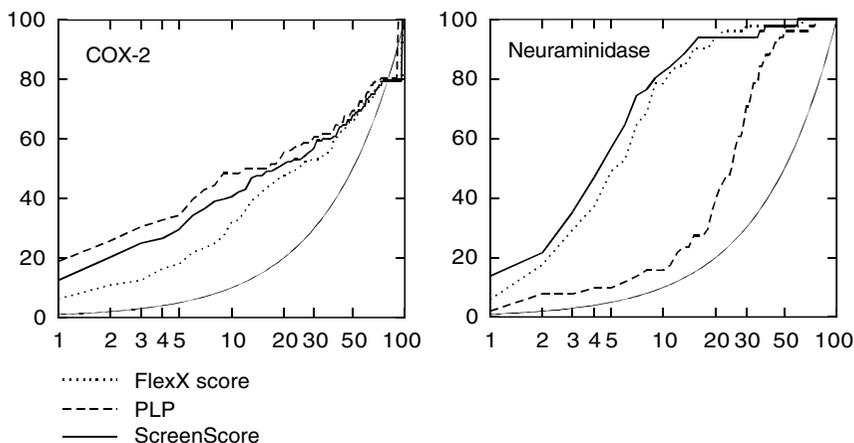
Figure 5 Results of seeding experiments on two targets with three different scoring functions. In both graphs, the accumulated percentages of active compounds are plotted against the percentage of the total ranked library. The smooth exponential curve in each graph corresponds to the hypothetical case of no enrichment and assumes a random ordering of the database.

scoring function by including directed hydrogen bonding terms,[63] which may lead to a similarly robust scoring function as ScreenScore.

Another example where the balance between H-bonding and hydrophobic contributions is important involves the performance of the knowledge-based scoring function DrugScore.[92] The estrogen receptor binding site is a large lipophilic cavity with acceptor groups at either end that can form hydrogen bonds with ligand hydroxyl groups as present in the agonists **1** and **2** or the antagonists **3** and **4** (Figure 6). The narrow binding pocket and relatively rigid nature of the ligands restrains possible binding modes significantly. Accordingly, it can be assumed that FlexX is capable of generating reasonable solutions likely to be in agreement with experiment. Therefore we can expect the present example to represent a valuable test for scoring functions. For both agonists and antagonists, lipophilic interactions largely determine the binding energy. The majority of antagonists, however, differ from the agonists in an additional side chain bearing a tertiary amino group. This difference is reflected in the bound structures of the receptor. In the agonist-bound state the binding pocket is not accessible to solvent, whereas in the antagonist-bound state it opens up and allows the positively charged antagonist side chain to form a salt bridge with the carboxylate group of Glu 351. Agonists should bind equally well to both forms of the receptor. A 7500 compound subset from the World Drug Index (WDI) and a library of 20 agonists and 16 antagonists were docked into both agonist (PDB code 1ere) and antagonist (PDB code 1err) forms of the receptor. FlexX scores obtained from both structures are plotted against each other in Figure 6(**a**). Due to the large contribution of
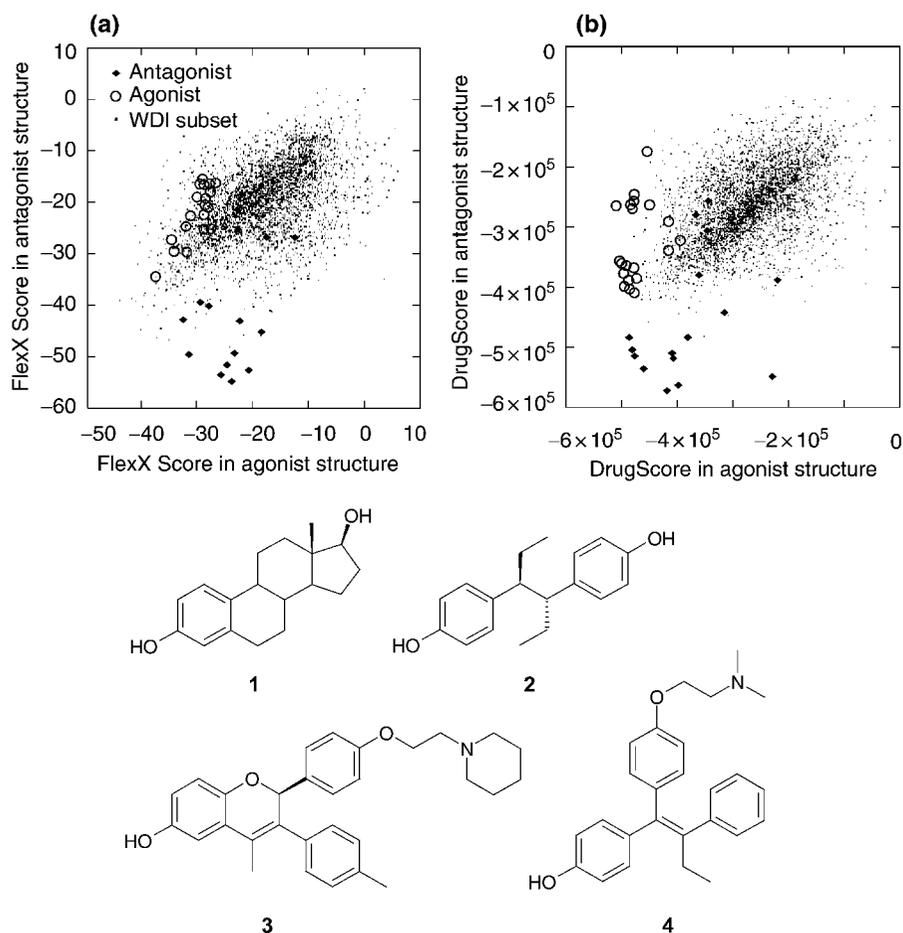
Figure 6 Docking estrogen receptor agonists and antagonists into two crystal structures of the estrogen receptor, the agonist-bound conformation and the antagonist-bound conformation. Scores for both docking results are plotted against each other. Compounds **1** and **2** are examples of agonists, compounds **3** and **4** are typical antagonists.

the surface-exposed salt bridge formed with Glu 351 to the total score the antagonists are clearly separated from the WDI compounds, whereas the agonists are ranked among the bulk of the WDI entries in the antagonist structure. In the agonist form, the formation of a salt bridge is not possible, resulting in a lower average score for all molecules. Almost the same result as with the FlexX score is obtained with the ChemScore function[71,199] by Protherics.[201] The new DrugScore function[92] performs better in this situation. Using this scoring function, results shown in Figure 6(**b**) are obtained. Not only are the agonists significantly better separated from the WDI subset

when docked into the agonist structure, but more importantly, about half of the agonists are also among the 10% top ranked molecules in the database when docked into the open, antagonist structure, where they have to compete with many structures forming salt bridges.

Finding Weak Inhibitors

Seeding experiments are often carried out with a handful of active compounds that have already been optimized for binding to a given target. Enrichment factors achieved in this way are often misleading, because finding potent inhibitors from among a number of random molecules is significantly easier than distinguishing weakly binding inhibitors from nonbinders. In practice, virtual screening will find, at best, inhibitors in the low micromolar range, simply because no chemical database will be large enough, diverse enough, and lucky enough to find optimized leads right away.

The difficulties associated with weak binders are illustrated in Figure 7 with thrombin as a target. The 7500 compound subset of the WDI mentioned above was docked into the thrombin active site together with three sets of 100
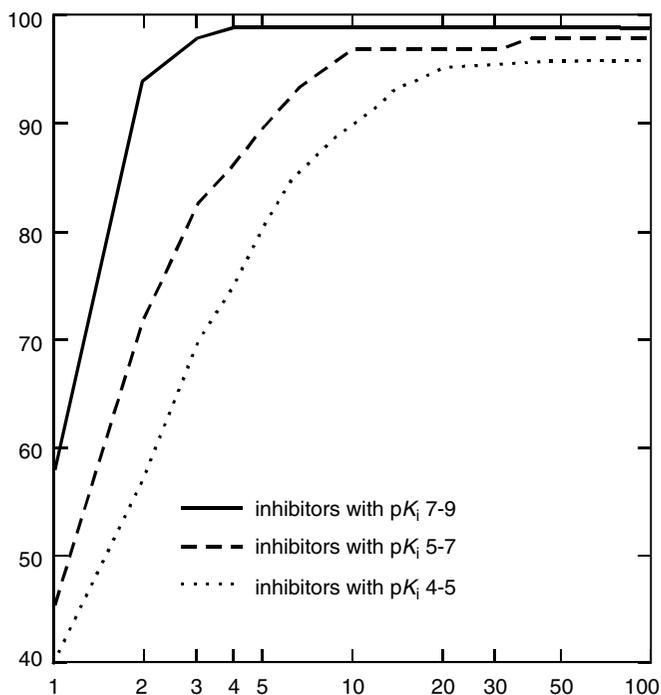
Figure 7 Enrichment of three sets of 100 thrombin inhibitors that cover different ranges of activity. Less active compounds are more difficult to enrich.
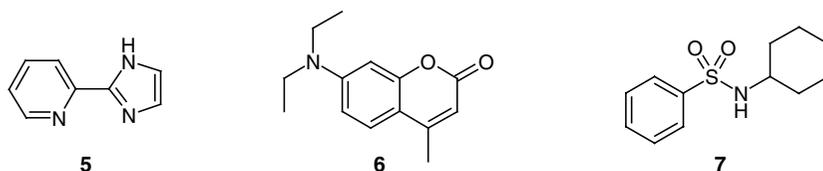
Figure 8 Weak binders to FKBP.

known inhibitors in different activity ranges. It can be clearly seen in Figure 7 that enrichment decreases as the binding affinity of the active compounds decreases. Note that thrombin is a relatively easy target for most virtual screening methods (at least to identify compounds with charged moieties binding to the S1 subsite), and thus the separation of actives and inactives is still good for the low micromolar inhibitors. According to the authors' experience, the situation is worse for many other targets.

Nevertheless, library ranking can successfully be applied to enrich even very weak ligands. A database of approximately 4000 commercially available compounds was screened against FKBP by means of the SAR-by-NMR technique[202] and was found to contain 31 compounds with activities below 2 mM. Three examples of these compounds are shown in Figure 8. Compounds **5**, **6**, and **7** have measured dissociation constants of 0.1, 0.13 and 0.5 mM, respectively. This set of structures was flexibly docked into the FKBP binding site using DOCK 4.0 in conjunction with the PMF scoring function.[87] For the top 20% of the ranked database, enrichment factors between 2 and 3 were achieved. Enrichment factors were twice as large as those obtained with the standard AMBER score implemented in DOCK.

Consensus Scoring

Different scoring schemes emphasize different physical phenomena that are important for ligand binding. Differences between scoring schemes might not be obvious in the calculation of binding affinities for known active compounds, but they can be very pronounced in the assessment of nonbinding molecules. The computational group at Vertex has reported good experience with a concept called "consensus scoring," whereby libraries of molecules are docked and assessed with several scoring functions and only those molecules are retained that score well with the majority of those functions. This can lead to a significant decrease in false positives,[198] but invariably a number of true positives is also lost in the process (see, e.g., Ref. 77).

One should keep in mind that in consensus scoring the number of false positives can be reduced, but one runs the risk of eliminating a number of active compounds that only one of the scoring functions has ranked high. Consider the example of the p38 MAP kinase inhibitors in Figure 4: consensus
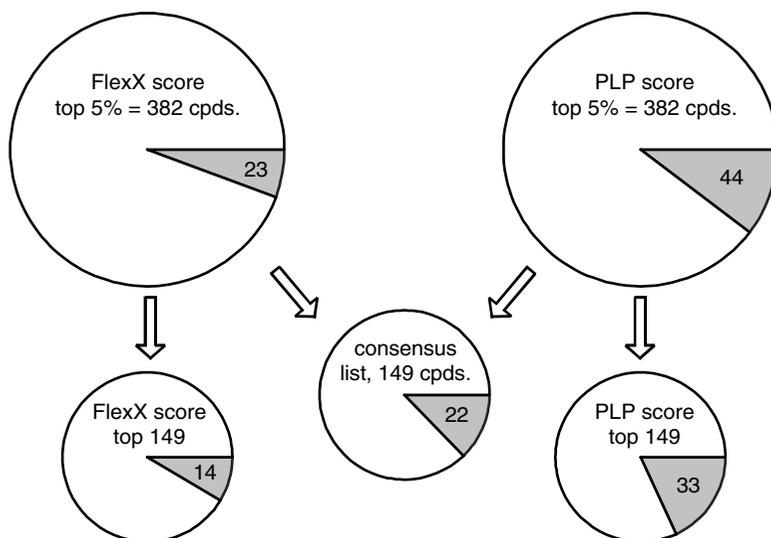
Figure 9 Analysis of the consensus scoring concept with COX-2 as an example. Numbers in the shaded areas are numbers of active compounds. The larger pie charts at the top show the numbers of inhibitors in the top 5% of the database in terms of scores. The smaller pie charts refer to fewer top ranking compounds for better comparison with the smaller size of the consensus list.

scoring for the top 100 compounds by means of PLP and FlexX scores would eliminate all but one of the singly hydrogen-bonded inhibitors.

Figure 9 shows a worked consensus scoring example for a virtual screening experiment on COX-2. (Figure 5 shows the corresponding FlexX and PLP enrichment curves.) There are 23 inhibitors in the top 5% of the FlexX score rank list and roughly twice as many in the PLP rank list. Consensus scoring retains 22 of the actives. Because many inactive compounds are filtered out, the ratio of actives to false positives increases relative to either of the original lists. A different picture is obtained when one regards only the top 149 compounds from the individual FlexX and PLP rank lists—the same number of compounds that are in the consensus list. It becomes clear that the PLP function alone performs significantly better than does consensus scoring.

Thus, if one does not know in advance which scoring function will work better, more robust results can be obtained with consensus scoring. If one has a rough idea which function works better, one can decrease the number of false positives more effectively by testing fewer compounds from the top of a single rank list. Experience from seeding experiments with known inhibitors or an analysis of the type of binding site of the target can help to identify a suitable scoring function.

Successful Identification of Novel Leads through Virtual Screening

It has been shown that virtual screening is an efficient way of finding novel leads. The program DOCK, one of the most widely used docking programs, has been applied in many published studies.[101,158,161,163,258,302,316] Usually the DOCK AMBER force field score has been applied. Other docking tools such as GREEN[203] also use the AMBER force field as a scoring function, and a successful screening application has been published.[204] The docking program SANDOCK[205] uses an empirical scoring function that evaluates steric complementarity, hydrophobic contacts, and hydrogen bonding. SANDOCK has been used to find a variety of novel FKBP inhibitors.[12]

Docking routines in the program packages DOCK and ICM[206] have been used to identify novel nuclear hormone receptor antagonists[207] and, for an RNA target, the transactivation response element (TAR) of HIV-1.[208] In both studies, the virtual screening protocol started with 153,000 compounds from the Available Chemicals Directory (ACD),[209] and the researchers employed increasingly elaborate docking and scoring schemes for smaller groups of selected compounds. In the HIV-1 TAR study, the ACD library was first rigidly docked into the binding site with DOCK. Only a simple contact scoring scheme was used in this step. The 20% best-scoring compounds were then subjected to flexible docking with ICM in combination with an empirical scoring function derived specifically for RNA targets, leading to a set of about 5000 compounds. Two more steps of longer sampling for the conformational analysis of these remaining compounds within the binding site led to 350 selected candidates. Two of the compounds that were experimentally tested significantly reduced the binding of the Tat protein to HIV-1 TAR.

A study by Grueneberg, Wendt, and Klebe[15] resulted in subnanomolar inhibitors of carbonic anhydrase II (CAII). The study is a textbook example of virtual screening focusing on successively smaller subsets of the initial database in several steps and employing different methods at each step. Carbonic anhydrase II is a metalloenzyme that catalyzes the reversible hydration of CO_2 to HCO_3^-.[210] In the human eye, an isoform of the enzyme is involved in water removal. Inhibitors of CAII can thus be used to reduce intraocular pressure in the treatment of glaucoma. The CAII binding site is a rather rigid, funnel-shaped binding pocket. Known inhibitors such as dorzolamide 8 (Figure 10; see also Ref. 8) bind to the catalytic zinc ion via a sulfonamide group.

An initial database of 90,000 entries in the Maybridge[211] and Lead-Quest[212] libraries was converted to 3D structures with the 3D structure generation program Corina.[213,214] In a first filtering step, all compounds were passed through a UNITY[215] pharmacophore query. The pharmacophore query was constructed from an analysis of available X-ray structures of the enzyme and incorporated donor, acceptor, and hydrophobic features of the binding site. Compounds passing this filter also had to contain a known zinc-binding
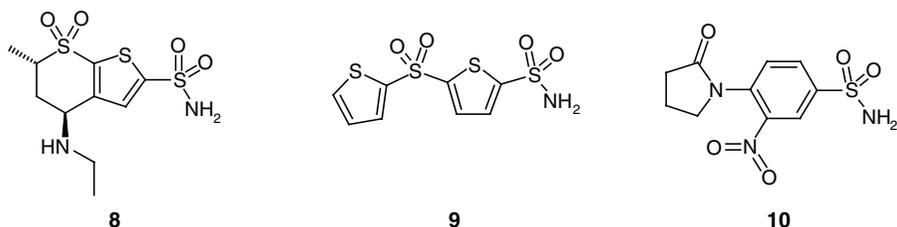
Figure 10 Inhibitors of carbonic anhydrase II. Compounds **9** and **10** are subnanomolar inhibitors identified through virtual screening. Compound **8** is the marketed drug dorzolamide.

group. A set of 3314 compounds passed these requirements. In the second filtering step, the known CAII inhibitor dorzolamide **8** was used as a template onto which all potential candidates were flexibly superimposed by means of the program FlexS.[172] The top-ranking compounds from this step were then docked into the binding site with FlexX[64–66] taking into account four conserved water molecules in the active site. The top-ranking 13 hits were chosen for experimental testing. Nine of these compounds showed activities below 1 µM, and the sulfonamides **9** and **10** (Figure 10) have K_i values below 1 nM.

The de novo design of inhibitors of the bacterial enzyme DNA gyrase is another example for a successful application of structure-based virtual screening.[14] DNA gyrase is a well-established antibacterial target.[216] It is an essential, prokaryotic type II topoisomerase with no mammalian counterpart involved in the vital processes of DNA replication, transcription, and recombination. DNA gyrase catalyzes the ATP-dependent introduction of negative supercoils into bacterial DNA as well as the decatenation and unknotting of DNA. The enzyme consists of two subunits A and B with the active enzyme being an A_2B_2 complex. Subunit A of DNA gyrase is involved in DNA breakage and reunion, whereas the B subunits catalyze the hydrolysis of ATP. Quinolones (e.g., the now famous ciprofloxacin), which inhibit DNA gyrase by binding to the subunit A, are successfully used as broad-spectrum antibacterial agents in the clinic. Unfortunately, resistance to quinolones emerged some time ago. The two other classes of DNA gyrase inhibitors, cyclothialidines and coumarins (e.g., novobiocin), bind to the ATP binding site of subunit B. Novobiocin was clinically used against *Staphylococcus aureus*, but it suffers from toxicity effects and resistance against it is developing rapidly. As demonstrated by the cyclothialidines, this type of resistance can be overcome. Unfortunately, the cyclothialidines have insufficient in vivo activities due to a class specific rapid and extensive glucuronidation of the essential phenol moiety.

To overcome the limitations of known DNA gyrase inhibitors, a new drug discovery project was inititated at Roche. Searching for novel inhibitors by screening the Roche compound library provided no suitable lead structures.

Therefore, a new rational approach was developed to generate lead structures by using the detailed 3D structural information of the ATP binding site located on subunit B. At the time of project initiation, the X-ray structures of the DNA gyrase subunit B complexed with a nonhydrolyzable ATP analogue, with novobiocin, and with cyclothialidine were available. In the inner part of the pocket they all share a common binding motif: each donates a hydrogen-bond to an aspartic acid side chain (Asp 73) and accepts a hydrogen bond from a conserved water molecule. It was reasoned that a novel inhibitor should have the ability to form these two key hydrogen bonds and a lipophilic part to pick up some lipophilic interactions with the enzyme.

A computational search of the ACD[209] and the Roche Compound Inventory, employing the SCORE1 function also implemented in LUDI, was carried out to identify molecules with a low molecular weight meeting the above criteria. Relying on the results of the in silico screening, just 600 compounds were tested initially. Then, analogues similar to the first hits were assayed. Overall, assay results for 3000 compounds gave rise to 150 hits clustered into 14 chemical classes. Seven of those classes could be validated as true, novel DNA gyrase inhibitors that act by binding to the ATP binding site located on the B subunit. The maximum noneffective concentration (MNEC) was in the 5–64 µg/mL range, that is, two to three orders of magnitude higher than the MNEC of novobiocin or cyclothialidine. Subsequent structure-based optimization of the hits led to compounds with potencies equal or up to 10 times better than novobiocin. Compound **11** (Figure 11; MNEC < 0.03 µg/mL) is an example of a novel potent inhibitor of DNA gyrase B resulting from structure-based virtual screening.

An important factor contributing to the success of the project was a new assay that allowed detecting not only highly potent inhibitors but also weak ones, so as to allow testing compounds at high concentrations. Instead of a supercoiling assay usually used to test DNA gyrase inhibitory activity, a coupled spectrophotometric ATPase assay was employed. Compounds could be assayed in concentrations up to 0.5 mM due to a higher tolerance of the solubilizing agent DMSO in this assay.
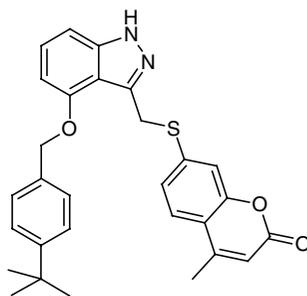
Figure 11 An inhibitor of DNA gyrase B, discovered at Roche by means of virtual screening and subsequent structure-based optimization.

11

OUTLOOK

The first scoring functions were published about 10 years ago. Since then, much experience has been gained in their application and in assessing their accuracy. Significant progress in the development of better functions has been made over the last few years, and it appears as if there now exist scoring functions that can be applied to a wide range of different proteins and which consistently yield considerable enrichment of active compounds. Consequently, many large and small pharmaceutical companies are increasingly using virtual screening techniques to identify possible leads.

In fact, structure-based ligand design is now seen as a very important approach to drug discovery that nicely complements HTS.[217] High throughout screening has a number of serious disadvantages: it is expensive,[218] and it leads to many false positives and few real leads.[22,219] Furthermore, not all biactivity tests are amenable to HTS techniques. And finally, despite the large size of the chemical libraries available to the pharmaceutical industry, it is far from possible to cover the whole universe of drug-like organic molecules. Because of these limitations, and given the current aggressive patenting strategies, the focused design of novel compounds and compound libraries will continue to gain importance.

Thus, there is every reason to believe that the value of structure-based approaches will continue to grow and become even more embraced by the pharmaceutical, agricultural, and related industries than it now is. The development of improved scoring functions is certainly vital for their success.

The major challenges to be overcome in the further development of scoring functions include

1. Polar interactions are still not treated adequately. It is somewhat strange to find that while the role of hydrogen bonds in biology has been well known for a long time and hydrogen bonds are qualitatively well understood, a quantitative treatment of hydrogen bonds in protein–ligand interactions is still missing. Therefore, hydrogen bonds have been referred to as "the last mystery in structure-based design."[38]
2. All scoring functions are essentially simple analytical functions fitted to experimental binding data. Presently, there exists a heavy bias in the public domain data toward peptidic ligands, which in turn leads to an overestimation of polar interactions in many scoring functions. The development of better scoring function clearly requires access to more data on nonpeptidic, low molecular weight, drug-like ligands.
3. Unfavorable interactions and unlikely docking solutions are not penalized strongly enough. General and robust methods that account for undesired features of complex structures in the derivation of scoring functions are still lacking.

4. So far, fast scoring functions only cover part of the whole receptor–ligand binding process. A more detailed picture could be obtained by taking into account properties of the unbound ligand, that is, solvation effects and energetic differences between the low-energy solution conformations and the bound conformation.

ACKNOWLEDGMENTS

REFERENCES

1. J. Greer, J. W. Erickson, and J. J. Baldwin, *J. Med. Chem.,* **37,** 1035 (1994). Application of the Three-Dimensional Structures of Protein Target Molecules in Structure-Based Drug Design.

2. M. von Itzstein, W.-Y. Wu, G. B. Kok, M. S. Pegg, J. C. Dyason, B. Jin, T. V. Phan, M. L. Smythe, H. F. White, S. W. Oliver, P. M. Colmant, J. N. Varghese, D. M. Ryan, J. M. Woods, R. C. Bethell, V. J. Hotham, J. M. Cameron, and C. R. Penn, *Nature (London),* **363,** 418 (1993). Rational Design of Potent Sialidase-Based Inhibitors of Influenza Virus Replication.

3. W. Lew, X. Chen, and U. Choung, *Curr. Med. Chem.,* **7,** 663 (2000). Discovery and Development of GS 4104 (Oseltamivir): An Orally Active Influenza Neuraminidase Inhibitor.

4. S. W. Kaldor, V. J. Kalish, J. F. Davies, V. S. Bhasker, J. E. Fritz, K. Appelt, J. A. Burgess, K. M. Campanale, N. Y. Chirgadze, D. K. Clawson, B. A. Dressman, S. D. Hatch, D. A. Khalil, M. B. Kosa, P. P. Lubbehusen, M. A. Muesing, A. K. Patick, S. H. Reich, K. S. Su, and J. H. Tatlock, *J. Med. Chem.,* **40,** 3979 (1997). Viracept (Nelfinavir Mesylate, AG1343): A Potent, Orally Bioavailable Inhibitor of HIV-1 Protease.

5. R. S. Bohacek, C. McMartin, and W. C. Guida, *Med. Res. Rev.,* **16,** 3 (1996). The Art and Practice of Structure-Based Drug Design: A Molecular Modeling Perspective.

6. R. E. Babine and S. L. Bender, *Chem. Rev.,* **97,** 1359 (1997). Molecular Recognition of Protein–Ligand Complexes: Applications to Drug Design.

7. R. E. Hubbard, *Curr. Opin. Biotechnology,* **8,** 696 (1997). Can Drugs Be Designed?

8. D. B. Boyd, in *Rational Molecular Design in Drug Research,* Proceedings of the *Alfred Benzon Symposium No. 42* (Copenhagen, June 8–12, 1997), T. Liljefors, F. S. Jørgensen, and P. Krogsgaard-Larsen, Eds., Munksgaard, Copenhagen, 1998, pp. 15–23. Progress in Rational Design of Therapeutically Interesting Compounds.

9. M. A. Murcko, P. R. Caron, and P. S. Charifson, *Annu. Reports Med. Chem.,* **34,** 297 (1999). Structure-Based Drug Design.

10. D. A. Gschwend, W. Sirawaraporn, D. V. Santi, and I. D. Kuntz, *Proteins: Struct., Funct., Genet.,* **29,** 59 (1997). Specificity in Structure-Based Drug Design: Identification of a Novel, Selective Inhibitor of *Pneumocystis carinii* Dihydrofolate Reductase.

11. E. K. Kick, D. C. Roe, A. G. Skillman, G. Liu, T. J. A. Ewing, Y. Sun, I. D. Kuntz, and J. A. Ellman, *Chem. Biol.,* **4,** 297 (1997). Structure-Based Design and Combinatorial Chemistry Yield Low Nanomolar Inhibitors of Cathepsin D.

12. P. Burkhard, U. Hommel, M. Sanner, and M. D. Walkinshaw, *J. Mol. Biol.*, **287,** 853 (1999). The Discovery of Steroids and Other Novel FKBP Inhibitors Using a Molecular Docking Program.

13. J. H. Toney, P. M. D. Fitzgerald, N. Grover-Sharma, S. H. Olson, W. J. May, J. G. Sundelof, D. E. Vanderwall, K. A. Cleary, S. K. Grant, J. K. Wu, J. W. Kozarich, D. L. Pompliano, and G. G. Hammond, *Chem. Biol.*, **5,** 185 (1998). Antibiotic Sensitation Using Biphenyl Tetrazoles as Potent Inhibitors of *Bacteroides fragilis* Metallo-β-Lactamase.

14. H.-J. Boehm, M. Boehringer, D. Bur, H. Gmuender, W. Huber, W. Klaus, D. Kostrewa, H. Kuehne, T. Luebbers, N. Meunier-Keller, and F. Mueller, *J. Med. Chem.*, **43,** 2664 (2000). Novel Inhibitors of DNA Gyrase: 3D Structure-Based Needle Screening, Hit Validation by Biophysical Methods, and 3D Guided Optimization. A Promising Alternative to Random Screening.

15. S. Grueneberg, B. Wendt, and G. Klebe, *Angew. Chem. Int. Ed.*, **40,** 389 (2001). Subnanomolar Inhibitors From Computer Screening: A Model Study Using Human Carbonic Anhydrase II.

16. M. A. Murcko, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., Wiley-VCH, New York, 1997, Vol. 11, pp. 1–66. Recent Advances in Ligand Design Methods.

17. D. E. Clark, C. W. Murray, and J. Li, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., Wiley-VCH, New York, 1997, Vol. 11, pp. 67–125. Current Issues in De Novo Molecular Design.

18. I. Muegge and M. Rarey, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., Wiley-VCH, New York, 2001, Vol. 17, pp. 1–60. Small Molecule Docking and Scoring.

19. P. J. Goodford, *J. Med. Chem.*, **28,** 849 (1985). A Computational Procedure for Determining Energetically Favored Binding Sites on Biologically Important Macromolecules.

20. H.-J. Boehm, *J. Comput.-Aided Mol. Design,* **6,** 61 (1992). The Computer Program LUDI: A New Method for the De Novo Design of Enzyme Inhibitors.

21. H.-J. Boehm, *J. Comput.-Aided Mol. Design,* **6,** 593 (1992). LUDI: Rule-Based Automatic Design of New Substituents for Enzyme Inhibitor Leads.

22. C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, *Adv. Drug Delivery Rev.*, **23,** 3 (1997). Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings.

23. D. E. Clark and S. D. Pickett, *Drug Discovery Today,* **5,** 49 (2000). Computational Methods for the Prediction of "Drug-Likeness."

24. H.-J. Boehm and G. Klebe, *Angew. Chem. Int. Ed.*, **35,** 2588 (1996). What Can We Learn from Molecular Recognition in Protein–Ligand Complexes for the Design of New Drugs?

25. O. Roche, R. Kiyama, and C. L. Brooks III, *J. Med. Chem.*, **44,** 3592 (2001). Ligand–Protein Database: Linking Protein–Ligand Complex Structures to Binding Data.

26. I. K. McDonald and J. M. Thornton, *J. Mol. Biol.*, **238,** 777 (1994). Satisfying Hydrogen Bonding Potential in Proteins.

27. M. H. Parker, D. F. Ortwine, P. M. O'Brien, E. A. Lunney, C. A. Banotai, W. T. Mueller, P. McConnell, and C. G. Brouillette, *Bioorg. Med. Chem. Lett.*, **10,** 2427 (2000). Stereoselective Binding of an Enantiomeric Pair of Stromelysin-1 Inhibitors Caused by Conformational Entropy Factors.

28. J. D. Dunitz, *Chem. Biol.*, **2,** 709 (1995). Win Some, Lose Some: Enthalpy–Entropy Compensation in Weak Molecular Interactions.

29. P. Gilli, V. Ferretti, G. Gilli, and P. A. Brea, *J. Phys. Chem.*, **98,** 1515 (1994). Enthalpy–Entropy Compensation in Drug-Receptor Binding.

30. D. H. Williams, D. P. O'Brien, and B. Bardsley, *J. Am. Chem. Soc.*, **123,** 737 (2001). Enthalpy/Entropy Compensation as a Competition Between Dynamics and Bonding: The Relevance to Melting of Crystals and Biological Aggregates.

31. P. C. Weber, J. J. Wendoloski, M. W. Pantoliano, and F. R. Salemme, *J. Am. Chem. Soc.*, **114,** 3197 (1992). Crystallographic and Thermodynamic Comparison of Natural and Synthetic Ligands Bound to Streptavidin.

32. A. R. Fersht, J.-P. Shi, J. Knill-Jones, D. M. Lowe, A. J. Wilkinson, D. M. Blow, P. Brick, P. Carter, M. M. Y. Waye, and G. Winter, *Nature (London),* **314,** 235 (1985). Hydrogen Bonding and Biological Specificity Analysed by Protein Engineering.

33. Y. W. Chen and A. R. Fersht, *J. Mol. Biol.,* **234,** 1158 (1993). Contribution of Buried Hydrogen Bonds to Protein Stability—The Crystal Structure of Two Barnase Mutants.

34. P. R. Connelly, R. A. Aldape, F. J. Bruzzese, S. P. Chambers, M. J. Fitzgibbon, M. A. Fleming, S. Itoh, D. J. Livingston, M. A. Navia, J. A. Thomson, and K. P. Wilson, *Proc. Natl. Acad. Sci. U.S.A.,* **91,** 1964 (1994). Enthalpy of Hydrogen-Bond Formation in a Protein–Ligand Binding Reaction.

35. B. P. Morgan, J. M. Scholtz, M. D. Ballinger, I. D. Zipkin, and P. A. Bartlett, *J. Am. Chem. Soc.,* **113,** 297 (1991). Differential Binding Energy: A Detailed Evaluation of the Influence of Hydrogen Bonding and Hydrophobic Groups on the Inhibition of Thermolysin by Phosphorus-Containing Inhibitors.

36. B. A. Shirley, P. Stanssens, U. Hahn, and C. N. Pace, *Biochemistry,* **31,** 725 (1992). Contribution of Hydrogen Bonding to the Conformational Stability of Ribonuclease T1.

37. U. Obst, D. W. Banner, L. Weber, and F. Diederich, *Chem. Biol.,* **4,** 287 (1997). Molecular Recognition at the Thrombin Active Site: Structure-Based Design and Synthesis of Potent and Selective Thrombin Inhibitors and the X-Ray Crystal Structures of Two Thrombin-Inhibitor Complexes.

38. H. Kubinyi, in *Pharmacokinetic Optimization in Drug Research,* B. Testa, H. van de Waterbeemd, G. Folkers, and R. Guy, Eds., Wiley-VCH, Weinheim, 2001, pp. 513–524. Hydrogen Bonding, the Last Mystery in Drug Design?

39. H.-J. Schneider, T. Schiestel, and P. Zimmermann, *J. Am. Chem. Soc.,* **114,** 7698 (1992). The Incremental Approach to Noncovalent Interactions: Coulomb and van der Waals Effects in Organic Ion Pairs.

40. A. C. Tissot, S. Vuilleumier, and A. R. Fersht, *Biochemistry,* **35,** 6786 (1996). Importance of Two Buried Salt Bridges in the Stability and Folding Pathway of Barnase.

41. J. D. Dunitz, *Science,* **264,** 670 (1994). The Entropic Cost of Bound Water in Crystals and Biomolecules.

42. A. Ben-Naim, *Hydrophobic Interactions,* Plenum Press, New York, 1980.

43. C. Tanford, *The Hydrophobic Effect,* Wiley, New York, 1980.

44. C. Chothia, *Nature (London),* **254,** 304 (1975). Structural Invariants in Protein Folding.

45. F. M. Richards, *Annu. Rev. Biophys. Bioeng.,* **6,** 151 (1977). Areas, Volumes, Packing, and Protein Structure.

46. K. A. Sharp, A. Nicholls, R. Friedman, and B. Honig, *Biochemistry,* **30,** 9686 (1991). Extracting Hydrophobic Free Energies From Experimental Data: Relationship to Protein Folding and Theoretical Models.

47. H. Mack, T. Pfeiffer, W. Hornberger, H.-J. Boehm, and H. W. Hoeffken, *J. Enzyme Inhib.,* **1,** 73 (1995). Design, Synthesis and Biological Activity of Novel Rigid Amidino-Phenylalanine Derivatives as Inhibitors of Thrombin.

48. A. R. Khan, J. C. Parrish, M. E. Fraser, W. W. Smith, P. A. Bartlett, and M. N. G. James, *Biochemistry,* **37,** 16841 (1998). Lowering the Entropic Barrier for Binding Conformationally Flexible Inhibitors to Enzymes.

49. M. S. Searle and D. H. Williams, *J. Am. Chem. Soc.,* **114,** 10690 (1992). The Cost of Conformational Order: Entropy Changes in Molecular Association.

50. M. S. Searle, D. H. Williams, and U. Gerhard, *J. Am. Chem. Soc.,* **114,** 10697 (1992). Partitioning of Free Energy Contributions in the Estimation of Binding Constants: Residual Motions and Consequences for Amide–Amide Hydrogen Bond Strengths.

51. M. A. Hossain and H.-J. Schneider, *Chem. Eur. J.,* **5,** 1284 (1999). Flexibility, Association Constants, and Salt Effects in Organic Ion Pairs: How Single Bonds Affect Molecular Recognition.

52. K. P. Murphy, D. Xie, K. S. Thompson, L. M. Amzel, and E. Freire, *Proteins: Struct., Funct., Genet.,* **18,** 63 (1994). Entropy in Biological Binding Processes: Estimation of Translational Entropy Loss.

53. J. Hermans and L. Wang, *J. Am. Chem. Soc.,* **119,** 2702 (1997). Inclusion of Loss of Translational and Rotational Freedom in Theoretical Estimates of Free Energies of Binding. Application to a Complex of Benzene and Mutant T4 Lysozyme.

54. R. S. Bohacek and C. McMartin, *J. Med. Chem.,* **35,** 1671 (1992). Definition and Display of Steric, Hydrophobic, and Hydrogen-Bonding Properties of Ligand Binding Sites in Proteins Using Lee and Richards Accessible Surface: Validation of a High-Resolution Graphical Tool for Drug Design.

55. D. H. Williams, M. S. Searle, J. P. Mackay, U. Gerhard, and R. A. Maplestone, *Proc. Natl. Acad. Sci. U.S.A.,* **90,** 1172 (1993). Toward an Estimation of Binding Constants in Aqueous Solution: Studies of Associations of Vancomycin Group Antibiotics.

56. H.-J. Boehm, *J. Comput.-Aided Mol. Design,* **8,** 243 (1994). The Development of a Simple Empirical Scoring Function to Estimate the Binding Constant for a Protein–Ligand Complex of Known Three-Dimensional Structure.

57. H.-J. Boehm, D. W. Banner, and L. Weber, *J. Comput.-Aided Mol. Design,* **13,** 51 (1999). Combinatorial Docking and Combinatorial Chemistry: Design of Potent Non-Peptide Thrombin Inhibitors.

58. S.-S. So and M. Karplus, *J. Comput.-Aided Mol. Design,* **13,** 243 (1999). A Comparative Study of Ligand–Receptor Complex Binding Affinity Prediction Methods Based on Glycogen Phosphorylase Inhibitors.

59. M. D. Miller, S. K. Kearsley, D. J. Underwood, and R. P. Sheridan, *J. Comput.-Aided Mol. Design,* **8,** 153 (1994). FLOG: A System to Select "Quasi-Flexible" Ligands Complementary to a Receptor of Known Three-Dimensional Structure.

60. G. Jones, P. Willett, and R. C. Glen, *J. Mol. Biol.,* **245,** 43 (1995). Molecular Recognition of Receptor Sites Using a Genetic Algorithm With a Description of Desolvation.

61. G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor, *J. Mol. Biol.,* **267,** 727 (1997). Development and Validation of a Genetic Algorithm for Flexible Docking.

62. D. K. Gehlhaar, G. M. Verkhivker, P. A. Rejto, C. J. Sherman, D. B. Fogel, L. J. Fogel, and S. T. Freer, *Chem. Biol.,* **2,** 317 (1995). Molecular Recognition of the Inhibitor AG-1343 by HIV-1 Protease: Conformationally Flexible Docking by Evolutionary Programming.

63. G. M. Verkhivker, D. Bouzida, D. K. Gehlhaar, P. A. Reijto, S. Arthurs, A. B. Colson, S. T. Freer, V. Larson, B. A. Luty, T. Marrone, and P. W. Rose, *J. Comput.-Aided Mol. Design,* **14,** 731 (2000). Deciphering Common Failures in Molecular Docking of Ligand–Protein Complexes.

64. M. Rarey, B. Kramer, T. Lengauer, and G. Klebe, *J. Mol. Biol.,* **261,** 470 (1996). A Fast Flexible Docking Method Using an Incremental Construction Algorithm.

65. M. Rarey, B. Kramer, and T. Lengauer, *J. Comput.-Aided Mol. Design,* **11,** 369 (1997). Multiple Automatic Base Selection: Protein–Ligand Docking Based on Incremental Construction Without Manual Intervention.

66. M. Rarey, S. Wefing, and T. Lengauer, *J. Comput.-Aided Mol. Design,* **10,** 41 (1996). Placement of Medium-Sized Molecular Fragments into Active Sites of Proteins.

67. M. Rarey, B. Kramer, and T. Lengauer, *Bioinformatics,* **15,** 243 (1999). Docking of Hydrophobic Ligands With Interaction-Based Matching Algorithms.

68. R. D. Head, M. L. Smythe, T. I. Oprea, C. L. Waller, S. M. Green, and G. R. Marshall, *J. Am. Chem. Soc.,* **118,** 3959 (1996). VALIDATE: A New Method for the Receptor-Based Prediction of Binding Affinities of Novel Ligands.

69. A. N. Jain, *J. Comput.-Aided Mol. Design,* **10,** 427 (1996). Scoring Non-Covalent Protein Ligand Interactions: A Continuous Differentiable Function Tuned to Compute Binding Affinities.

70. W. Welch, J. Ruppert, and A. N. Jain, *Chem. Biol.,* **3,** 449 (1996). Hammerhead: Fast, Fully Automated Docking of Flexible Ligands to Protein Binding Sites.

71. M. D. Elridge, C. W. Murray, T. R. Auton, G. V. Paolini, and R. P. Mee, *J. Comput.-Aided Mol. Design,* **11,** 425 (1997). Empirical Scoring Functions. I. The Development of a Fast Empirical Scoring Function to Estimate the Binding Affinity of Ligands in Receptor Complexes.

72. C. W. Murray, T. R. Auton, and M. D. Elridge, *J. Comput.-Aided Mol. Design,* **12,** 503 (1999). Empirical Scoring Functions. II. The Testing of an Empirical Scoring Function for the Prediction of Ligand-Receptor Binding Affinities and the Use of Bayesian Regression to Improve the Quality of the Method.

73. H.-J. Boehm, *J. Comput.-Aided Mol. Design,* **12,** 309 (1998). Prediction of Binding Constants of Protein Ligands: A Fast Method for the Prioritization of Hits Obtained From De Novo Design or 3D Database Search Programs.

74. Y. Takamatsu and A. Itai, *Proteins: Struct., Funct., Genet.,* **33,** 62 (1998). A New Method for Predicting Binding Free Energy Between Receptor and Ligand.

75. R. Wang, L. Liu, L. Lai, and Y. Tang, *J. Mol. Model.,* **4,** 379 (1998). SCORE: A New Empirical Method for Estimating the Binding Affinity of a Protein–Ligand Complex.

76. D. Rognan, S. L. Lauemoller, A. Holm, S. Buus, and V. Tschinke, *J. Med. Chem.,* **42,** 4650 (1999). Predicting Binding Affinities of Protein Ligands from Three-Dimensional Models: Application to Peptide Binding to Class I Major Histocompatibility Proteins.

77. C. Bissantz, G. Folkers, and D. Rognan, *J. Med. Chem.,* **43,** 4759 (2000). Protein-Based Virtual Screening of Chemical Databases. 1. Evaluation of Different Docking/Scoring Combinations.

78. M. Stahl and M. Rarey, *J. Med. Chem.,* **44,** 1035 (2001). Detailed Analysis of Scoring Functions for Virtual Screening.

79. G. E. Kellogg, G. S. Joshi, and D. J. Abraham, *Med. Chem. Res.,* **1,** 444 (1991). New Tools for Modeling and Understanding Hydrophobicity and Hydrophobic Interactions.

80. E. C. Meng, I. D. Kuntz, D. J. Abraham, and G. E. Kellogg, *J. Comput.-Aided Mol. Design,* **8,** 299 (1994). Evaluating Docked Complexes With the HINT Exponential Function and Empirical Atomic Hydrophobicities.

81. C. Zhang, G. Vasmatzis, J. L. Cornette, and C. DeLisi, *J. Mol. Biol.,* **267,** 707 (1997). Determination of Atomic Solvation Energies from the Structure of Crystallized Proteins.

82. R. S. DeWitte and E. I. Shakhnovich, *J. Am. Chem. Soc.,* **118,** 11733 (1996). SMoG: De Novo Design Method Based on Simple, Fast and Accurate Free Energy Estimates. 1. Methodology and Supporting Evidence.

83. R. S. DeWitte, A. V. Ishchenko, and E. I. Shakhnovich, *J. Am. Chem. Soc.,* **119,** 4608 (1997). SMoG: De Novo Design Method Based on Simple, Fast, and Accurate Free Energy Estimates. 2. Case Studies in Molecular Design.

84. J. B. O. Mitchell, R. A. Laskowski, A. Alex, and J. M. Thornton, *J. Comput. Chem.,* **20,** 1165 (1999). BLEEP—A Potential of Mean Force Describing Protein–Ligand Interactions. I. Generating the Potential.

85. J. B. O. Mitchell, R. A. Laskowski, A. Alex, M. J. Forster, and J. M. Thornton, *J. Comput. Chem.,* **20,** 1177 (1999). BLEEP—A Potential of Mean Force Describing Protein–Ligand Interactions. II. Calculation of Binding Energies and Comparison With Experimental Data.

86. I. Muegge and Y. C. Martin, *J. Med. Chem.,* **42,** 791 (1999). A General and Fast Scoring Function for Protein–Ligand Interactions: A Simplified Potential Approach.

87. I. Muegge, Y. C. Martin, P. J. Hajduk, and S. W. Fesik, *J. Med. Chem.,* **42,** 2498 (1999). Evaluation of PMF Scoring in Docking Weak Ligands to the FK506 Binding Protein.

88. I. Muegge, *Med. Chem. Res.*, **9**, 490 (1999). The Effect of Small Changes in Protein Structure on Predicted Binding Modes of Known Inhibitors of Influenza Virus Neuraminidase: PMF-Scoring in DOCK4.

89. I. Muegge, *Perspect. Drug Discovery Design,* **20**, 99 (2000). A Knowledge-Based Scoring Function for Protein–Ligand Interactions: Probing the Reference State.

90. I. Muegge, *J. Comput. Chem.*, **22**, 418 (2001). Effect of Ligand Volume Correction on PMF Scoring.

91. M. Stahl, *Perspect. Drug Discovery Design,* **20**, 83 (2000). Modifications of the Scoring Function in FlexX for Virtual Screening Applications.

92. H. Gohlke, M. Hendlich, and G. Klebe, *J. Mol. Biol.,* **295**, 337 (2000). Knowledge-Based Scoring Function to Predict Protein–Ligand Interactions.

93. H. Gohlke, M. Hendlich, and G. Klebe, *Perspect. Drug Discovery Design,* **20**, 115 (2000). Predicting Binding Modes, Binding Affinities and "Hot Spots" for Protein-Ligand Complexes Using a Knowledge-Based Scoring Function.

94. E. C. Meng, B. K. Shoichet, and I. D. Kuntz, *J. Comput. Chem.,* **13**, 505 (1992). Automated Docking With Grid-Based Energy Evaluation.

95. I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, and T. Ferrin, *J. Mol. Biol.,* **161**, 269 (1982). A Geometric Approach to Macromolecule–Ligand Interactions.

96. B. K. Shoichet, D. L. Bodian, and I. D. Kuntz, *J. Comput. Chem.,* **13**, 380 (1992). Molecular Docking Using Shape Descriptors.

97. E. C. Meng, D. A. Gschwend, J. M. Blaney, and I. D. Kuntz, *Proteins: Struct., Funct., Genet.,* **17**, 266 (1993). Orientational Sampling and Rigid-Body Minimization in Molecular Docking.

98. T. J. A. Ewing and I. D. Kuntz, *J. Comput. Chem.,* **18**, 1175 (1997). Critical Evaluation of Search Algorithms for Automated Molecular Docking and Database Screening.

99. S. Makino and I. D. Kuntz, *J. Comput. Chem.,* **18**, 1812 (1997). Automated Flexible Ligand Docking Method and Its Application for Database Search.

100. M. Vieth, J. D. Hirst, A. Kolinski, and C. L. Brooks III, *J. Comput. Chem.,* **19**, 1612 (1998). Assessing Energy Functions for Flexible Docking.

101. G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, and A. J. Olson, *J. Comput. Chem.,* **14**, 1639 (1998). Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function.

102. M. Schapira, M. Trotov, and R. Abagyan, *J. Mol. Recognition,* **12**, 177 (1999). Prediction of the Binding Energy for Small Molecules, Peptides and Proteins.

103. B. K. Shoichet, A. R. Leach, and I. D. Kuntz, *Proteins: Struct., Funct., Genet.,* **34**, 4 (1999). Ligand Solvation in Molecular Docking.

104. B. N. Dominy and C. L. Brooks III, *Proteins: Struct., Funct., Genet.,* **36**, 318 (1999). Methodology for Protein–Ligand Binding Studies: Application to a Model for Drug Resistance, the HIV/FIV Protease System.

105. X. Zou, Y. Sun, and I. D. Kuntz, *J. Am. Chem. Soc.,* **121**, 8033 (1999). Inclusion of Solvation in Ligand-Binding Free Energy Calculations Using the Generalized Born Model.

106. I. Massova and P. A. Kollman, *J. Am. Chem. Soc.,* **121**, 8133 (1999). Computational Alanine Scanning to Probe Protein–Protein Interactions: A Novel Approach to Evaluate Binding Free Energies.

107. O. A. T. Donini and P. A. Kollman, *J. Med. Chem.,* **43**, 4180 (2000). Calculation and Prediction of Binding Free Energies for the Matrix Metalloproteinases.

108. B. Kuhn and P. A. Kollman, *J. Am. Chem. Soc.,* **122**, 3909 (2000). A Ligand That is Predicted to Bind Better to Avidin Than Biotin: Insights From Computational Fluorine Scanning.

109. J. Aquist, C. Medina, and J.-E. Samuelsson, *Protein Eng.,* **7**, 385 (1994). A New Method for Predicting Binding Affinity in Computer-Aided Drug Design.

110. T. Hansson, J. Marelius, and J. Aquist, *J. Comput.-Aided Mol. Design,* **12**, 27 (1998). Ligand Binding Affinity Prediction by Linear Interaction Energy Methods.

111. I. D. Wall, A. R. Leach, D. W. Salt, M. G. Ford, and J. W. Essex, *J. Med. Chem.,* **42,** 5142 (1999). Binding Constants of Neuraminidase Inhibitors: An Investigation of the Linear Interaction Energy Method.

112. R. C. Rizzo, J. Tirado-Rives, and W. L. Jorgensen, *J. Med. Chem.,* **44,** 145 (2001). Estimation of Binding Affinities for HEPT and Nevirapine Analogues With HIV-1 Reverse Transcriptase via Monte Carlo Simulations.

113. D. A. Pearlman, *J. Med. Chem.,* **42,** 4313 (1999). Free Energy Grids: A Practical Qualitative Application of Free Energy Perturbation to Ligand Design Using the OWFEG Method.

114. D. A. Pearlman and P. A. Charifson, *J. Med. Chem.,* **44,** 502 (2001). Improved Scoring of Ligand–Protein Interactions Using OWFEG Free Energy Grids.

115. D. N. A. Boobbyer, P. J. Goodford, P. M. McWhinnie, and R. C. Wade, *J. Med. Chem.,* **32,** 1083 (1989). New Hydrogen-Bond Potentials for Use in Determining Energetically Favorable Binding Sites on Molecules of Known Structure.

116. R. C. Wade, K. J. Clark, and P. J. Goodford, *J. Med. Chem.,* **36,** 140 (1993). Further Development of Hydrogen-Bond Functions for Use in Determining Energetically Favorable Binding Sites on Molecules of Known Structure. 1. Ligand Probe Groups With the Ability to Form Two Hydrogen Bonds.

117. R. C. Wade and P. J. Goodford, *J. Med. Chem.,* **36,** 148 (1993). Further Development of Hydrogen Bond Functions for Use in Determining Energetically Favorable Binding Sites on Molecules of Known Structure. 2. Ligand Probe Groups With the Ability to Form More Than Two Hydrogen Bonds.

118. M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCammon, *Biophys. J.,* **72,** 1047 (1997). The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review.

119. T. P. Straatsma, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1991, Vol. 9, pp. 81–127. Free Energy by Molecular Simulation.

120. P. A. Kollman, *Acc. Chem. Res.,* **29,** 461 (1996). Advances and Continuing Challenges in Achieving Realistic and Predictive Simulations of the Properties of Organic and Biological Molecules.

121. M. L. Lamb and W. L. Jorgensen, *Curr. Opin. Chem. Biol.,* **1,** 449 (1997). Computational Approaches to Molecular Recognition.

122. M. R. Reddy, M. D. Erion, and A. Agarwal, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., Wiley-VCH, New York, 2000, Vol. 16, pp. 217–304. Free Energy Calculations: Use and Limitations in Predicting Ligand Binding Affinities.

123. M. K. Gilson, J. A. Given, and M. S. Head, *Chem. Biol.,* **4,** 87 (1997). A New Class of Models for Computing Receptor–Ligand Binding Affinities.

124. A. E. Mark and W. F. van Gunsteren, *J. Mol. Biol.,* **240,** 167 (1994). Decomposition of the Free Energy of a System in Terms of Specific Interactions.

125. D. Williams and B. Bardsley, *Perspect. Drug Discovery Design,* **17,** 43 (1999). Estimating Binding Constants—The Hydrophobic Effect and Cooperativity.

126. P. R. Andrews, D. J. Craik, and J. L. Martin, *J. Med. Chem.,* **27,** 1648 (1984). Functional Group Contributions to Drug–Receptor Interactions.

127. H.-J. Schneider, *Chem. Soc. Rev.,* 227 (1994). Linear Free Energy Relationships and Pairwise Interactions in Supramolecular Chemistry.

128. T. J. Stout, C. R. Sage, and R. M. Stroud, *Structure,* **6,** 839 (1998). The Additivity of Substrate Fragments in Enzyme–Ligand Binding.

129. M. K. Holloway, J. M. Wai, T. A. Halgren, P. M. D. Fitzgerald, J. P. Vacca, B. D. Dorsey, R. B. Levin, W. J. Thompson, L. J. Chen, S. J. deSolms, N. Gaffin, T. A. Lyle, W. A. Sanders, T. J. Tucker, M. Wiggins, C. M. Wiscount, O. W. Woltersdorf, S. D. Young, P. L. Darke, and J. A. Zugay, *J. Med. Chem.,* **38,** 305 (1995). A Priori Prediction of Activity for HIV-Protease Inhibitors Employing Energy Minimization in the Active Site.

130. P. D. J. Grootenhuis and P. J. M. van Galen, *Acta Crystallogr., Sect. D,* **51,** 560 (1995). Correlation of Binding Affinities With Non-Bonded Interaction Energies of Thrombin-Inhibitor Complexes.

131. S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta Jr., and P. Weiner, *J. Am. Chem. Soc.,* **106,** 765 (1984). A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins.

132. S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, *J. Comput. Chem.,* **7,** 230 (1986). An All Atom Force Field for Simulations of Proteins and Nucleic Acids.

133. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.,* **4,** 187 (1983). CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations.

134. A. Nicholls and B. Honig, *Science,* **268,** 1144 (1995). Classical Electrostatics in Biology and Chemistry.

135. N. Majeux, M. Scarsi, J. Apostolakis, C. Ehrhardt, and A. Caflisch, *Proteins: Struct., Funct., Genet.,* **37,** 88 (1999). Exhaustive Docking of Molecular Fragments with Electrostatic Solvation.

136. W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, *J. Am. Chem. Soc.,* **112,** 6127 (1990). Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics.

137. M. L. Lamb, K. W. Burdick, S. Toba, M. M. Young, A. G. Skillman, X. Zou, J. R. Arnold, and I. D. Kuntz, *Proteins: Struct., Funct., Genet.,* **42,** 296 (2001). Design, Docking and Evaluation of Multiple Libraries Against Multiple Targets.

138. T. Zhang and D. E. Koshland Jr., *Protein Sci.,* **5,** 348 (1996). Computational Method for Relative Binding Free Energies of Enzyme–Substrate Complexes.

139. P. H. Huenenberger, V. Helms, N. Narayana, S. S. Taylor, and J. A. McCammon, *Biochemistry,* **38,** 2358 (1999). Determinants of Ligand Binding to cAMP-Dependent Protein Kinase.

140. H. Meirovitch, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., Wiley-VCH, New York, 1998, Vol. 11, pp. 1–74. Calculation of the Free Energy and the Entropy of Macromolecular Systems by Computer Simulation.

141. J. Greer, J. W. Erickson, J. J. Baldwin, and M. D. Varney, *J. Med. Chem.,* **37,** 1035 (1994). Application of the Three-Dimensional Structures of Protein Target Molecules in Structure-Based Design.

142. J. Bostrom, P.-O. Norrby, and T. Liljefors, *J. Comput.-Aided Mol. Design,* **12,** 383 (1998). Conformational Energy Penalties of Protein-Bound Ligands.

143. M. Vieth, J. D. Hirst, and C. L. Brooks III, *J. Comput.-Aided Mol. Design,* **12,** 563 (1998). Do Active Site Conformations of Small Ligands Correspond to Low Free-Energy Solution Structures ?

144. G. Klebe and T. Mietzner, *J. Comput.-Aided Mol. Design,* **8,** 583 (1994). A Fast and Efficient Method to Generate Biologically Relevant Conformations.

145. I. D. Kuntz, K. Chen, K. A. Sharp, and P. A. Kollman, *Proc. Natl. Acad. Sci. U.S.A.,* **96,** 9997 (1999). The Maximal Affinity of Ligands.

146. Ajay and M. A. Murcko, *J. Med. Chem.,* **38,** 4953 (1995). Computational Methods to Predict Binding Free Energy in Ligand–Receptor Complexes.

147. J. D. Hirst, *Curr. Opin. Drug Discovery Dev.,* **1,** 28 (1998). Predicting Ligand-Binding Energies.

148. R. M. A. Knegtel and P. D. J. Grootenhuis, *Perspect. Drug Discovery Design,* **9–11,** 99 (1998). Binding Affinities and Non-Bonded Interaction Energies.

149. T. I. Oprea and G. R. Marshall, *Perspect. Drug Discovery Design,* **9–11,** 3 (1998). Receptor-Based Prediction of Binding Activities.

150. J. R. H. Tame, *J. Comput.-Aided Mol. Design,* **13,** 99 (1999). Scoring Functions: A View From the Bench.

151. H.-J. Boehm and M. Stahl, *Med. Chem. Res.,* **9,** 445 (1999). Rapid Empirical Scoring Functions in Virtual Screening Applications.

152. P. J. Goodford, *J. Mol. Graphics*, **3**, 107 (1985). Favored Sites for Drug Design.

153. M. Matsumara, W. J. Becktel, and B. W. Matthews, *Nature (London)*, **334**, 406 (1988). Hydrophobic Stabilization in T4 Lysozyme Determined Directly by Multiple Substitutions of Ile 3.

154. V. Nauchatel, M. C. Villaverde, and F. Sussman, *Protein Sci.*, **4**, 1356 (1995). Solvent Accessibility as a Predictive Tool for the Free Energy of Inhibitor Binding to the HIV-1 Protease.

155. A. M. Davis and S. J. Teague, *Angew. Chem. Int. Ed.*, **38**, 736 (1999). Hydrogen Bonding, Hydrophobic Interactions, and Failure of the Rigid Receptor Hypothesis.

156. L. Wesson and D. Eisenberg, *Protein Sci.*, **1**, 227 (1992). Atomic Solvation Parameters Applied to Molecular Dynamics of Proteins in Solution.

157. S. Vajda, Z. Weng, R. Rosenfeld, and C. DeLisi, *Biochemistry*, **33**, 13977 (1994). Effect of Conformational Flexibility and Solvation on Receptor-Ligand Binding Free Energies.

158. S. Miyazawa and R. L. Jernigan, *Macromolecules*, **18**, 534 (1985). Estimation of Effective Interresidue Contact Energies From Protein Crystal Structures: Quasi-Chemical Approximation.

159. P. Rose, Computer Assisted Molecular Design Course, Jan 16–18, 1997, University of California, San Francisco, CA.

160. E. Gallicchio, M. M. Kubo, and R. M. Levy, *J. Am. Chem. Soc.*, **120**, 4526 (1998). Entropy–Enthalpy Compensation in Solvation and Ligand Binding Revisited.

161. F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, *J. Mol. Biol.*, **112**, 535 (1977). The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures. http://www.rcsb.org/pdb.

162. H. Gohlke and G. Klebe, *Curr. Opin. Struct. Biol.*, **11**, 231 (2001). Statistical Potentials and Scoring Functions Applied to Protein–Ligand Binding.

163. M. J. Sippl, *J. Comput.-Aided Mol. Design*, **7**, 473 (1993). Boltzmann's Principle, Knowledge-Based Mean Fields and Protein Folding. An Approach to the Computational Determination of Protein Structures.

164. G. Verkhivker, K. Appelt, S. T. Freer, and J. E. Villafranca, *Protein Eng.*, **8**, 677 (1995). Empirical Free Energy Calculations of Ligand–Protein Crystallographic Complexes. I. Knowledge-Based Ligand Protein Interaction Potentials Applied to the Prediction of Human Immunodeficiency Virus 1 Protease Binding Affinity.

165. A. Wallqvist, R. L. Jernigan, and D. G. Covell, *Protein Sci.*, **4**, 1181 (1995). A Preference-Based Free-Energy Parameterization of Enzyme-Inhibitor Binding. Applications to HIV-1-Protease Inhibitor Design.

166. A. Wallqvist and D. G. Covell, *Proteins: Struct., Funct., Genet.*, **25**, 403 (1996). Docking Enzyme-Inhibitor Complexes Using a Preference-Based Free-Energy Surface.

167. Y. Sun, T. J. A. Ewing, A. G. Skillman, and I. D. Kuntz, *J. Comput.-Aided Mol. Design*, **12**, 579 (1998). CombiDOCK: Structure-Based Combinatorial Docking and Library Design.

168. E. J. Martin, R. E. Critchlow, D. C. Spellmeyer, S. Rosenberg, K. L. Spear, and J. M. Blaney, *Pharmacochem. Libr.*, **29**, 133 (1998). Diverse Approaches to Combinatorial Library Design.

169. DOCK User Manual, Regents of the University of California, San Francisco, CA. http://www.cmpharm.ucsf.edu/kuntz/kuntz.html.

170. R. M. A. Knegtel, D. M. Bayada, R. A. Engh, W. von der Saal, V. J. van Geerestein, and P. D. J. Grootenhuis, *J. Comput.-Aided Mol. Design*, **13**, 167 (1999). Comparison of Two Implementations of the Incremental Construction Algorithm in Flexible Docking of Thrombin Inhibitors.

171. M. Stahl and H.-J. Boehm, *J. Mol. Graphics Modell.*, **16**, 121 (1998). Development of Filter Functions for Protein–Ligand Docking.

172. C. Lemmen, T. Lengauer, and G. Klebe, *J. Med. Chem.,* **41,** 4502 (1998). FlexS: A Method for Fast Flexible Ligand Superposition.

173. C. Lemmen, A. Zien, R. Zimmer, and T. Lengauer, *Abstracts of the Pacific Symposium on Biocomputing,* Big Island, Hawaii, January 4–9, 1999, pp. 482–493. Application of Parameter Optimization to Molecular Comparison.

174. G. R. Desiraju and T. Steiner, *The Weak Hydrogen Bond in Chemistry and Biology,* Oxford University Press, Oxford, UK, 1999.

175. G. Parkinson, A. Gunasekera, J. Vojtechovsky, X. Zhang, T. Kunkel, H. Berman, and R. H. Ebright, *Nature Struct. Biol.,* **3,** 837 (1996). Aromatic Hydrogen Bond in Sequence-Specific Protein DNA Recognition.

176. J. P. Gallivan and D. A. Dougherty, *Proc. Natl. Acad. Sci. U.S.A.,* **96,** 9459 (1999). Cation–π Interactions in Structural Biology.

177. J. P. Gallivan and D. A. Dougherty, *J. Am. Chem. Soc.,* **122,** 870 (2000). A Computational Study of Cation–π Interactions vs. Salt Bridges in Aqueous Media: Implications for Protein Engineering.

178. G. B. McGaughey, M. Gagné, and A. K. Rappé, *J. Biol. Chem.,* **273,** 15458 (1998). π-Stacking Interactions.

179. C. Chipot, R. Jaffe, B. Maigret, D. A. Pearlman, and P. A. Kollman, *J. Am. Chem. Soc.,* **118,** 11217 (1996). The Benzene Dimer: A Good Model for π–π Interactions in Proteins? A Comparison Between the Benzene and the Toluene Dimers in the Gas Phase and in Aqueous Solution.

180. T. G. Davies, R. E. Hubbard, and J. R. H. Tame, *Protein Sci.,* **8,** 1432 (1999). Relating Structure to Thermodynamics: The Crystal Structures and Binding Affinity of Eight OppA-Peptide Complexes.

181. G. Klebe, F. Dullweber, and H.-J. Boehm, in *Drug-Receptor Thermodynamics: Introduction and Applications,* R. B. Raffa, Ed., Wiley, Chichester, UK, 2001, pp. 83–103. Thermodynamic Models of Drug-Receptor Interactions: A General Introduction.

182. S. Ha, R. Andreani, A. Robbins, and I. Muegge, *J. Comput.-Aided Mol. Design,* **14,** 435 (2000). Evaluation of Docking/Scoring Approaches: A Comparative Study Based on MMP3 Inhibitors.

183. I. Massova and P. A. Kollman, *Perspect. Drug Discovery Design,* **18,** 113 (2000). Combined Molecular Mechanical and Continuum Solvent Approach (MM-PBSA/GBSA) to Predict Ligand Binding.

184. B. Kuhn and P. A. Kollman, *J. Med. Chem.,* **43,** 3786 (2000). Binding of a Diverse Set of Ligands to Avidin and Streptavidin: An Accurate Quantitative Prediction of Their Relative Affinities by a Combination of Molecular Mechanics and Continuum Solvent Models.

185. N. Froloff, A. Windemuth, and B. Honig, *Protein Sci.,* **6,** 1293 (1997). On the Calculation of Binding Free Energies Using Continuum Methods: Application to MHC Class I Protein–Peptide Interactions.

186. J. Shen, *J. Med. Chem.,* **40,** 2953 (1997). A Theoretical Investigation of Tight-Binding Thermolysin Inhibitors.

187. C. J. Woods, M. A. King, and J. W. Essex, *J. Comput.-Aided Mol. Design,* **15,** 129 (2001). The Configurational Dependence of Binding Free Energies: A Poisson–Boltzmann Study of Neuraminidase Inhibitors.

188. G. Archontis, T. Simonson, and M. Karplus, *J. Mol. Biol.,* **306,** 307 (2001). Binding Free Energies and Free Energy Components from Molecular Dynamics and Poisson–Boltzmann Calculations. Application to Amino Acid Recognition by Aspartyl–tRNA Synthetase.

189. A. R. Leach, *J. Mol. Biol.,* **235,** 345 (1994). Ligand Docking to Proteins With Discrete Side-Chain Flexibility.

190. R. M. A. Knegtel, I. D. Kuntz, and C. M. Oshiro, *J. Mol. Biol.,* **266,** 424 (1997). Molecular Docking to Ensembles of Protein Structures.

191. C. McMartin and R. S. Bohacek, *J. Comput.-Aided Mol. Design,* **11,** 333 (1997). QXP: Powerful, Rapid Computer Algorithms for Structure-Based Drug Design.

192. J. Apostolakis, A. Plückthun, and A. Caflisch, *J. Comput. Chem.,* **19,** 21 (1998). Docking Small Ligands in Flexible Binding Sites.

193. V. Schnecke, C. A. Swanson, E. D. Getzoff, J. A. Tainer, and L. A. Kuhn, *Proteins: Struct., Funct., Genet.,* **33,** 74 (1998). Screening a Peptidyl Database for Potential Ligands to Proteins With Side-Chain Flexibility.

194. I. Kolossváry, and W. C. Guida, *J. Comput. Chem.,* **20,** 1671 (1999). Low-Mode Conformational Search Elucidated: Application to $C_{39}H_{80}$ and Flexible Docking of 9-Deazaguanine Inhibitors to PNP.

195. H. B. Broughton, *J. Mol. Graphics Modell.,* **18,** 247 (2000). A Method for Including Protein Flexibility in Protein–Ligand Docking: Improving Tools for Database Mining and Virtual Screening.

196. H. Claussen, C. Buning, M. Rarey, and T. Lengauer, *J. Mol. Biol.,* **308,** 377 (2001). FlexE: Efficient Molecular Docking Considering Protein Structure Variations.

197. M. D. Miller, R. P. Sheridan, S. K. Kearsley, and D. J. Underwood, in *Methods in Enzymology,* L. C. Kuo and J. A. Shafer, Eds., Academic Press, San Diego, 1994, Vol. 241, pp. 354–370. Advances in Automated Docking Applied to Human Immunodeficiency Virus Type 1 Protease.

198. P. S. Charifson, J. J. Corkery, M. A. Murcko, and W. P. Walters, *J. Med. Chem.,* **42,** 5100 (1999). Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins.

199. C. A. Baxter, C. W. Murray, D. E. Clark, D. R. Westhead, and M. D. Eldridge, *Proteins: Struct., Funct., Genet.,* **33,** 367 (1998). Flexible Docking Using Tabu Search and an Empirical Estimate of Binding Affinity.

200. WDI: World Drug Index, 1996, Derwent Information. http://www.derwent.com.

201. Protherics PLC, Runcorn, Cheshire, UK. The data set was formerly available at http://www.protherics.com/crunch/.

202. S. B. Shuker, P. J. Hajduk, R. P. Meadows, and S. W. Fesik, *Science,* **274,** 1531 (1996). Discovering High-Affinity Ligands for Proteins: SAR by NMR.

203. N. Tomioka and A. Itai, *J. Comput.-Aided Mol. Design,* **8,** 347 (1994). GREEN: A Program Package for Docking Studies in Rational Drug Design.

204. T. Toyoda, R. K. B. Brobey, G.-I. Sano, T. Horii, N. Tomioka, and A. Itai, *Biochem. Biophys. Res. Commun.,* **235,** 515 (1997). Lead Discovery of Inhibitors of the Dihydrofolate Reductase Domain of *Plasmodium falciparum* Dihydrofolate Reductase–Thymidylate Synthase.

205. P. Burkhard, P. Taylor, and M. D. Walkinshaw, *J. Mol. Biol.,* **277,** 449 (1998). An Example of a Protein Ligand Found by Database Mining: Description of the Docking Method and Its Verification by a 2.3 Å X-Ray Structure of a Thrombin–Ligand Complex.

206. R. Abagyan, M. Trotov, and D. Kuznetsov, *J. Comput. Chem.,* **15,** 488 (1994). ICM—A New Method for Protein Modeling and Design: Applications to Docking and Structure Prediction from the Distorted Native Conformation.

207. M. Schapira, B. M. Raaka, H. H. Samuels, and R. Abagyan, *Proc. Natl. Acad. Sci. U.S.A.,* **97,** 1008 (2000). Rational Discovery of Novel Nuclear Hormone Receptor Antagonists.

208. A. V. Filikov, V. Monan, T. A. Vickers, R. H. Griffey, P. D. Cook, R. A. Abagyan, and T. L. James, *J. Comput.-Aided Mol. Design,* **14,** 593 (2000). Identification of Ligands for RNA Targets via Structure-Based Virtual Screening: HIV-1 TAR.

209. Available Chemicals Directory, MDL Information Systems, Inc., San Leandro, CA, USA. http://www.mdl.com.

210. D. W. Christianson and C. A. Fierke, *Acc. Chem. Res.,* **29,** 331 (1996). Carbonic Anhydrase: Evolution of the Zinc-Binding Site by Nature and by Design.

211. Maybridge Database, Maybridge Chemical Co. Ltd., UK, 1999. http://www.maybridge.com and http://www.daylight.com.

212. LeadQuest Chemical Compound Libraries, Vol. 1–3, 2000, Tripos Inc., St. Louis, MO, USA. http://www.tripos.com.

213. J. Sadowski, C. Rudolph, and J. Gasteiger, *Tetrahedron Comput. Methodol.,* **3,** 537 (1990). Automatic Generation of 3D Atomic Coordinates for Organic Molecules.

214. J. Sadowski, C. H. Schwab, and J. Gasteiger, Corina, 1998, Molecular Networks GmbH Computerchemie, Erlangen, Germany. http://www.mol-net.de/.

215. Unity Chemical Information Software, Version 4.1.1., Tripos, Inc., St. Louis, MO, USA.

216. A. Maxwell, *Biochem. Soc. Trans.,* **27,** 48 (1999). DNA Gyrase as a Drug Target.

217. D. Bailey and D. Brown, *Drug Discovery Today,* **6,** 57 (2001). High-Throughput Chemistry and Structure-Based Design: Survival of the Smartest.

218. R. M. Eglen, G. Scheider, and H.-J. Boehm, in *Virtual Screening for Bioactive Molecules,* G. Schneider and H.-J. Boehm, Eds., Wiley-VCH, Weinheim, 2000, pp. 1–14. High-Throughput Screening and Virtual Screening: Entry Points to Drug Discovery.

219. R. Lahana, *Drug Discovery Today,* **4,** 447 (1999). How Many Leads from HTS?

CHAPTER 3

Potentials and Algorithms for Incorporating Polarizability in Computer Simulations

Steven W. Rick* and Steven J. Stuart[†]

*Department of Chemistry, University of New Orleans, New Orleans, Louisiana 70148 and Chemistry Department, Southern University of New Orleans, New Orleans, Louisiana 70126, and [†]*Department of Chemistry, Clemson University, Clemson, South Carolina 29634*

INTRODUCTION

Polarization refers to the redistribution of a particle's electron density due to an electric field. In terms of molecular interactions, polarization leads to nonadditivity, since a molecule polarized by another molecule will interact differently with a third molecule than it would if it were not polarized. The change in the electron density can be characterized by changes in the monopole charges, dipole moments, or higher order moments. Methods for treating polarizability in molecular dynamics or Monte Carlo simulations achieve this goal through inducible dipole moments (the polarizable point dipole and shell models) or through fluctuating charges (the electronegativity equalization and semiempirical models). This chapter describes these models, with a focus on those methods that have been developed for molecular dynamics and Monte

Carlo computer simulations, and reviews some of the systems that have been simulated with polarizable potentials.

NONPOLARIZABLE MODELS

Before discussing polarizable models, a useful starting point is to consider nonpolarizable models. A typical nonpolarizable potential for molecular systems is[1]

$$U = \sum_{\text{bonds}} K_B(r - r_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} \sum_n \frac{V_n}{2}(1 + \cos(n\phi - \gamma))$$

$$+ \sum_{\text{nonbonded pairs}} \left\{ 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right] + \frac{q_i q_j}{r_{ij}}\right\} \qquad [1]$$

where U represents the potential energy of the system. There are terms for the bond length, r, with a force constant, K_B, and an equilibrium bond length, r_0; the bond angle, θ, with a force constant K_θ and an equilibrium angle, θ_0; and the dihedral angle, ϕ, with barrier heights, V_n, and equilibrium angles, γ. The intermolecular interactions are described with a Lennard–Jones (LJ) interaction,

$$U_{\text{LJ}}(r) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right] \qquad [2]$$

in which ε and σ are parameters describing the energy and distance scale of the interactions, respectively, and r_{ij} is the distance between nonbonded atoms i and j. The Coulomb interaction between charged atoms is given by $q_i q_j/r_{ij}$, where q_i is the partial charge on atom i. These interactions are illustrated in Figure 1.

The Lennard–Jones interaction contains a short-range repulsive part, falling off as r^{-12}, and a longer range attractive part, falling off as r^{-6}. The attractive part has the same dependence as the (dipole–dipole) London dispersion energy, which for two particles with polarizabilities α is proportional to $-\alpha^2/r^6$ (Ref. 2). The Lennard–Jones parameters are not typically assigned[3] using known values of α, but this interaction is one way in which polarizability, in an average sense, is included in nonpolarizable models.

Another way in which polarizability is included implicitly is in the value of the partial charges, q_i, that are assigned to the atoms in the model. The charges used in potential energy models for condensed phases are often enhanced from the values that would be consistent with the gas-phase dipole moment, or those that would best reproduce the electrostatic potential (ESP)

$$\text{H} \quad \text{H}$$
$$\text{O}$$

$$(q_i q_j / r_{ij}) + 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

$$\text{H}$$

$$\frac{V_n}{2}(1 + \cos(n\omega + \gamma)) \quad \text{N}$$

$$\text{C}$$

$$K_\theta(\theta - \theta_0)^2 \qquad K_B(r - r_0)^2$$

$$\text{O}$$
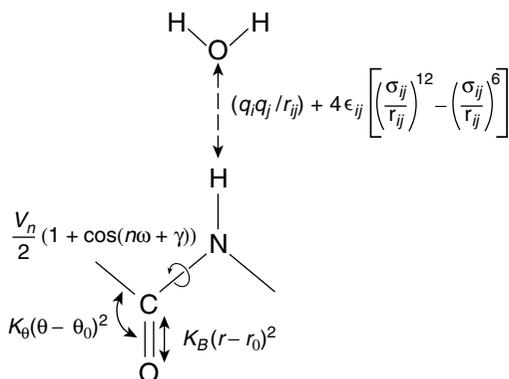
Figure 1 Schematic of the interactions between an amino group and a water showing the Lennard–Jones and electrostatic nonbonded interactions along with the bond length, bond angle, and dihedral angle (torsional) interactions.

from gas-phase ab initio calculations. Enhanced charge values are a means of accounting for the strong polarization of electron distributions by the electric fields of the other particles in a condensed phase environment. The enhanced charges are obtained either through explicit parameterization[4,5] or by using charges obtained via quantum chemical methods that are known to overestimate charge values.[6] Although the enhanced charge values treat polarization in an effective way, they cannot correctly reflect the dependence of charge distributions on the system's state, nor can they respond dynamically to fluctuations in the electric field due to molecular motion. The average electric field, and therefore the charge distribution and dipole moment, will depend on the physical state and composition of the system. For example, a molecule in a solution with a high ionic strength may feel a field different from a molecule in a pure solvent; even in the bulk liquid state, the polarization of a water molecule will depend on the density, and thus on the system's temperature and pressure. In addition, conformational changes may influence the charge distribution of a molecule.[7–13] Molecular motions in the system will result in conformational changes and fluctuations in the electric field, causing the electrostatic distribution to change on a subpicosecond time scale. Treating these effects requires a polarizable model.

POLARIZABLE POINT DIPOLES

One method for treating polarizability is to add point inducible dipoles on some or all atomic sites. This polarizable point dipoles (PPD) method has been applied to a wide variety of atomic and molecular systems, ranging from noble gases to water to proteins. The dipole moment, $\boldsymbol{\mu}_i$, induced on a site i is

proportional to the electric field at that site, $\mathbf{E}_i$. The proportionality constant is the polarizability tensor, $\boldsymbol{\alpha}_i$. The dipole feels an electric field both from the permanent charges of the system and from the other induced dipoles. The expression for the $\boldsymbol{\mu}_i$ is

$$\boldsymbol{\mu}_i = \boldsymbol{\alpha}_i \cdot \mathbf{E}_i = \boldsymbol{\alpha}_i \cdot \left[\mathbf{E}_i^0 - \sum_{j \neq i} \mathbf{T}_{ij} \boldsymbol{\mu}_j \right] \qquad [3]$$

where $\mathbf{E}^0$ is the field from the permanent charges. (There also may be permanent dipoles or other multipoles present contributing to $\mathbf{E}^0$.) The induced dipoles interact through the dipole field tensor, $\mathbf{T}_{ij}$,

$$\mathbf{T}_{ij} = \frac{1}{r^3}\mathbf{I} - \frac{3}{r^5} \begin{pmatrix} x^2 & xy & xz \\ yx & y^2 & yz \\ zx & zy & z^2 \end{pmatrix} \qquad [4]$$

where $\mathbf{I}$ is the identity matrix, r is the distance between i and j, and x, y, and z are the Cartesian components of the vector between i and j.

The energy of the induced dipoles, U_{ind}, can be split into three terms,

$$U_{\text{ind}} = U_{\text{stat}} + U_{\mu\mu} + U_{\text{pol}} \qquad [5]$$

The energy U_{stat} is the interaction energy of the N induced dipoles with the permanent, or static, field

$$U_{\text{stat}} = -\sum_{i=1}^{N} \boldsymbol{\mu}_i \cdot \mathbf{E}_i^0 \qquad [6]$$

the energy $U_{\mu\mu}$ is the induced dipole–induced dipole interaction

$$U_{\mu\mu} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j \neq i} \boldsymbol{\mu}_i \cdot \mathbf{T}_{ij} \cdot \boldsymbol{\mu}_j \qquad [7]$$

and the polarization energy, U_{pol},

$$U_{\text{pol}} = \frac{1}{2}\sum_{i=1}^{N} \boldsymbol{\mu}_i \cdot \mathbf{E}_i \qquad [8]$$

is that required to distort the electron distribution to create the dipoles.[4,14] Any polarizable model in which dipole moments, charges, or other multipoles

are modified by their environment will have a polarization energy corresponding to U_{pol}. Even nonpolarizable models that are parameterized to have charges enhanced from the gas-phase values should include such a term, and U_{pol} has been called the "missing term" in many pair potentials.[4,5] By using Eq. [3], the electric field can be replaced by $\boldsymbol{\alpha}^{-1} \cdot \boldsymbol{\mu}_i$ and U_{pol} can be written as

$$U_{pol} = \frac{1}{2} \sum_{i=1}^{N} \boldsymbol{\mu}_i \cdot \boldsymbol{\alpha}_i^{-1} \cdot \boldsymbol{\mu}_i \qquad [9]$$

where $\boldsymbol{\alpha}_i^{-1}$ is the inverse of the polarization tensor. If the polarization matrix is isotropic ($\alpha_{xx} = \alpha_{yy} = \alpha_{zz} = \alpha_i$) and diagonal, then

$$U_{pol} = \sum_{i=1}^{N} \frac{\mu_i^2}{2\alpha_i} \qquad [10]$$

Combining the three energy terms gives

$$U_{ind} = \sum_{i=1}^{N} \boldsymbol{\mu}_i \cdot \left[-\mathbf{E}_i^0 + \frac{1}{2} \sum_{j \neq i} \mathbf{T}_{ij} \cdot \boldsymbol{\mu}_j + \frac{1}{2} \mathbf{E}_i \right] \qquad [11]$$

which, using $\mathbf{E}_i = \mathbf{E}_i^0 - \sum \mathbf{T}_{ij} \cdot \boldsymbol{\mu}_j$ (Eq. [3]), reduces to the relationship for the static field, $\mathbf{E}^0$, and

$$U_{ind} = -\frac{1}{2} \sum_{i=1}^{N} \boldsymbol{\mu}_i \cdot \mathbf{E}_i^0 \qquad [12]$$

Note that the energy is the dot product of the induced dipole and the *static* field, not the total field.[15–19] Without a static field, there are no induced dipoles. Induced dipoles alone do not interact strongly enough to overcome the polarization energy it takes to create them (except when they are close enough to polarize catastrophically).

The static field at site i due to permanent charges is

$$\mathbf{E}_i^0 = \sum_{j \neq i} \frac{q_j \mathbf{r}_{ij}}{r_{ij}^3} \qquad [13]$$

where q_j is the charge at site j and r_{ij} is the distance between i and j. A point charge at a site is generally assumed not to contribute to the field at that site. For rigid models of water and other small molecules, charges on the same molecule contribute a constant amount to the electric field at each site

(in internal coordinates). These effects are often incorporated into the fixed charge distribution and are not explicitly included in the static field, which is calculated using only charges from different molecules.[15,19–37]

Some water models use a shielding function, $S(r)$, that changes the contribution to $\mathbf{E}_i$ from the charge at j,[20,26,30,31,37]

$$\mathbf{E}_i = \sum_{j \neq i} \frac{S(r_{ij})q_j\mathbf{r}_{ij}}{|r_{ij}|^3} \tag{14}$$

The shielding function differs from 1 only at small distances and accounts for the fact that at small separations the electric field will be modified by the spatial extent of the electron cloud. For larger molecules, the interactions from atoms that are directly bonded to atom i and are separated by two bonds or less (termed 1–2 and 1–3 bonded interactions) do not typically contribute to $\mathbf{E}_i$.[32,38]

In the most general case, all the dipoles will interact through the dipole field tensor. The method of Applequist et al.[39,40] for calculating molecular polarizabilities uses this approach. One problem with coupling all the dipoles with the interaction given by Eq. [4] is the "polarization catastrophe". As pointed out by Applequist, Carl, and Fung[39] and Thole,[41] the molecular polarization, and therefore the induced dipole moment, may become infinite at small distances. The mathematical origins of such singularities are made more evident by considering a simple system consisting of two atoms (A and B) with isotropic polarizabilities, α_A and α_B. The molecular polarizability, which relates the molecular dipole moment ($\mu = \mu_A + \mu_B$) to the electric field, has two components, one parallel and one perpendicular to the bond axis between A and B,

$$\alpha_{\|} = [\alpha_A + \alpha_B + (4\alpha_A\alpha_B/r^3)]/[1 - (4\alpha_A\alpha_B/r^6)] \tag{15}$$

$$\alpha_{\perp} = [\alpha_A + \alpha_B - (2\alpha_A\alpha_B/r^3)]/[1 - (\alpha_A\alpha_B/r^6)] \tag{16}$$

The parallel component, $\alpha_{\|}$, becomes infinite as the distance between the two atoms approaches $(4\alpha_A\alpha_B)^{1/6}$. The singularities can be avoided by making the polarizabilities sufficiently small so that at the typical distances between the atoms ($>1\,\text{Å}$) the factor $(4\alpha_A\alpha_B)/r^6$ is always less than one. The Applequist polarizabilities are in fact small compared to ab initio values.[41,42] Applequist's atomic polarizabilities were selected to optimize the molecular polarizabilities for a set of 41 molecules (see Table 1). Note that careful choice of polarizabilities can move the singularities in Eqs. [15] and [16] to small distances, but not eliminate them completely, thus causing problems for simulation techniques such as Monte Carlo (MC), which tend to sample these nonphysical regions of configuration space.

Table 1 Polarizability Parameters for Atoms

Atom	Polarizability (Å^3)		
	Applequist et al.[a]	Thole[b]	Experimental or ab initio[c]
H (alkane)	0.135	0.514	0.667
H (alcohol)	0.135	—	—
H (aldehyde)	0.167	—	—
H (amide)	0.161	—	—
C (alkane)	0.878	1.405	1.76
C (carbonyl)	0.616	—	—
N (amide)	0.530	1.105	1.10
N (nitrile)	0.52	—	—
O (alcohol)	0.465	0.862	0.802
O (ether)	0.465	—	—
O (carbonyl)	0.434	—	—
F	0.32	—	0.557
Cl	1.91	—	2.18
Br	2.88	—	3.05
I	4.69	—	5.35

[a]Ref. 39.
[b]Ref. 41
[c]Ref. 42.

Alternatively, the polarization catastrophe can be avoided by screening (attenuating) the dipole–dipole interaction at small distances.[41] As with the screening of the static field, screening of the dipole–dipole interaction can be physically interpreted as correcting for the fact that the electronic distribution is not well represented by point charges and dipoles at small distances.[39,41,43] Mathematically, screening avoids the singularities such as those in Eqs. [15] and [16]. The Thole procedure for screening is to introduce a scaling distance $s_{ij} = 1.662(\alpha_i\alpha_j)^{1/6}$. This results in a charge density radius of 1.662 Å, for example, between atoms with a polarizability of 1 Å^3. The dipole field tensor is thus changed to

$$\mathbf{T}_{ij} = (4v^3 - 3v^4)\frac{1}{r^3}\mathbf{I} - v^4\frac{3}{r^5}\begin{pmatrix} x^2 & xy & xz \\ yx & y^2 & yz \\ zx & zy & z^2 \end{pmatrix} \qquad [17]$$

where $v = r/s_{ij}$. $\mathbf{T}_{ij}$ is unchanged if r is greater than s_{ij}. Thole's polarizability parameters, together with the scale factor 1.662, were selected to optimize the molecular polarizabilities for a set of 16 molecules (Table 1). Unlike Applequist, Thole assigns only one polarizability per atom independent of its valence state and does not assign polarizabilities to halide atoms. The Thole parameters are closer to the experimental and ab initio polarizabilities.[42] Although the atomic polarizabilities of Applequist and Thole are different, the resulting

molecular polarizabilities are not that far off, with the Applequist method tending to overestimate the polarization anisotropies.

Various computer simulation models have used either the Applequist parameters and no screening[15,16,24,27,28,32,34] or the Thole parameters and screening of $\mathbf{T}_{ij}$.[19,31,38] Different screening functions have been used as well. A large number of polarizable models have been developed for water, many of them with one polarizable site (with $\alpha = 1.44$ Å^3) on or near the oxygen position.[20–23,26,29,30,33,35–37] For these models, the polarizable sites do not typically get close enough for polarization catastrophes {$(4\alpha\alpha)^{1/6} = 1.4$ Å, see comments after Eq. [16]}, so screening is not as necessary as it would be if polarization sites were on all atoms. However, some water models with a single polarizable site do screen the dipole field tensor.[20,22,37] Another model for water places polarizable sites on bonds.[25] Other polarizable models have been used for monatomic ions and used no screening of $\mathbf{T}$ or $\mathbf{E}^0$.[15,16,27,34] Polarizable models have been developed for proteins as well, by Warshel and co-workers (with screening of $\mathbf{T}$ but not $\mathbf{E}^0$),[44,45] and by Wodak and co-workers (with no screening).[46]

An attractive feature of the dipole polarizable model is that the assignment of electrostatic potential parameters is more straightforward than for nonpolarizable models. Charges can be assigned on the basis of experimental dipole moments or ab initio electrostatic potential charges for the isolated molecule. The polarizabilities can be assigned from the literature (as in Table 1) or calculated. Contrarily, with nonpolarizable models, charges may have some permanent polarization to reflect their enhanced values in the condensed phase.[6,47] The degree of enhancement is part of the art of constructing potentials and limits the transferability of these potentials. By explicitly including polarizability, the polarizable models are a more systematic approach for potential parameterization and are therefore more transferrable.

Using Eqs. [9] and [11], the energy can be rewritten as

$$U_{\mathrm{ind}} = -\sum_{i=1}^{N} \boldsymbol{\mu}_i \cdot \mathbf{E}_i^0 + \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i} \boldsymbol{\mu}_i \cdot \mathbf{T}_{ij} \cdot \boldsymbol{\mu}_j + \frac{1}{2}\sum_{i=1}^{N} \boldsymbol{\mu}_i \boldsymbol{\alpha}_i^{-1} \boldsymbol{\mu}_i \qquad [18]$$

and the derivative of U_{ind} with respect to the induced dipoles is

$$\boldsymbol{\nabla}_{\boldsymbol{\mu}_i} U_{\mathrm{ind}} = -\mathbf{E}_i^0 + \sum_{j\neq i} \mathbf{T}_{ij} \cdot \boldsymbol{\mu}_j + \boldsymbol{\alpha}_i^{-1} \cdot \boldsymbol{\mu}_i = 0 \qquad [19]$$

The derivative in Eq. [19] is zero because $\boldsymbol{\alpha}_i^{-1} \cdot \boldsymbol{\mu}_i = \mathbf{E}_i^0 - \sum \mathbf{T}_{ij} \cdot \boldsymbol{\mu}_j$, according to Eq. [3]. The values of the induced dipoles are therefore those that minimize the energy. Other polarizable models also have auxiliary variables, analogous to $\boldsymbol{\mu}$, which likewise adjust to minimize the energy.

The polarizable point dipole models have been used in molecular dynamics (MD) simulations since the 1970s.[48] For these simulations, the forces, or spatial derivatives of the potential, are needed. From Eq. [18], the force[23] on atomic site k is

$$\mathbf{F}_k = -\nabla_k U_{\text{ind}} = \sum_{i=1}^{N} \boldsymbol{\mu}_i \nabla_k \mathbf{E}_i^0 + \sum_{i \neq k} \boldsymbol{\mu}_k \cdot (\nabla_k \mathbf{T}_{ki}) \cdot \boldsymbol{\mu}_i \qquad [20]$$

All contributions to the forces from terms involving derivatives with respect to the dipoles are zero from the extremum condition of Eq. [19].[23,49]

Finding the inducible dipoles requires a self-consistent method, because the field that each dipole feels depends on all of the other induced dipoles. There exist three methods for determining the dipoles: matrix inversion, iterative methods, and predictive methods. We describe each of these in turn.

The dipoles are coupled through the matrix equation,

$$\mathbf{A} \cdot \boldsymbol{\mu} = \mathbf{E}^0 \qquad [21]$$

where the diagonal elements of the matrix, A_{ii}, are $\boldsymbol{\alpha}_i^{-1}$ and the off-diagonal elements A_{ij} are T_{ij}. For a system with N dipoles, solving for each of them involves inverting the $N \times N$ matrix, $\mathbf{A}$—an $O(N^3)$ operation that is typically too computationally expensive to perform at each step of an $O(N)$ or $O(N^2)$ simulation. Consequently, this method has been used only rarely.[31] Note that since Eq. [21] for $\boldsymbol{\mu}$ is linear, there is only one solution for the dipoles.

In the iterative method, an initial guess for the field is made by, for example, just using the static field, $\mathbf{E}^0$, or by using the dipoles from the previous time step of the MD simulation.[48,49] The dipole moments resulting from this field are evaluated using Eq. [3], which can be iterated to self-consistency. Typical convergence limits on the dipoles range from 1×10^{-2} D to 1×10^{-6} D.[21,27,34–36,50] Long simulations require very strict convergence limits or mild thermostatting[50] to prevent problems due to poor energy conservation. Alternatively, the energy U_{pol} can be monitored for convergence.[19,51] The level of convergence, and therefore the number of iterations required, varies considerably. Between 2 and 10 iterations are typically required. For some calculations, including free energy calculations, a high level of convergence may be necessary.[38] The iterative method is the most common method for finding the dipoles.

The predictive methods determine $\boldsymbol{\mu}$ for the next time step based on information from previous time steps. Ahlström et al.[23] used a first-order predictor algorithm, which uses the $\boldsymbol{\mu}$ values from the two previous times steps to predict $\boldsymbol{\mu}$ at the next time step,

$$\boldsymbol{\mu}_i(t) = 2\boldsymbol{\mu}_i(t - \Delta t) - \boldsymbol{\mu}_i(t - 2\Delta t) \qquad [22]$$

where Δt is the time step, and t is time. This method is not stable for long times, but can be combined with an iterative solution, either by providing the initial iteration of the electric field values,[52,53] or by allowing the iteration to be performed less frequently than every step.[23,54] Higher-order predictor algorithms have been used as well.[23,52,55]

A different predictive procedure is to use the extended Lagrangian method, in which each dipole is treated as a dynamical variable and given a mass M_μ and velocity $\dot{\boldsymbol{\mu}}$. The dipoles thus have a kinetic energy, $\frac{1}{2}\sum_i M_\mu \dot{\boldsymbol{\mu}}_i^2$ and are propagated using the equations of motion just like the atomic coordinates.[22,56–58] The equation of motion for the dipoles is

$$M_\mu \ddot{\boldsymbol{\mu}}_i = -\boldsymbol{\nabla}_{\boldsymbol{\mu}_i} U_{\text{ind}} = \mathbf{E}_i - \boldsymbol{\alpha}_i^{-1} \cdot \boldsymbol{\mu}_i \qquad [23]$$

Here $\ddot{\boldsymbol{\mu}}_i$ is the second derivative with respect to time, that is, the acceleration. The dipole mass does not correspond to any physical mass of the system; it is chosen for numerical convenience, by, for example, comparing the trajectories with those from the iterative method.[57] It is desirable to keep the kinetic energy of the dipoles small so that the dipole degrees of freedom are cold and near the potential energy minimum (corresponding to the exact solution of Eq. [3]).

Because this method avoids iterations, which require recalculating $\mathbf{E}_i$ multiple times for every sampled configuration, the extended Lagrangian method is a more efficient way of calculating the dipoles at every time step. But even with methods that allow for only a single evaluation of the energy and force per time step, polarizable point dipole methods are more computationally intensive than nonpolarizable simulations. Evaluating the dipole–dipole interactions in Eqs. [7] and [20] is several times more expensive than evaluating the Coulombic interactions between point charges in Eq. [1]. A widely used rule of thumb is that polarizable simulations based on a point dipole model take roughly four times longer than a nonpolarizable simulation of the same system.

The polarizable point dipole model has also been used in Monte Carlo simulations with single particle moves.[19,21,24,59–62] When using the iterative method, a whole new set of dipoles must be computed after each molecule is moved. These updates can be made more efficient by storing the distances between all the particles, since most of them are unchanged, but this requires a lot of memory. The many-body nature of polarization makes it more amenable to molecular dynamics techniques, in which all particles move at once, compared to Monte Carlo methods where typically only one particle moves at a time. For nonpolarizable, pairwise-additive models, MC methods can be efficient because only the interactions involving the moved particle need to be recalculated [while the other $(N-1) \times (N-1)$ interactions are unchanged]. For polarizable models, all $N \times N$ interactions are, in principle, altered when one particle moves. Consequently, exact polarizable MC calculations can be

two to three orders of magnitude slower than comparable nonpolarizable calculations.[63] Various approximate methods, involving incomplete convergence or updating only a subset of the dipoles, have been suggested.[59] Unfortunately, these methods result in significant errors in computed physical properties.[19,63] Monte Carlo methods are capable of moving more than one particle at a time, with good acceptance ratios,[64,65] using, for example, the hybrid MC technique, but this method has not been applied to polarizable models, as far as we are aware.

One final point concerns the long-range nature of the interactions in dipole-based models. Dipole–dipole and dipole–charge interactions are termed long range because they do not decrease faster than volume grows—that is, as r^3. If periodic boundary conditions are used, some treatment of the long-range interactions is needed. The most complete treatment of the long-range forces is the Ewald summation technique.[64,66] All models, whether polarizable or not, face this problem if they have long-range forces, but for polarizable models this is a more significant issue. The use of cut-offs or other truncation schemes will change both the static field and the dipole field tensor. These changes to the electric field will modify the value of the induced dipole, which in turn will change the field at other sites. Accordingly, the treatment of long-range forces feeds back on itself in a way that does not occur with nonpolarizable models. It is thus crucial to treat the long-range interactions as accurately as possible in polarizable simulations. Nevertheless, a large number, if not most, of the simulations using polarizable potentials have not used Ewald sums. Recently, Nymand and Linse[67] showed that different boundary conditions (including Ewald sums, spherical cut-off, and reaction field methods) lead to more significant differences in equilibrium, dynamical, and structural properties for polarizable water models than for nonpolarizable models.

Conventional methods for performing the Ewald sum scale as $O(N^{3/2})$ or $O(N^2)$,[68] and formulations specifically designed to include dipole–dipole interactions[66] are in fairly wide use. Faster scaling methods, such as the fast multipole and particle–mesh algorithms, have also been extended to the treatment of point dipoles.[50,69]

SHELL MODELS

A defining feature of the models discussed in the previous section, regardless of whether they are implemented via matrix inversion, iterative techniques, or predictive methods, is that they all treat the polarization response in each polarizable center using point dipoles. An alternative approach is to model the polarizable centers using dipoles of finite length, represented by a pair of point charges. A variety of different models of polarizability have used this approach, but especially noteworthy are the shell models frequently used in simulations of solid-state ionic materials.

The shell model has its origin in the Born theory of lattice dynamics, used in studies of the phonon dispersion curves in crystals.[70,71] Although the Born theory includes the effects of polarization at each lattice site, it does not account for the short-range interactions between sites and, most importantly, neglects the effects of this interaction potential on the polarization behavior. The shell model, however, incorporates these short-range interactions.[72,73] The earliest applications of the shell model, as with the Born model, were to analytical studies of phonon dispersion relations in solids.[74] These early applications have been well reviewed elsewhere.[71,75–77] In general, lattice dynamics applications of the shell model do not attempt to account for the dynamics of the nuclei and typically use analytical techniques to describe the statistical mechanics of the shells. Although the shell model continues to be used in this fashion,[78] lattice dynamics applications are beyond the scope of this chapter. In recent decades, the shell model has come into widespread use as a model Hamiltonian for use in molecular dynamics simulations; it is these applications of the shell model that are of interest to us here.

The shell model to be described in detail below is essentially identical to the Drude oscillator model;[79,80] both treat polarization via a pair of charges attached by a harmonic spring. The different nomenclature results largely from the use of these models in recent decades by two different scientific communities. The term *Drude model* is used more frequently in simulations of the liquid state, whereas the term *shell model* is used more often in simulations of the solid state. As polarizable models become more common in both fields, the terms are beginning to be used indistinguishably. In this chapter, we will use the term shell model exclusively to describe polarizable models in which the dipoles are treated adiabatically; they are always at or near their minimum-energy conformation. We reserve the term Drude oscillator specifically for applications where the dipole oscillates either thermally or with a quantum mechanical zero-point energy, and this oscillating dipole gives rise to a dispersion interaction. The literature has not been entirely consistent on this point of terminology, but it is a useful distinction to make.

The shell model describes each polarizable ion or atom as a pair of point charges separated by a variable distance, as illustrated in Figure 2. These charges consist of a positive, "core" charge located at the site of the nucleus, and a negative, "shell" charge. These charges are connected by a harmonic spring. To some extent, these charges can be justified physically as an effective (shielded) nuclear charge and a corresponding effective charge in the valence shell that is responsible for most of the polarization response of the atom. This interpretation should not be taken literally, however; the magnitude of the charges are typically treated as adjustable parameters of the model rather than true shielded charge values. As such, they should be viewed primarily as an empirical method for representing the dipolar polarization of the site.

The magnitudes of both the core and shell charges are fixed in this model. The polarization thus occurs via relative displacement of the core
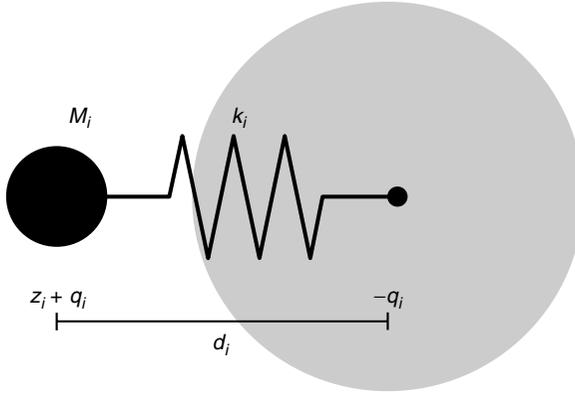
Figure 2 In the shell model, a "core" charge $z_i + q_i$ is attached by a harmonic spring with spring constant k_i to a "shell" charge $-q_i$. For a neutral atom, $z_i = 0$. The center of mass is at or near the core charge, but the short-range interactions are centered on the shell charge. (Not drawn to scale; the displacement d_i between the charges is much smaller than the atomic radius.)

and shell charges. For a neutral atom i with a core charge of $+q_i$, an equal and opposite shell charge of $-q_i$, and a shell charge that is displaced by a distance $\mathbf{d}_i$ from the core charge, the dipole moment is

$$\boldsymbol{\mu}_i = -q_i \mathbf{d}_i \qquad [24]$$

As with any model involving inducible dipoles, the potential energy of the induced dipoles contains terms representing the interaction with any static field, the interaction with other dipoles, and the polarization energy, that is,

$$U_{\text{ind}} = U_{\text{stat}} + U_{\mu\mu} + U_{\text{pol}} \qquad [25]$$

The polarization energy arises in this case from the harmonic spring separating the two charges,

$$U_{\text{pol}} = \frac{1}{2} \sum_{i=1}^{N} k_i d_i^2 \qquad [26]$$

for a collection of N polarizable atoms with spring constants k_i and charge displacements $d_i = |\mathbf{d}_i|$. Using $\mathbf{d}_i$ from Eq. [24] and comparing it with Eq. [10] for polarizable point dipoles, we see that the polarizability of an isotropic shell model atom can be written as

$$\alpha_i = q_i^2 / k_i \qquad [27]$$

The electrostatic interaction between independent polarizable atoms is simply the sum of the charge–charge interactions between the four charge sites,

$$U_{\mu\mu} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i} q_i q_j \left[\frac{1}{|\mathbf{r}_{ij}|} - \frac{1}{|\mathbf{r}_{ij}-\mathbf{d}_j|} - \frac{1}{|\mathbf{r}_{ij}+\mathbf{d}_i|} + \frac{1}{|\mathbf{r}_{ij}-\mathbf{d}_j+\mathbf{d}_i|} \right] \quad [28]$$

Typically, no Coulombic interactions are included between the core and shell charges on a single site. Note that the electrostatic interaction in this model is implemented using only the charge–charge terms already present in Eq. [1]. No new interaction types, such as the dipole field tensor $\mathbf{T}_{ij}$ of Eq. [7], are required. The computational advantage of avoiding dipole–dipole interactions is almost exactly nullified by the necessity of calculating four times as many charge–charge interactions, however.

The interaction of the induced dipoles with the static field is the sum of the terms for each individual charge site,

$$U_{\text{stat}} = -\sum_{i=1}^{N} q_i[\mathbf{r}_i \cdot \mathbf{E}_i^0 - (\mathbf{r}_i+\mathbf{d}_i) \cdot \mathbf{E}_i^{0\prime}] \quad [29]$$

where $\mathbf{E}_i^0$ is the static field at the location of the core charge, $\mathbf{r}_i$, and $\mathbf{E}_i^{0\prime}$ is the static field at the location of the shell charge, $\mathbf{r}_i + \mathbf{d}_i$. Note that $\mathbf{E}_i^0 \neq \mathbf{E}_i^{0\prime}$, in general.

Equations [28] and [29] correspond directly to Eqs. [7] and [6], but for the case of dipoles with finite extent. In that sense, models based on point dipoles can be seen as idealized versions of the shell model, in the limit of infinitely small dipoles. That is, the magnitude of the charges q_i and spring constants k_i approach infinity in such a way as to keep the atomic polarizabilities α_i constant. Indeed, in that limit, the displacements will approach zero in the shell model, and the two models will be entirely equivalent.

To the extent that the polarization of physical atoms results in dipole moments of finite length, it can be argued that the shell model is more physically realistic (the section on Applications will examine this argument in more detail). Of course, both models include additional approximations that may be even more severe than ignoring the finite electronic displacement upon polarization. Among these approximations are (1) the representation of the electronic charge density with point charges and/or dipoles, (2) the assumption of an isotropic electrostatic polarizability, and (3) the assumption that the electrostatic interactions can be terminated after the dipole–dipole term.

In describing the shell model, the charge q was described as an effective valence charge of the atom. In some applications of the shell model, the shell charge is indeed interpreted physically in this manner, and q is assigned based

on estimates of shielded charge values. More typically, however, this physical interpretation is relaxed, and q is used as an adjustable parameter in the fitting of the model. Recall that both q and k determine the polarizability of the atom (Eq. [27]). These parameters are often obtained from experimental values for the polarizability, as well as from the elastic and dielectric constants. There is some redundancy in the model, however, as q and k are not independent.[81] In simulations, the shell can either be treated adiabatically (as in the iterative methods) or dynamically (as in the extended Lagrangian method). In the case where the shell is modeled dynamically, the spring constant k affects the characteristic frequency of the spring oscillations, and thus can be chosen from physical arguments or for numerical convenience.[82]

In the model described above, the core and shell charges have equal magnitudes, such that the polarizable atom remains electrically neutral. The original application of the shell model[72,73] and the majority of applications since then[77,83–94] have been to ionic systems. Charged species can easily be accommodated through the introduction of a permanent charge z_i coincident with the core (nuclear) charge (see Figure 2). This permanent charge then contributes to the static electric field experienced by the core and shell charges on other sites. The charge z_i can represent either an integer charge on a simple ion, or the effective partial charge on an atom in a molecular species.[95–100] Assigning charges this way is equivalent to allowing unequal core and shell charges, which is how the model is usually implemented in practice. Conceptually, however, it is useful to consider the permanent charge as a separate component of the model, so that the polarizable component is neutral, and thus has a dipole moment that is independent of the choice of origin.

We should remain cognizant of the fact that there is a conceptual difference between the polarizable point dipole models and the shell model. In the former, the point dipoles can be (and often are) assumed to be merely one term in an infinite series of multipoles that is used in a mathematical expansion of the electric field external to the molecule. In the shell model, on the other hand, the dipole moment is assumed to arise physically from the electron cloud's displacement from the molecular center. Because of the finite length of this dipole, it is important to specify whether the nonelectrostatic interaction centers are located at the cores (nuclei) or the shells (center of electron density). The nonelectrostatic interactions—including short-range repulsion (exchange) and van der Waals terms—are purely electronic in nature. Consequently, these interactions are typically taken to act between the shells, rather than the cores. The specific functional form used for the short-range interactions varies with the implementation, ranging from Buckingham or Born–Mayer potentials for ions to Lennard–Jones potentials for neutral species. Because a steep repulsive potential is an integral part of the shell model, polarization catastrophe is typically not an issue for these models.[91]

Several different methods exist for treating the motion of the polarizable degrees of freedom in dynamic simulations. As with the models based on point

dipoles, there are iterative, adiabatic techniques as well as fully dynamic methods. In the adiabatic methods, the correspondence between the shell charge and the effective electronic degrees of freedom is invoked, along with the Born–Oppenheimer approximation. In this case, the slow-moving nuclei and core charges are said to move adiabatically in the field generated by the shell charges. In other words, the positions of the shell charges are assumed to update instantaneously in response to the motion of the nuclei, and thus always occupy the positions in which they feel no net force (i.e., the positions that minimize the total energy of the system). The forces on the core charges are then used to propagate the dynamics, using standard numerical integration methods. The other alternative is to treat the charges fully dynamically, allowing them to occupy positions away from the minimum-energy position dictated by the nuclei, and thus experience nonzero forces.

When the charges are treated adiabatically, a self-consistent method must be used to solve for the shell displacements, $\{\mathbf{d}_i\}$ (just as with the dipoles $\{\boldsymbol{\mu}_i\}$ in the previous section). Combining Eqs. [26], [28], and [29], we can write the total energy of the shell model system as

$$U_{\text{ind}}(\{\mathbf{r}_i\}, \{\mathbf{d}_i\}) = \sum_{i=1}^{N} \left\{ \frac{1}{2} k_i d_i^2 + q_i [\mathbf{r}_i \cdot \mathbf{E}_i^0 - (\mathbf{r}_i + \mathbf{d}_i) \cdot \mathbf{E}_i^{0'}] \right.$$
$$\left. + \left[\frac{1}{2} \sum_{i=1}^{n} \sum_{j \neq i} \left(\frac{1}{r_{ij}} - \frac{1}{|\mathbf{r}_{ij} - \mathbf{d}_j|} - \frac{1}{|\mathbf{r}_{ij} + \mathbf{d}_i|} + \frac{1}{|\mathbf{r}_{ij} - \mathbf{d}_i + \mathbf{d}_j|} \right) \right] \right\}$$

[30]

which is the equivalent of Eq. [18] for a model with polarizable point dipoles, but with one important difference: Eq. [18] is a quadratic function of the $\{\boldsymbol{\mu}_i\}$, guaranteeing that its derivative (Eq. [19]) is linear and that a standard matrix method can be used to solve for the $\{\boldsymbol{\mu}_i\}$. Equation [30] is not a quadratic function of the $\{\mathbf{d}_i\}$. Moreover, the dependence of the short-range interactions on the displacements of the shell particles further complicates the matter. Consequently, matrix methods are typically not used to find the shell displacements that minimize the energy.

Iterative methods are used instead. In one such approach,[101] the nuclear (core) positions are updated, and the shell displacements from the previous step are used as the initial guess for the new shell displacements. The net force, $\mathbf{F}_i$, on the shell charge is calculated from the gradient of Eq. [30], together with any short-range interactions. Because the harmonic spring interaction is, by far, the fastest varying component of the potential felt by the shell charge, the incremental shell displacement $\delta \mathbf{d}_i = \mathbf{F}_i / k_i$ represents a very good estimate of the equilibrium (energy minimizing) position of the shells. The forces are recalculated at this position, and the procedure is iterated until a (nearly) force-free configuration is obtained. Alternatively, this steepest descent style

minimization can be replaced by more sophisticated minimization techniques,[102] such as conjugate gradients.[103] Depending on the convergence criterion used, these iterative methods typically require between 3 and 10 iterations.[77,99,101,103]

The dynamic approach to solving for the shell displacements was first proposed by Mitchell and Fincham.[90] In this method, the mass of each atom or ion is partitioned between the core and the shell. The mass of the shell charge is typically taken to be less than 10% of the total particle mass, and often as light as 0.2 amu.[82,90,97,104] No physical significance is attributed to the charge mass, as it is not meant to represent the mass of the electronic degrees of freedom whose polarization the shell charge represents. Rather, it is a parameter chosen solely for the numerical efficiency of the integration algorithm. Choosing a very light shell mass allows the shells (i.e., the dipole moments) to adjust very quickly in response to the electric field generated by the core (nuclear) degrees of freedom. In the limit of an infinitely light shell mass, the adiabatic limit would be recovered. The choice of shell mass also has a direct effect on the characteristic frequency of the oscillating shell model. For a particle with total mass M, a fraction f of which is attributed to the shell and $1 - f$ to the core, the reduced mass will be $\mu = f(1 - f)M$, resulting in an oscillation frequency of

$$\omega = \sqrt{\frac{k}{f(1 - f)M}} \qquad [31]$$

An overly small shell mass would thus result in high oscillation frequencies, requiring the use of an exceedingly small time step for integration of the dynamics—an undesirable situation for lengthy simulations. In practice, the shell mass is chosen to be (1) light enough to ensure adequate response times, (2) heavy enough that reasonable time steps may be used, and (3) away from resonance with any other oscillations in the system.

The dynamic treatment of the charges is quite similar to the extended Lagrangian approach for predicting the values of the polarizable point dipoles, as discussed in the previous section. One noteworthy difference between these approaches, however, is that the positions of the shell charges are ordinary physical degrees of freedom. Thus the Lagrangian does not have to be "extended" with fictitious masses and kinetic energies to encompass their dynamics.

With an appropriate partitioning of the particle masses between core and shell, this dynamic method for integrating the dynamics of the shell model can become more efficient than iterative methods. The lighter masses in the system require time steps 2–5 times smaller than those required in an iterative shell model simulation (or a nonpolarizable simulation).[82,90,97,99,104] But because the iterative methods require 3–10 force evaluations per time step to achieve

comparable energy conservation,[90,99,101,105] the dynamic methods can have a computational advantage, in some cases by as much as a factor of two to four.[90] Because the shell model represents each polarizable site with two point charges, it replaces the dipole–dipole interactions in the polarizable point dipole models with four charge–charge interactions. The greater number of pair distances largely offsets the computational advantage of the simpler interaction, and energy and force evaluations in the two methods are comparable in speed. Between the reduced time step and the greater number of interactions, shell models typically require 10 times more CPU time than corresponding nonpolarizable simulations.[105,106]

In the shell model, as mentioned above, the short-range repulsion and van der Waals interactions are taken to act between the shell particles. This finding has the effect of coupling the electrostatic and steric interactions in the system: in a solid-state system where the nuclei are fixed at the lattice positions, polarization can occur not only from the electric field generated by neighboring atoms, but also from the short-range interactions with close neighbors (as, e.g., in the case of defects, substitutions, or surfaces). This ability to model both electrical and mechanical polarizability is one reason for the success of shell models in solid-state ionic materials.[73,107]

There exist a variety of extensions of the basic shell model. One variation for molecular systems uses an anisotropic oscillator to couple the core and shell charges,[99,108] thus allowing for anisotropic polarizability in nonspherical systems. Other modifications of the basic shell model that account for explicit environment dependence include a deformable or "breathing" shell[75,76,109] and shell models allowing for charge transfer between neighboring sites.[75,76,110]

Shell models have been used successfully in a wide variety of systems. The greatest number of applications have been in the simulation of ionic materials,[86–88,90,111] especially systems including alkali halides,[83] oxides,[85,89,91,92,112–115] and zeolites.[93,94] The shell model is also commonly used for the simulation of molten salts,[77,84,90,101,116–120] and shell-type models have been developed for various molecular[95–99] and polymeric species.[100,121]

ELECTRONEGATIVITY EQUALIZATION MODELS

Polarizability can also be introduced into standard potentials (Eq. [1]) by allowing the values of the partial charges to respond to the electric field of their environment. A practical advantage of this approach is that it introduces polarizability without introducing new interactions. And unlike the shell model, this can be accomplished using the same number of charge–charge interactions as would be present in a nonpolarizable simulation. Another more conceptual advantage is that this treats the polarizable and permanent electrostatic interactions with the same multipoles. One way to couple the charges to their environment is by using electronegativity equalization.

The energy required to create a charge, q, on an atom can be expressed as a Taylor series expansion,

$$U(q) = E^0 + \chi^0 q + \frac{1}{2} J q^2 \qquad [32]$$

which has been truncated after the second-order terms. If the Taylor series is valid for charges of up to ± 1 e, then, because the ionization potential, IP, is equal to $U(1) - U(0)$ and the electron affinity, EA, is $U(-1) - U(0)$, the Taylor series coefficients are

$$\chi^0 = (\text{IP} + \text{EA})/2 \qquad [33]$$
$$J = \text{IP} - \text{EA} \qquad [34]$$

Equation [33] is Mulliken's definition of electronegativity,[122] so the linear coefficient in the Taylor series is the electronegativity of the atom. Mulliken's definition is consistent with other electronegativity scales. The second-order coefficient, $\frac{1}{2} J$, is the "hardness" of the atom, η.[123] For semiconductors, the hardness is half the band gap, and η is an important property in inorganic and acid–base chemistry.[124] Physically, IP $-$ EA is the energy required to transfer an electron from one atom to another atom of the same type,

$$2A(g) \rightarrow A^+(g) + A^-(g) \qquad \Delta E = \text{IP} - \text{EA} \qquad [35]$$

This energy is always positive (in fact, it is positive even if the two atoms are not the same element), so $J \geq 0$. Figure 3 shows $U(q)$ for chlorine and sodium, as calculated from the experimental IP and EA. The energies of the ions, χ^0, and J are all calculated using the experimental IP and EA. Chlorine is more electronegative than sodium ($\chi^0_{\text{Na}} = 2.84$ eV, $\chi^0_{\text{Cl}} = 8.29$ eV) and also harder ($J_{\text{Na}} = 4.59$ eV, $J_{\text{Cl}} = 9.35$ eV). This means that both the slope and the second derivative of $U(q)$ are larger for Cl than for Na.

When atoms are brought together to form molecules, the energy of the charges is described in the EE model as

$$U(\mathbf{q}) = \sum_i \left(E^0_i + \chi^0_i q_i + \frac{1}{2} J_{ii} q_i^2 \right) + \sum_i \sum_{j > i} J_{ij}(r_{ij}) q_i q_j \qquad [36]$$

The vector $\mathbf{q}$ represents the set of q_i. The second-order coefficient, $J_{ij}(r_{ij})$, depends on the distance between the two atoms i and j, and at large distances should equal $1/r_{ij}$. At shorter distances, there may be screening of the interactions, just as for the dipole–dipole interactions in the earlier section on Polarizable Point Dipoles. This screened interaction is typically assumed to arise
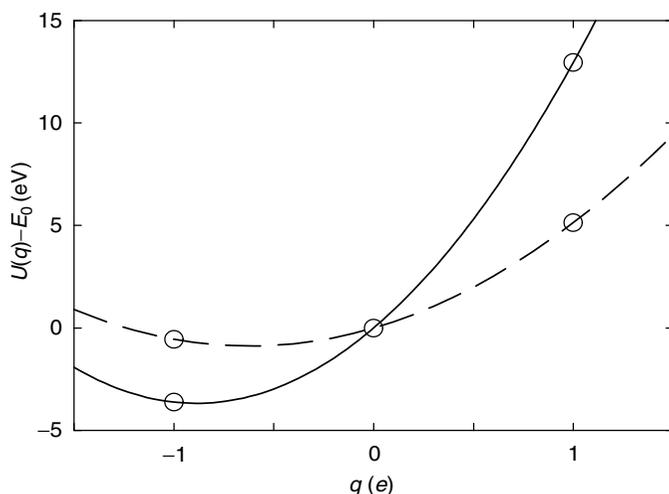
Figure 3 Energy versus charge for chlorine (solid line) and sodium (dashed line). The lines are a quadratic fit through the energies of the ions relative to the neutral atom.

from the Coulomb interaction between delocalized charge distributions $\rho(\mathbf{r})$, rather than point charges,

$$J_{ij}(\mathbf{r}) = \int \frac{\rho_i(\mathbf{r}_i)\rho_j(\mathbf{r}_j)}{|\mathbf{r}_i - \mathbf{r}_j - \mathbf{r}|} d\mathbf{r}_i \, d\mathbf{r}_j \qquad [37]$$

The charge distributions are frequently assumed to be spherical, for simplicity.[125–128] Directional interactions can be incorporated with nonspherical charge distributions, at some added computational expense.[129–131]

The partial charges on each atom of the molecule are found by minimizing the energy, subject to a constraint that the total charge is conserved.

$$\sum_i q_i = q_{\text{tot}} \qquad [38]$$

The charge conservation constraint can be enforced using an undetermined multiplier,

$$U(\mathbf{q}) = U(\mathbf{q}) - \lambda \left(\sum_i q_i - q_{\text{tot}} \right) \qquad [39]$$

Minimizing this expression for the energy with respect to each of the q_i under the assumption that the molecule in question is neutral ($q_{\text{tot}} = 0$) gives

for all i (i.e., $\forall\, i$):

$$\left(\frac{\partial U}{\partial q_i}\right) - \lambda = 0 \qquad \forall\, i \qquad [40]$$

Because $(\partial U/\partial q_i)$ for each atom is equal to the same undetermined multiplier λ, this quantity must be identical for all atoms in the molecule,[132]

$$\left(\frac{\partial U}{\partial q_i}\right) = \left(\frac{\partial U}{\partial q_j}\right) \qquad \forall\, i,j \qquad [41]$$

Through Mulliken's identification of $\partial U/\partial q$ as the electronegativity, we see that minimizing the energy with respect to the charges is equivalent to equalizing the electronegativities,

$$\chi_i \equiv \left(\frac{\partial U}{\partial q_i}\right) = \chi_i^0 + J_{ii}q_i + \sum_{j \neq i} J_{ij}(r_{ij})q_j \qquad [42]$$

for all atoms. Notice that the electronegativity of atom i in a molecule, χ_i, differs from the electronegativity of the isolated atom, χ_i^0, and now depends on its charge, the charge of the other atoms, its hardness, and the interactions with other atoms through $J_{ij}(r_{ij})$. In addition, Parr et al.[132] identified the chemical potential of an electron as the negative of the electronegativity, $\mu = -\partial U/\partial q$. So electronegativity equalization is equivalent to chemical potential equalization. Thus, this model effectively moves charge around a molecule to minimize the energy or to equalize the electronegativity or chemical potential. These interpretations are all equivalent (for a dissenting opinion, see Ref. 133).

Electronegativity equalization (EE) was first proposed by Sanderson.[134] The EE model, with appropriate parameterization, has been successful in predicting the charges of a variety of molecules.[125,135–138] The parameters χ^0 and J are not typically assigned from Eqs. [33] and [34], but instead are taken as parameters to be optimized and can be viewed as depending on the valence state of the atom, as indicated by electronic structure calculations.[139,140] Some models[136] set $J_{\alpha\beta}(r_{ij}) = 1/r_{ij}$, and others use some type of screening.[125,135,137] In addition, some models have an expression for the energy that is not quadratic.[125,135,141] Going beyond the quadratic term in the Taylor expansion of Eq. [32] can possibly extend the validity of the model, but it introduces complications in the methods available for treating the charge dynamics, as will be discussed below.

For a collection of molecules, the overall energy is comprised of the energy given by Eq. [36] for each molecule and an interaction between charge

sites on different molecules,

$$U(\{q\}, \{r\}) = \sum_{\alpha} \left(\sum_{i \in \alpha} \chi_i^0 q_i + \frac{1}{2} \sum_{i \in \alpha} \sum_{j \in \alpha} q_i q_j J_{ij}(r_{ij}) - E_{\alpha}^{\text{gp}} \right)$$
$$+ \sum_{\alpha} \sum_{\beta > \alpha} \sum_{i \in \alpha} \sum_{j \in \beta} q_i q_j J_{ij}(r_{ij}) \qquad [43]$$

where α and β label the molecules, and i and j represent atoms (or other charge sites) in these molecules. The E_{α}^{gp} term represents the gas-phase energy of molecule α and defines the zero of energy as corresponding to infinitely separated molecules. The energy given by Eq. [43] replaces the Coulomb energy $q_i q_j / r_{ij}$ in Eq. [1]. The charges q_i are now treated as independent variables, and the polarization response is determined by variations in the charge values. These charges depend on the interactions with other molecules as well as other charge sites on the same molecule, and will change for every time step or configuration sampled during a simulation. The charge values used for each configuration are, in principle, those that minimize the energy given by Eq. [43]. This method for treating polarizability has thus been called the *fluctuating charge* method[126] and has been applied to a variety of systems.[10,82,104,126,142–148] The $J_{ij}(r)$ interaction between different molecules is typically taken to be $1/r$, although the interactions between atoms on the same molecule may be screened. Therefore, this method does not modify the intermolecular interactions.

Charge conservation can be imposed in either of two ways. A charge neutrality constraint can be applied to the entire system, allowing charge to move from atomic site to atomic site until the electronegativities are equal on all the atoms of the system. Alternatively, charge can be constrained independently on each molecule (or other subgroup), so that charge flows only between atoms on the same molecule until the electronegativities are equalized within each molecule, but not between distinct molecules.[126] In most cases, charge is taken to be conserved for each molecule, so there is no charge transfer between molecules.

Variations, including the atom–atom charge transfer (AACT)[149] and the bond-charge increment (BCI)[146,150] model, only allow for charge to flow between two atoms that are directly bonded to each other, guaranteeing that the total charge of each set of bonded atoms is conserved. In some situations, charge transfer is an important part of the interaction energy, so there are reasons to remove this constraint.[151–154] However, this can lead to some nonphysical charge transfer, as illustrated in the simple example of a gas-phase sodium chloride molecule. The energy for one Na atom and one Cl atom is

$$U(q) = E_{\text{Na}}^0 + E_{\text{Cl}}^0 + (\chi_{\text{Na}} - \chi_{\text{Cl}}) q_{\text{Na}} + \frac{1}{2} \left[J_{\text{Na}} + J_{\text{Cl}} - 2 J_{\text{NaCl}}(r_{\text{NaCl}}) \right] q_{\text{Na}}^2 \quad [44]$$

where we have used $q_{Cl} = -q_{Na}$. The charge that minimizes this energy is

$$q_{Na} = \frac{-(\chi_{Na} - \chi_{Cl})}{J_{Na} + J_{Cl} - 2J_{NaCl}(r_{NaCl})} \quad [45]$$

At large distances, $J_{NaCl}(r_{NaCl})$ approaches zero, and, if the χ and J parameters are taken from Eqs. [33] and [34], then

$$q_{Na} = \frac{-(\chi_{Na} - \chi_{Cl})}{J_{Na} + J_{Cl}} = \frac{-\frac{1}{2}(IP_{Na} + EA_{Na} - IP_{Cl} - EA_{Cl})}{IP_{Na} - EA_{Na} + IP_{Cl} - EA_{Cl}} \quad [46]$$

which gives $q_{Na} = 0.391 \, e$. Thus the model predicts a significant amount of charge transfer, even at large distances. Similar errors in the dissociation limit are seen with certain electronic structure methods.[155,156] A significant amount of charge separation, and a consequent overestimation of the dipole moment, can be found for large polymers as well. Reducing this charge transfer along the polymer can be accomplished with the AACT and BCI models.[146,149,150] In addition, when comparing fluctuating charge models with ab initio results for water trimers, agreement was found to be much better for the model without charge transfer, even after the charge-transfer model was reparameterized by fitting to the ab initio three-body energies.[145]

These and associated problems with overestimated charge transfer are a general characteristic of EE-based models. Unfortunately, such errors cannot be eliminated through parameterization; the problem is a side effect of attempting to treat quantum mechanical charge-transfer effects in a purely classical way. As with all empirical potentials, the use of fractional charges is necessary for an accurate description of the electrostatic potential. Yet by allowing fractional charge transfer, the EE model has no means of enforcing the transfer of only an integral number of electrons between distant species. Indeed, the neutral dissociation products for NaCl are correctly predicted by the EE model, if the infinitely separated ions are constrained to have integer charge (see Figure 3). This constraint is difficult to apply in practice, however. As discussed recently by Morales and Martinez,[157] the EE-based models can be viewed as analytically differentiable approximations to a more rigorous statistical interpretation of $U(q)$, which is discontinuous at integer values of charge transfer and correctly predicts zero charge transfer at infinite distance. In chemically bonded systems, the assumption of partial charge transfer is not as unrealistic as in ionic compounds, as electrons are delocalized across covalent bonds. However, in these covalent cases the EE model effectively assumes that the coherence length of a delocalized electron is infinite and does not depend on the surroundings. It is for this reason that the polarizability of polymers, for example, increases too quickly with chain length under the EE model.[149] Molecular charge constraints can avoid problems at the dissociation

limit, and methods constraining the charge based on the bonding network are an extension of EE models that appear to be successful at controlling the coherence lengths.[149] There now exist classical models that can describe the charge transfer reasonably across the full range of a dissociating bond, but these are currently less well developed.[157]

The polarization energy in the EE models can be compared directly to that in the polarizable point dipole and shell models. Consider the first term in Eq. [43],

$$\sum_{i \in \alpha} \chi_i^0 q_i + \frac{1}{2} \sum_{i \in \alpha} \sum_{j \in \alpha} q_i q_j J_{ij}(r_{ij}) - E_\alpha^{\text{gp}} \qquad [47]$$

This term represents the energy required to induce charges q_i on the atoms of molecule α in the electric field of its neighbors, relative to the energy of the isolated molecule. This quantity is simply the polarization energy of the molecule. The polarization energy of the full system can thus be written

$$U_{\text{pol}} = \sum_\alpha \left[\sum_{i \in \alpha} \chi_i^0 q_i + \frac{1}{2} \sum_{i \in \alpha} \sum_{j \in \alpha} q_i q_j J_{ij}(r_{ij}) - E_\alpha^{\text{gp}} \right] \qquad [48]$$

which can be compared to Eqs. [9] and [26].[10]

There exist other models that treat polarizability using variable charges in a way similar to the fluctuating charge model.[22,53,127,143,158] In the Sprik and Klein[22] model for water, four charge sites are located near the oxygen atom in a tetrahedral geometry, in addition to the three atom-centered permanent charges. The tetrahedron of charges is used to represent an induced dipole moment on the oxygen center. This approach differs from a polarizable point dipole model in using a dipole of finite extent. It also differs from a shell model in that the point charges are fixed in the molecular frame. Consequently, the Sprik–Klein model should perhaps best be considered an entirely different type of model. The model of Zhu, Singh, and Robinson[158] is similar to the Sprik–Klein model, but it has no permanent charges. The four charge sites, two on hydrogen atoms and two on lone-pair positions 1 Å from the oxygen atom, are all variables coupled to the electric field. For both these models, the coupling is described by the polarizability, α, just as with other dipole polarizable models. Wilson and Madden[159] described a model for ions in which charge is transferred between ends of a rigid, rotating rod. In the model of Perng et al.,[143] the charge, q_i, on an atom is equal to a permanent value, q_i^0, plus an induced part, δq_i. The induced charge is dependent on the electrostatic potential at that site and all the induced charges are coupled through Coulombic interactions, similar to the fluctuating charge models. In the polarizable point charge (PPC) model of Svishchev et al.,[53] charges are coupled directly to the electric field at that site, so this model is slightly different from the fluctuating charge model.

Although a valence-type force field of the type illustrated by Eq. [1] is most suitable for modeling molecular systems, the electronegativity equalization approach to treating polarization can be coupled equally well to other types of potentials. Streitz and Mintmire[127] used an EE-based model in conjunction with an embedded atom method (EAM) potential to treat polarization effects in bulk metals and oxides. The resulting ES + EAM model has been parameterized for aluminum and titanium oxides, and has been used to study both charge-transfer effects and reactivity at interfaces.[127,128,160,161]

In most electronegativity equalization models, if the energy is quadratic in the charges (as in Eq. [36]), the minimization condition (Eq. [41]) leads to a coupled set of linear equations for the charges. As with the polarizable point dipole and shell models, solving for the charges can be done by matrix inversion, iteration, or extended Lagrangian methods.

As with other polarizable models, the matrix methods tend to be avoided by most researchers because of their computational expense. And when they are used, the matrix inversion is typically not performed at every step.[160,162] Some EE applications have relied on iterative methods to determine the charges.[53,127] For very large-scale systems, multilevel methods are available.[161,163] As with the dipole polarizable models, the proper treatment of long-range electrostatic interactions is especially important for fluctuating charge models.[164] Monte Carlo methods have also been developed for use with fluctuating charge models.[162,165] Despite this variety of available techniques, the most common approach is to use a matrix inversion or iterative method only to obtain the initial energy-minimizing charge distribution; an extended Lagrangian method is then used to propagate the charges dynamically in order to take advantage of its computational efficiency.

In the extended Lagrangian method, as applied to a fluctuating charge system,[126] the charges are given a fictitious mass, M_q, and evolved in time according to Newton's equation of motion, analogous to Eq. [23],

$$M_q \ddot{q}_i = -\frac{\partial U}{\partial q_i} - \lambda_\alpha \qquad [49]$$

where λ_α is the average of the negative of the electronegativity of the molecule α containing atom i,

$$\lambda_\alpha = -\frac{1}{N_\alpha} \sum_{i \in \alpha} \chi_i \equiv -\bar{\chi}_\alpha \qquad [50]$$

Here, N_α is the number of atoms in molecule α. Combining Eq. [49], [50], and [42], we have

$$M_q \ddot{q}_i = \bar{\chi}_\alpha - \chi_i \qquad [51]$$

In other words, the force experienced by each charge is proportional to the difference between the electronegativity at that site and the average electronegativity in the charge-constrained molecule that contains the charge site.

Equations [50] and [51] assume that the total charge of each molecule is conserved. They also assume that all of the charge masses are identical. If charge is allowed to transfer between molecules, then λ_α and $\bar{\chi}_\alpha$ are independent of the molecule, α, and are given by[126]

$$\lambda = \frac{1}{N_{\text{mol}}} \sum_{\alpha=1}^{N_{\text{mol}}} \frac{1}{N_\alpha} \sum_{i \in \alpha} \chi_i \equiv -\bar{\chi} \qquad [52]$$

A short trajectory of the fluctuating charge on a water molecule using the TIP4P-FQ model[126] comparing the extended Lagrangian model with the exact minimum energy value is shown in Figure 4. The extended Lagrangian values oscillate around the exact values, until near the end of the interval at which time the two trajectories begin to diverge from each other, due to the chaotic nature of the system. The charges also oscillate with small magnitude around the exact solution, demonstrating that they remain quite close to the true electronegativity equalizing (energy minimizing) values. The small oscillations also imply that the charges are at a much colder temperature (≈ 1 K in Figure 4) than the rest of the system. One drawback of the extended Lagrangian method is that it contains an additional parameter, the charge mass. This mass must be chosen to be small enough that the charges respond promptly to changes in the
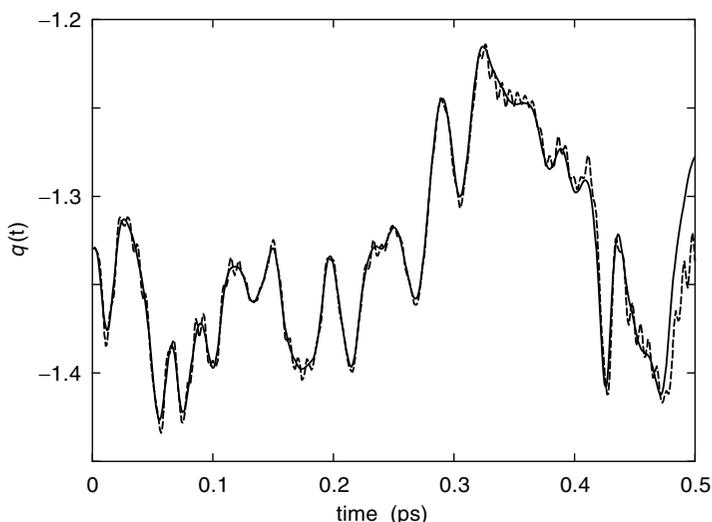

Figure 4 Negative charge near the oxygen atom versus time for the TIP4P-FQ water model, comparing the exact (solid line) and extended Lagrangian value (dashed line).

electronic potential (i.e., a large frequency for the oscillations about the exact trajectory in Figure 4), but large enough so that reasonable length time steps can still be used. In addition, the mass should be chosen so that the coupling between the charge and nuclear degrees of freedom is relatively weak. Any such coupling enables the cold charges to absorb energy from the rest of the system, eventually reaching equilibrium with the warmer parts of the system. Weak coupling results in relatively slow energy transfer, taking hundreds of picoseconds or longer before the charge temperature and amplitude of the charge oscillations become large enough to require reminimization. For many applications, standard 1 femtosecond time steps can be used, and the charges will remain at a temperature less than 6 K for a 50-ps simulation with-out thermostatting. Thus EE combined with the extended Lagrangian method is not much more computationally demanding than nonpolarizable simula-tions.[126,148] Finding the optimum masses can be difficult for systems with many different atom types, each fluctuating on a different time scale.[10] For these cases, different M_q must be used for the different charge masses. The expression for λ_α becomes

$$\lambda_\alpha = -\frac{\sum_i \chi_i / M_{q,i}}{\sum_i 1 / M_{q,i}} \qquad [53]$$

For more complex systems, thermostatting may be required to keep the charges near 0 K.[10]

Polarizable models based on EE implement the electrostatic interactions using either point or diffuse charges, and can thus be combined quite easily with other methods of treating polarizability to create hybrid models. The EE and the dipole polarizable models have some features in common, but they are not equivalent. They have, for example, different distance depen-dences and polarizability responses. Some hybrid models have included both dipole polarizability and fluctuating charges.[131,144,150,166] The fluctuating charge model has also been combined with a shell-type model, as a method of allowing polarization in single-atom species such as simple ions, without having to introduce the added complication of the dipole field tensor.[82,104,167]

The χ_i and J_{ij} parameters for the EE models can be optimized so that the resulting charges match gas-phase values as determined from either ab initio quantum mechanical calculations or the experimental dipole moment.[125,126,136-138,148] Parameters derived along these lines can give accu-rate gas-phase charge values. Information about many-body interactions can be included in the parameterization in several ways. First, quantities including ESP charges, geometries, and the strength of many-body interactions can be obtained from ab initio calculations on clusters.[142,145,150,166] Second, the polarization response from an applied electric field can be used.[146] Third, one can optimize the parameters to give the optimal charges both in the gas phase and in the presence of a solvent, as modeled using reaction field

methods.[10,168] Finally, the parameters themselves can be directly calculated using density functional theory (DFT) methods.[169,170]

As presented, the EE approach given by Eq. [43] is a simple mathematical model resulting from a Taylor series; it can be given a more rigorous foundation using electronic density functional theory.[169] Using DFT, and making simplifying approximations for the exchange and kinetic energy functionals, expressions analogous to Eq. [43] can be derived.[130,131,171] This approach has been termed chemical potential equalization (CPE).[130] Efforts like CPE or even parameterizations of fluctuating charge models using electronic structure calculations represent a step away from empirical potential models toward ab initio simulation methods. However, even with a sophisticated treatment of the charges, empirical terms in the potential such as the Lennard–Jones interaction still remain. A standard method is to set the Lennard–Jones parameters so that the energies and geometries of important dimer conformations (e.g., hydrogen-bonded dimers) are close to ab initio values.[10,144,145,166] In some cases, the remaining potential parameters have been taken from existing force fields.[146,150] One interesting extension of the fluctuating charge model has been developed by Siepmann and co-workers.[147] In their model, the Lennard–Jones size parameter becomes a variable that is coupled to the charge on a given atom. The size of the atom increases as the atom becomes more negatively charged and obtains greater electronic density. This increase in size is thus consistent with physical intuition. Other models in which some of the remaining potential parameters are treated as variables are described in the next section.

SEMIEMPIRICAL MODELS

A number of quantum polarizable models have been developed.[144,172–177] These treatments of polarizability represent a step toward full ab initio methods. The models can be characterized by a small number of electronic states or potential energy surfaces, which are coupled to each other. For the purposes of this tutorial, our description is of the method of Gao.[173,174] In his method, molecular orbitals, ϕ_A, for each molecule are defined as a linear combination of N_b atomic orbitals, χ_μ,

$$\phi_A = \sum_{\mu=1}^{N_b} c_{\mu A}\chi_\mu \qquad [54]$$

As is standard in semiempirical methods,[178] the molecular orbitals are orthonormal, so the overlap matrix, S_{AB}, is assumed to be diagonal. The molecular wave function, Ψ_a, is a Hartree[144] or Hartree–Fock[173] product of the molecular orbitals. For a (closed-shell) molecule with $2M$ electrons, there will be M doubly occupied molecular orbitals. The wave function of a system comprised

of N molecules is taken as a Hartree product of the individual molecular wave functions,

$$\Phi = \prod_{i=1}^{N} \Psi_i \qquad [55]$$

The Hartree product neglects exchange correlation interactions between molecules. To include proper exchange would make these models inefficient and impractical.

The Hamiltonian for the system

$$\hat{H} = \sum_{i=1}^{N} \hat{H}_i^0 + \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N} \hat{H}_{ij} \qquad [56]$$

contains the isolated molecular Hamiltonian, $\hat{H}_i^0$, given, in atomic units, by

$$\hat{H}_i^0 = \sum_{a=1}^{2M} \hat{T}_a - \sum_{\alpha=1}^{A}\sum_{a=1}^{2M} \frac{Z_\alpha(i)}{R_{\alpha a}} + \sum_{a=1}^{2M}\sum_{b>a} \frac{1}{r_{ab}} \qquad [57]$$

where $\hat{T}$ is the kinetic energy operator, $Z_\alpha(i)$ is the nuclear charge of atom α on molecule i, $R_{\alpha a}$ is the distance between the nucleus of atom α and electron a, and r_{ab} is the distance between two electrons. The interaction Hamiltonian between molecules i and j, $\hat{H}_{ij}$, is

$$\hat{H}_{ij} = \sum_{a=1}^{2M}\sum_{b=1}^{2M} \frac{1}{r_{ab}} + \sum_{\alpha=1}^{A}\sum_{\beta=1}^{A} \frac{Z_\alpha(i)Z_\beta(j)}{R_{\alpha\beta}} \qquad [58]$$

where $R_{\alpha\beta}$ is the distance between atom α on molecule i and atom β on molecule j. The interaction energy of the system is

$$E = \langle\Phi|\hat{H}|\Phi\rangle - N\langle\Phi^0|\hat{H}_i^0|\Phi^0\rangle \qquad [59]$$

where Φ^0 is the ground-state wave function of the isolated molecule and $\langle\Phi^0|\hat{H}_i^0|\Phi^0\rangle$ is the energy of the isolated molecule. The polarization energy is

$$E_{pol} = \sum_{i=1}^{N}(\langle\Phi|\hat{H}_i^0|\Phi\rangle - \langle\Phi^0|\hat{H}_i^0|\Phi^0\rangle) \qquad [60]$$

or, equivalently,

$$E_{pol} = \sum_{i=1}^{N}(\langle\Psi_i|\hat{H}_i^0|\Psi_i\rangle - \langle\Psi_i^0|\hat{H}_i^0|\Psi_i^0\rangle) \qquad [61]$$

which is the difference between the molecular energy of wave function Φ (or Ψ_i), which is the expectation value of the Hamiltonian in Eq. [56], and the molecular energy of the isolated molecule wave function. This expression for the polarization energy is comparable to Eqs. [9], [26], and [48] for the other models.

To avoid calculating the two-electron integrals in Eq. [59], the assumption is made that no electron density is transferred between molecules. The interaction Hamiltonian is then

$$\hat{H}_{ij}(\Psi_j) = -\sum_{a=1}^{2M} V_a(\Psi_j) + \sum_{\alpha=1}^{A} Z_\alpha(i) V_\alpha(\Psi_j) \qquad [62]$$

where $V_x(\Psi_j)$ is the electrostatic potential[179] from molecule j at the position of electron a or nuclei α of molecule i,

$$V_x(\Psi_j) = -\int \frac{\Psi_j^2(\mathbf{r})}{|\mathbf{r}_x - \mathbf{r}|} d\mathbf{r} + \sum_{\beta=1}^{A} \frac{Z_\beta(j)}{|\mathbf{r}_x - \mathbf{R}_\beta|} \qquad [63]$$

If the $V_x(\Psi_j)$ coming from the electrons and nuclei of molecule j is represented just by point charges on atomic sites, then

$$V_x(\Psi_j) = \sum_{\beta=1}^{A} \frac{q_\beta(\Psi_j)}{|\mathbf{r}_x - \mathbf{R}_\beta|} \qquad [64]$$

and

$$\hat{H}_{ij}(\Psi_j) = -\sum_{a=1}^{2M} \sum_{\beta=1}^{A} \frac{q_\beta(\Psi_j)}{r_{a\beta}} + \sum_{\alpha=1}^{A} \sum_{\beta=1}^{A} \frac{Z_\alpha(i) q_\beta(\Psi_j)}{R_{\alpha\beta}} \qquad [65]$$

where $q_\beta(\Psi_j)$ is the partial atomic charge on atom β in molecule j derivable from the wave function Ψ_j. (Other semiempirical models have charges offset from the atomic sites.)[172,175,176] The energy of molecule i is then changed by the partial charges from the other molecules. Since exchange correlation interactions are neglected as mentioned above in regard to Eq. [55], the short-range repulsive interactions need to be added, which can be done with a Lennard–Jones potential. The interaction energy between molecules i and j is then

$$E_{ij} = \frac{1}{2} \left(\langle \Psi_i | \hat{H}_{ij} | \Psi_i \rangle + \langle \Psi_j | \hat{H}_{ji} | \Psi_j \rangle \right) + E_{\mathrm{LJ}} \qquad [66]$$

which is used so that E_{ij} is equal to E_{ji}. The interactions between molecules then consist of only Lennard–Jones and Coulombic components. Polarizability

is treated using variable charges. The total energy is then

$$E = E_{\text{pol}} + \frac{1}{2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} E_{ij} \qquad [67]$$

and the forces on the nuclear coordinates are provided by the derivative of E with respect to the positions.

The charges can be found through Mulliken population analysis,[180] which, because the overlap matrix is diagonal, is

$$q_\alpha = K \left(Z_\alpha - 2 \sum_{a=1}^{M} \sum_{\mu} c_{\mu a}^2 \right) \qquad [68]$$

where the sum over μ is over atomic orbitals centered on atom α, and K is an empirical scaling parameter correcting for errors in the Mulliken charges (K is about 2). The Lennard–Jones parameters are assumed to be independent of the electronic states and all applications to date have been for rigid molecular geometries, so the models do not need to include nonbonded interactions.

The electronic structure of molecules can be described at the semiempirical level using, for example, the Austin model (AM1)[181] or at the ab initio level with a Gaussian basis set.[182] Other quantum theoretical methods can be used, however, as illustrated the method of Kim and co-workers[175,176] who use a ''truncated adiabatic basis'' consisting of the ground and first few excited states of the isolated molecule. For water, these methods introduce about 7–10 basis functions per molecule.[144,176] The wave function coefficients in these models are found using an iterative method.[144,172–176] An interesting variant of the empirical valence bond (EVB) approach has recently been introduced by Lefohn, Ovchinnikov and Voth.[177] In this approach, as applied to water, there are only three EVB states per molecule, and all potential parameters, rather than being derived from ab initio or semiempirical methods, are parameterized against experimental data.

Another method for treating polarizability is to have more than one potential surface with different electronic properties coupled together. This method is applicable to systems that can be represented by a few electronic states, like those with resonance. Each of these states can have its own potential energy parameters. One such model was developed for the peptide bond.[183] The peptide bond can be described as consisting of the resonance structures of two states, one with a N—C single bond and no formal charges and the other with a N=C double bond and formal charges on the nitrogen and oxygen. Each of these states is coupled to the environment, which in turn can shift the energies of the states. The potential parameters for these states can be different, but in the peptide bond model only the charges of the peptide group atoms and the dihedral force constant for rotations about

the peptide bond were taken to be state dependent. All other parameters were taken from existing force fields. Each peptide group, i, has a coefficient for each state, C_{iA} and C_{iB}, with the constraint that

$$C_{iA}^2 + C_{iB}^2 = 1 \qquad [69]$$

The charge for atom α that would go into the potential (e.g., Eq. [1]) is given by

$$q_{i\alpha} = C_{iA}^2 q_{\alpha A} + C_{iB}^2 q_{\alpha B} \qquad [70]$$

where $q_{\alpha X}$ is the charge of atom α for state X. Similarly, the dihedral force constant (for the $n = 2$ term) is given by

$$V_i = C_{iA}^2 V_A + C_{iB}^2 V_B \qquad [71]$$

The charges and dihedral force constants thus vary between the values for state A and the values for state B. This model provides a method for treating polarizability in which both the electrostatic parameters and the bonded parameters are coupled to the environment. It would be straightforward to couple the short-range potential to the electrostatic variables, like in the shell models and the fluctuating charge model of Siepmann and co-workers.[147] The two states are coupled with a term, $C_{iA}C_{iB}E_{AB}$. The coefficients for residue i are coupled to those of other residues through the Coulomb interactions. The coefficients are found by minimizing the energy, subject to the constraint of Eq. [69], and they are propagated using the extended Lagrangian method. Since the method treats the bonded parameters as variables too, it can also handle the amino group pyramidalization. In addition, the two-state empirical model enforces a charge conservation constraint on all peptide groups. Consequently, like the AACT and BCI electronegativity equalization models,[146,149,150] it will not overestimate the charge flow along the polymer.

One feature of the semiempirical models is that because the polarization is described by a set of coefficients that have a normalization condition, for example, Eq. [69], there will be no polarization catastrophe like there can be with dipole polarizable or fluctuating charge models. With a finite basis set, the polarization response is limited and can become only as large as the state with the largest dipole moment.

APPLICATIONS

Water

Water is the most common substance to be studied with polarizable potentials. An extremely large number of polarizable potentials for liquid

water have been developed, including those that treat the polarizability using polarizable point dipoles,[15,19–21,23–31,33,35,36,52,54,58,184] shell models,[97,99] charge-transfer models,[22,126,130,144,145,147,158,185,186] semiempirical models,[144,172,174,176,177] and hybrid methods.[166] The available literature on the simulation of water is extensive enough to deserve separate reviews.[187,188] Here, we concentrate primarily on general conclusions drawn from polarizable simulations of water.

Considerable latitude exists in choosing the nonelectrostatic features of a water model, including the functional form for the van der Waals interactions, the modeling of the intramolecular bonds and angles (flexible or rigid), and the inclusion or omission of an explicit hydrogen-bonding term. The electrostatic features of the model vary considerably as well. Although many polarizable models are constrained to reproduce the gas-phase dipole, the molecular polarizability, and sometimes the gas-phase quadrupole moment, these replications of the real data can all be accomplished in several ways with different placement of charge sites. Because of this freedom, as well as the facts that different experimental properties were used for the parameterization of the various models and different boundary conditions were used in the various simulations, it is difficult to compare different models on an equal footing. Nevertheless, the large variety of available water models does permit some general conclusions.

One of the principal purposes for using a polarizable model (of any type) is the ability to model a system under a variety of experimental conditions. For water models, a truly transferable model should cover the full range of states from gas phase to condensed phases, including ice, liquid water at ambient conditions, and even the supercritical fluid. It should also be capable of modeling heterogeneous environments by incorporating the varying polarization responses of water at interfaces,[189] around highly charged solutes, and in highly hydrophobic environments (as in the interior of proteins or lipid bilayers). Because water is in fact found under such a wide variety of conditions, and because of its anomalous properties, a fully transferable water model unfortunately remains a holy grail. Nonetheless, polarizable potentials have had considerable success in improving the transferability of water potentials in general.

Most nonpolarizable water models are actually fragile in this regard; they are not transferable to temperatures or densities far from where they were parameterized.[190] Because of the emphasis on transferability, polarizable models are typically held to a higher standard and are expected to reproduce monomer and dimer properties for which nonpolarizable liquid-state models are known to fail. Consequently, several of the early attempts at polarizable models were in fact less successful at ambient conditions than the benchmark nonpolarizable models, SPC[191] (simple point charge) and TIP4P[192] (transferable interaction potential, 4 points). Nonetheless, there is now a large collection of models that reproduce many properties of both the gas phase

(monomer and dimer geometry, dipole moment, and/or polarizability; second virial coefficient) and the bulk liquid (thermodynamic, structural, and dynamic properties).[30,36,52,53,126,166,185] The expectation is typically that such models will also be able to perform well at conditions intermediate between gas and liquid phases, such as clusters and interfaces. It is also assumed that a reasonably correct treatment of polarization will allow for some extrapolation beyond these conditions, so that systems where the electric field is not as homogeneous as in bulk water can be treated.

Even so, there are properties of small clusters and the bulk liquid that remain fairly elusive. For example, many models, both polarizable and nonpolarizable, do a poor job of reproducing the geometry of the water dimer. The methods typically predict a dimer that is too "flat", that is, with too small an angle between the donated O—H bond on the donor and the C_{2v} axis of the acceptor. This lack of tetrahedral coordination at the oxygen acceptor is usually attributed to the lack of lone pairs in the model; the electrostatic potential is insufficiently anisotropic on the oxygen end of the molecule when only atom-centered charges and dipoles are used. Models with off-atom charge sites,[54,166] higher order multipoles,[21,193] or explicitly anisotropic potentials[15,193] can be used to avoid this problem.

For gas-phase properties, the second virial coefficient, $B(T)$, provides one of the most sensitive tests of a water model.[186,194] Both polarizable and nonpolarizable models are capable of reproducing experimental values of $B(T)$, and some models have even been parameterized to do so explicitly.[15,24,29] Polarizable models appear to provide significant improvements in reproducing not only the second virial coefficient,[24,25] but also the third coefficient, $C(T)$.[186,195]

In the liquid phase, calculations of the pair correlation functions, dielectric constant, and diffusion constant have generated the most attention. There exist nonpolarizable and polarizable models that can reproduce each quantity individually; it is considerably more difficult to reproduce all quantities (together with the pressure and energy) simultaneously. In general, polarizable models have no distinct advantage in reproducing the structural and energetic properties of liquid water, but they allow for better treatment of dynamic properties.

It is now well understood that the static dielectric constant of liquid water is highly correlated with the mean dipole moment in the liquid, and that a dipole moment near 2.6 D is necessary to reproduce water's dielectric constant of $\varepsilon = 78$.[4,5,185,196] This holds for both polarizable and nonpolarizable models. Polarizable models, however, do a better job of modeling the frequency-dependent dielectric constant than do nonpolarizable models.[126] Certain features of the dielectric spectrum are inaccessible to nonpolarizable models, including a peak that depends on translation-induced polarization response, and an optical dielectric constant that differs from unity. The dipole moment of 2.6 D should be considered as an optimal value for typical (i.e.,

classical and rigid) water models; it is not necessarily the best estimate of the actual dipole moment. The dipole moment of liquid water cannot be measured experimentally, nor can it even be defined unambiguously, since the electronic density is not zero between molecules.[197,198] Ab initio simulations of liquid water predict that the average dipole moment varies from 2.4 to 3.0 D depending on how the density is partitioned, so a value of 2.6 D is consistent with these studies.[199–201]

Dynamic properties, such as the self-diffusion constant, are likewise strongly correlated with the dipole moment.[5,23] This coupling between the translational motion and the dipole moment is indicated in the dielectric spectrum.[126] Models that are overpolarized tend to undergo dynamics that are significantly slower than the real physical system. The inclusion of polarization can substantially affect the dynamics of a model, although the direction of the effect can vary. When a nonpolarizable model is reparameterized to include polarizability, the new model often exhibits faster dynamics, as with polarizable versions of TIP4P,[202] Reimers–Watts–Klein (RWK),[185,203] and reduced effective representation (RER)[30] potentials. There are exceptions, however, such as the polarizable simple point charge (PSPC)[23,57] and fluctuating charge (FQ)[126] models. The usual explanation for faster dynamics in polarizable models is that given by Sprik.[202] Events governing dynamical properties, such as translational diffusion and orientational relaxation, are activated processes—they depend on relatively infrequent barrier-crossing events. Adiabatic dynamics of the polarizable degrees of freedom allows for relaxation of the polarization at the transition, through means that are inaccessible to nonpolarizable models. This in turn lowers the activation barrier and increases the number of successful transition attempts. The nonunanimity of published simulation results concerning dynamic properties is likely due to such factors as: inconsistent parameterization procedures between the polarizable and nonpolarizable models; a strong dependence of dynamic properties on the system pressure (which is often insufficiently controlled during simulations); and the effects of using point versus diffuse charge distributions.

Transferability to different temperatures is a particularly difficult task for polarizable water models. This statement is illustrated by the problems in predicting the *PVT* and phase coexistence properties. There are a handful of polarizable water models—including both dipole- and EE-based models— that are reasonably successful in predicting some of the structural and energetic changes in liquid water over a range of several hundred degrees.[53,61,204] Many models fail to capture this behavior, however, so temperature transferability is far from an automatic feature of polarizable models.[35,52,61,62] Indeed, it has been demonstrated by several authors[35,52,61] that a point dipole-based model designed specifically to reproduce properties of the gas-phase monomer and the bulk liquid at 298 K is doomed to fail at higher temperatures. This failure could arise from insufficiencies in the Lennard–Jones function typically

used for the short-range repulsion, as well as from the use of point charges or dipoles rather than diffuse charge distributions. Evidence exists showing that diffuse charge distributions are necessary to ensure transferability, in both polarizable and nonpolarizable models.[35,53,205]

Predicting phase coexistence behavior near the critical point seems to be a particularly difficult task, even for the best polarizable models. Almost no existing model that works well at ambient conditions has been demonstrated to predict the critical temperature and density to better than 10% accuracy.[61,206] And those that are specifically designed to work well near the critical point seem to do a poor job of reproducing the liquid structure at lower temperatures.[61] Part of the problem is that the simulations required to measure phase coexistence properties are computationally expensive due to the extensive sampling required. Because of this expense, phase coexistence properties have not typically been included in the list of target properties when parameterizing new water models. Thus, it is only now becoming clear how to construct a model that is transferable across hundreds of degrees, from supercooled liquid to supercritical fluid. It is not yet clear whether one particular type of polarizable model is better able to capture the variation of water properties under varying temperatures and densities than another. However, the current situation clearly underscores the considerable flexibility and ambiguities involved in parameterizing polarizable potentials.

Transferability from the solid state to the liquid state is equally problematic. A truly transferable potential in this region of the phase diagram must reproduce not only the freezing point, but also the temperature of maximum density and the relative stability of the various phases of ice. This goal remains out of reach at present, and few existing models demonstrate acceptable transferability from solid to liquid phases.[33,52,207] One feature of water that has been demonstrated by both an EE model study[207] and an ab initio study[200] is that the dipole moments of the liquid and the solid are different, so polarization is likely to be important for an accurate reproduction of both phases. In addition, while many nonpolarizable water models exhibit a computed temperature of maximum density for the liquid, the temperature is not near the experimental value of 277 K.[53,62,208–215] For example, TIP4P[192] and SPC/E[4] models have a temperature of maximum density, T_{MD}, near 248 K.[211,213,215] Several EE models[53,147,207] and one EE–PPD hybrid model[166] yield a T_{MD} right at 277 K, suggesting that polarizability may be an important factor for this property as well. However, PPD models do not reproduce the T_{MD} maximum density very well; one model does not even have a T_{MD}[212] and another has a temperature dependence on the density that is much too strong.[62] One nonpolarizable model, the TIP5P model, which includes lone-pair interaction sites, has been successfully constructed to have the correct T_{MD}.[216]

The successful transferability of water models from the bulk phases to more heterogeneous conditions is another important goal for scientists

developing polarizable models. A vast literature exists in this area, with applications ranging from the solvation of simple ions[27,82,202,217,218] and biomolecules[10] to hydrophobic hydration and the structure of water at interfaces[104,219,220] and in external electric fields.[206,220] Due to the wide variation in electrostatic environments encountered, it is not surprising to find that polarizable models generally (but not always) provide significant improvements over nonpolarizable models.

Proteins and Nucleic Acids

For both proteins and nucleic acids, there exist significant structure-determining, hydrogen-bonding interactions between groups with π electrons: the peptide group for proteins and bases for nucleic acids. The extensive network of peptide hydrogen bonds in α-helices and β-sheets in proteins and the base-pair stacking in the double helix of nucleic acids are stablized by polarization of electrons with some π character. This stabilization has been labeled π-bond cooperativity or resonance-assisted hydrogen bonding.[221,222] The polarization of the π electrons in amides can be represented by the usual two dominant resonance structures
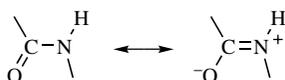
and in the nucleic acid bases (shown here for uracil),
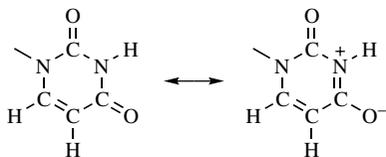
Resonance structures like these are commonly cited as leading to the planar geometry of the peptide bond and nucleic acid bases.

A number of quantum mechanical studies on the molecules *N*-methylactamide (NMA) and *N*-methylformamide (NMF), have addressed the importance of cooperative, or nonadditive, effects on hydrogen-bond formation.[223–225] Aggregates of NMA or NMF may be considered prototypes of the protein backbone. For these systems, the cooperative effects were found to add about 12–20% to the stabilization energy. Most of that energy can be decomposed into the polarization energy, with charge transfer making only a modest contribution, although the size of each component depends on the method of decomposition.[225] Experimental studies on NMA aggregates also

indicate cooperativity in the hydrogen-bond energies,[226] and dielectric measurements on polypeptide chains show an enhancement of the dipole moment of the peptide group in an α-helix.[227] Other quantum mechanical studies have addressed the importance of polarizability on protein folding,[228] enzyme catalysis,[229] DNA base pair stacking,[230] and nucleic acid interactions with ions.[231]

   Several polarizable models for proteins and the peptide group have been developed, using polarizable point dipoles,[32,44,45,232] electronegativity equalization models,[10,146] and the two-state empirical model.[183] Simulations using point polarizable dipole models by Warshel and co-workers[44,45] and by Wodak and co-workers[46] examined the role of polarizability on protein stability, dielectric properties, and enzymatic activity. For example, Van Belle et al.[46] found that the helix dipoles are enhanced, in agreement with the dielectric measurements of Wada,[227] and, further, the helix dipoles are enhanced not only through hydrogen bonds to the backbone, but also through association with side chain atoms. Polarization has also been shown to influence the folding time scales for small polypeptides.[183] For nucleic acids, a point polarizable dipole model was recently introduced.[232] Despite these studies and acknowledgment of the importance of polarizability from both electronic structure and experimental studies, not many simulations of proteins or nucleic acids using polarizable models have been done to date.

   An implication of resonance-assisted hydrogen bonding is that as the charges are polarized, through hydrogen bonds or other interactions, the hybridization of the atoms involved can change. For example, studies of crystal structures of formamide reveal that the C=O bond length increases and the C—N bond length decreases due to the formation of hydrogen bonded dimers.[233] Other crystal structures and ab initio quantum calculations on amides further validate the fact that hydrogen bonds can change those bond lengths.[234] The hydrogen bonds in these structures are in the amide plane and promote the double bond, zwitterionic state. On the other hand, the interactions in which the amino nitrogen serves as a hydrogen-bond acceptor would stabilize the single bond form. Partial sp^3 hybridization of the amino nitrogen leads to pyramidalization. Indeed, nonplanarities of some peptide bonds have been observed in atomic-resolution structures of proteins.[183,235,236] In addition, the planarity of the peptide bond is dependent on a protein's secondary structure, with residues in an α-helix being more planar than elsewhere.[183] For nucleic acids, ab initio calculations indicate that the amino group can be pyramidalized through interactions with neighboring molecules or ions.[237] For both the peptide bond and nucleic acid bases, there is reason to believe that a significant degree of nonplanarity can be induced by the environment. To treat these effects, the polarization of the electrostatic degrees of freedom—charges or dipoles—would have to be coupled to the bonded interactions, as has been developed for the peptide bond.[183]

COMPARISON OF THE POLARIZATION MODELS

Mechanical Polarization

One important difference between the shell model and polarizable point dipole models is in the former's ability to treat so-called mechanical polarization effects. In this context, mechanical polarization refers to any polarization of the electrostatic charges or dipoles that result from causes other than the electric field of neighboring atoms. In particular, mechanical interactions such as steric overlap with nearby molecules can induce polarization in the shell model, as further described below. These mechanical polarization effects are physically realistic and are quite important in some condensed-phase systems.

As mentioned earlier, the shell model is closely related to those based on polarizable point dipoles; in the limit of vanishingly small shell displacements, they are electrostatically equivalent. Important differences appear, however, when these electrostatic models are coupled to the nonelectrostatic components of a potential function. In particular, these interactions are the nonelectrostatic repulsion and van der Waals interactions—short-range interactions that are modeled collectively with a variety of functional forms. Point dipole- and EE-based models of molecular systems often use the Lennard–Jones potential. On the other hand, shell-based models frequently use the Buckingham or Born–Mayer potentials, especially when ionic systems are being modeled.

Regardless of the specific potential used, PPD- and EE-based models typically lack coupling between the short-range potential and the long-range electrostatic degrees of freedom. The dipoles and fluctuating charges respond solely to the local electric field (see Eq. [3]), with no regard for local short-range interactions. In other words, the polarizability, α, of each point dipole in a PPD model is independent of the local environment. The situation is different for the shell-based models. Because the van der Waals and exchange-repulsion interactions being modeled by the short-range nonelectrostatic part of the potential are electron–electron interactions, the interaction sites are almost always taken to be coincident with the shell (electronic) charge, rather than the core (nuclear) charge or center of mass. The short-range interactions in the shell model couple with only one end of the finite dipole, rather than with both "ends" of the point dipole. Consequently, the shell model includes a coupling between the short-range interactions and the orientation of the dipole—a coupling that is not present in point dipole-based models. The coupling of short-range interactions and dipole orientations is in fact quite realistic physically, and the lack of such a coupling is a disadvantage of the PPD models. One way to better understand this coupling is to recognize that the shell models have two mechanisms for polarization: a purely electrostatic induction effect, governed by the fixed polarizability in Eq. [27], and a

mechanical polarization effect that depends on the specific implementation of the dispersion and short-range repulsive interactions. Thus each polarizable site has an effective polarizability that depends on the local environment. When a shell-model atom is confined in a condensed phase, the steric interactions with neighboring ions will generally reduce the effective polarizability compared to the gas-phase value. In a crystalline environment, there are additional effects to consider: the anions and cations will polarize by different amounts in an applied electric field (due to the more diffuse electron density in the anions). The mechanical polarization effects will act to increase the effective polarizability in cations, and decrease it in anions.[73] These effects are completely realistic; the polarizabilities of atoms and ions do change with their environment in just these ways,[238–240] and shell models have at times been specifically parameterized to include this effect quantitatively.[73,96] Indeed, the inclusion of this mechanical polarizability effect has been shown to be crucial for reproducing condensed-phase properties such as phonon dispersion curves.[74,75]

Another coupling of the short-range repulsive and long-range electrostatic interactions has been developed by Chen, Xing, and Siepmann.[147] In their EE model, the repulsive part of the Lennard–Jones potential is coupled to the charge. This coupling is consistent with ab initio quantum calculations that find that the size of an atom increases with its negative charge.[241] Studies of gas–liquid[61] and solid–liquid[207] coexistence of water also suggest that models that couple the volume of an atom (through the Lennard–Jones interaction) to the size of the atom's charge may be best suited for prediction of molecular properties in the three phases. Empirical and semiempirical methods provide a natural way to link the charges to other parts of the potentials, as is done in the empirical valence bond approach[242] and the two-state peptide bond model.[183]

To further illustrate the importance of coupling the electrostatic and short-ranged repulsion interactions, we consider the example of a dimer of polarizable rare gas atoms, as presented by Jordan et al.[96] In the absence of an external electric field, a PPD model predicts that no induced dipoles exist (see Eq. [12]). But the shell model correctly predicts that the rare gas atoms polarize each other when displaced away from the minimum-energy (force-free) configuration. The dimer will have a positive quadrupole moment at large separations, due to the attraction of each electron cloud for the opposite nucleus, and a negative quadrupole at small separations, due to the exchange-correlation repulsion of the electron clouds. This result is in accord with ab initio quantum calculations on the system, and these calculations can even be used to help parameterize the model.[96]

In essence, this difference between shell models and PPD models arises from the former's treatment of the induced dipole as a dipole of finite length. Polarization in physical atoms results in a dipole moment of a small, but finite, extent. Approximating this dipole moment as an idealized point dipole, as in the PPD models, is an attractive mathematical approximation and produces

negligible errors in such properties as the electric field generated outside the molecule. Unfortunately, there are some physical effects that this idealization obscures, such as the environment-dependent polarizability.

All polarizable models share the ability to polarize, by varying their charge distribution in response to their environment. In addition, shell models and EE models with charge-dependent radii have the ability to modify their polarizability—the magnitude of this polarization response—in response to their local environment. Consequently, it is reasonable to expect that shell models and mechanically coupled EE models may be slightly more transferable to different environments than more standard PPD and EE models. To date, it is not clear whether this expectation has been fully achieved. Although some shell-based models for both ionic and molecular compounds have been demonstrated to be transferable across several phases and wide ranges of phase points,[73,96,99,243] it is not clear that the transferability displayed by these models is better than that demonstrated in PPD- or EE-based models. And even with an environment-dependent polarizability, it has been demonstrated that the basic shell model cannot fully capture all of the variations in ionic polarizabilities in different crystal environments.[85]

Computational Efficiency

One significant difference between the different methods of incorporating polarization is their computational efficiency. For energy evaluations, the electronegativity equalization-based methods are considerably more efficient than the dipole or shell models. Dipole-based methods require evaluation of the relatively CPU-expensive dipole–dipole interactions (Eq. [7]). The charge–charge interactions used in shell models are much cheaper, by about a factor of three. But this advantage is eliminated by the need to represent each polarizable center by two point charges, thus quadrupling the total number of interactions that need to be computed. Methods based on electronegativity equalization typically represent each polarizable site by a single charge (either point or diffuse), and energy evaluations are thus three-to-four times faster than with the other models, for direct summation. Semiempirical methods have 4–10 basis functions per atom, and each energy evaluation requires solving large matrices, thereby decreasing the computational efficiency of these models.[144,172–176] In the simpler two-state empirical model, the additional computational requirements are comparable to the EE models.[183]

Energy evaluation for any collection of point charges and dipoles can be accelerated significantly by using fast-multipole[244,245] or particle-mesh[246,247] methods. The computational advantages of these methods are proportionally much greater for the dipole-based models, because they avoid the direct evaluation of a more expensive interaction. In large systems, the overhead associated with using dipoles can be reduced to about a third more than the cost of using point charges. Algorithms for performing conventional,[66]

fast-multipole,[69] and particle-mesh Ewald[50] summation on point dipoles are available, and even quite efficient, but are considerably more complex than the comparable methods for monopole charges.[244,247–249]

Regardless of the type of model used, a method must be chosen for the self-consistent solution of the polarizable degrees of freedom. Direct solution via matrix inversion is nearly always avoided by most researchers in the field, because of the prohibitive $O(N^3)$ scaling with system size, N. Both iterative and predictive methods reduce the scaling to match that of the potential evaluation [$O(N^2)$ for direct summation; $O(N \ln N)$ for Ewald-based methods;[50,68] $O(N)$ if interactions are neglected beyond some distance cutoff], but the cost of the iterations means that the predictive methods are always more efficient. Extended Lagrangian methods have been implemented for all four types of polarizable potential.[10,22,56–58,82,90,97,99,104,126,148,183] The extended Lagrangian methods are least popular for PPD-based models; as a general rule, simulations with these models still tend to use iterative methods. The extended Lagrangian approach is perhaps most natural for the shell model, for which it is physically reasonable to assign a mass to the polarizable degrees of freedom (the shell charges) and treat them dynamically. However, the small mass of the shell charge usually requires an MD time step smaller than would be chosen in a nonpolarizable simulation.[82,90,97,99,104] The fluctuating charge and PPD models usually do not require a reduction in time step, thus making them somewhat more efficient in this regard.

Multiple time step methods[250,251] can also be used to reduce the computational cost of simulations with polarizable models. Such methods have been used successfully with shell and fluctuating charge models.[82,104] However, it is more problematic to apply these multiple time scale integrators in simulations using iterative integrators. The multiple time scale integrators work by calculating updated values for only a fraction of the system's interactions during some of the time steps; but since all of the interactions are needed in order to provide well-converged values for the polarizable degrees of freedom, the bulk of the expensive electrostatic interactions must still be evaluated at every step.

Hyperpolarizability

Note that linearly polarizable point dipoles provide only an approximation to the true polarization response in two different ways. First, polarization can include terms that are nonlinear in the electric field. Thus, Eq. [3] represents only the first term in an infinite series,

$$\mu = \alpha \cdot \mathbf{E} + \frac{1}{2}\mathbf{E} \cdot \beta \cdot \mathbf{E} + \cdots \qquad [72]$$

where β is a third-rank tensor representing the first hyperpolarizability of the system.[14] In water, for example, the nonlinear polarization effects begin to

become significant at field strengths[252,253] of 2–3 V/Å, which is comparable to the mean field strength in an aqueous solution.[53] This finding indicates that perhaps there are improvements to be made by going beyond the approximation of linear polarization. Only occasional attempts have been made to include these effects.[55]

Charge-Transfer Effects

The EE-based and semiempirical models implement polarization via charge transfer between atoms on the same molecule. These models are fundamentally different from the treatment of shell and PPD models, which include point polarization but no charge transfer. There are important differences between the two approaches.

As pointed out in the section on Electronegativity Equalization Models, the implementation of charge transfer in current EE models tends to lead to overpolarization in large molecules or when intermolecular charge transfer is allowed. In contrast, the lack of charge transfer in point-polarizable models can sometimes lead to underpolarization. In general, the point-polarizable models predict that the polarizability of a single molecule or a system of molecules will increase linearly with its size, in proportion to the number of (linearly polarizable and weakly interacting) dipoles.[254] This behavior is exactly correct for systems without charge transfer, such as saturated hydrocarbon molecules and most biomolecules. The PPD models severely underpredict, however, the increase of polarization with system size for conductive systems such as unsaturated hydrocarbons. An EE-based model does significantly better at predicting the size-dependent polarization of conductive systems, but exaggerates the polarization in large systems with no charge transfer.[149] Thus we emphasize that it is important to choose the method of treating polarization that is most appropriate for the system being studied. Hybrid models containing both point-polarizable and charge-transfer sites are perhaps the most flexible approach.[145,146,150,166]

Another side effect of the EE and semiempirical models' reliance on charge transfer for treating polarization is a geometry dependence that is absent in point-polarizable models. The charge redistribution in an EE model can arise only as a result of charge transfer from one site to another. Consequently, the polarization response is constrained by the geometry of the charge sites. This constraint is most severe for highly symmetric species. For planar molecules such as benzene and water, the EE model unrealistically predicts that the out-of-plane component of the polarizability tensor is zero. Linear molecules cannot be polarized in the transverse direction. Atomic or ionic species suffer the most dramatic limitation: they have no polarization response at all under the EE approximation. Whereas this can be a severe limitation in some circumstances, an EE model for water with purely planar polarizability somewhat surprisingly performs as well as or better than PPD and shell models

with three-dimensional polarizability tensors.[126] Off-atom charge sites have been successfully used to address this limitation in some cases, as have hybrid models.[82,96,104,145,146,150,166]

The Electrostatic Potential

In addition to treating the polarization response in different ways, the various methods considered here also provide different levels of approximation to the external electric field. Accurate simulation of intermolecular interactions requires that the electrostatic potential be correctly represented everywhere outside the molecular surface. The correct electrostatic potentials can be reproduced, of course, by the physically correct nuclear and electronic charge distribution. At points outside the molecular surface, however, it can also be reproduced to arbitrary accuracy by a series of point monopoles, dipoles, quadrupoles, and so on. This approach is taken in most computer simulations. The simplest level of approximation is to include point charges (monopoles) at the atomic sites. The accuracy of this approximation can be improved by (1) adding more charge sites (off-atom sites); (2) increasing the number of terms in the series (dipoles, quadrupoles, etc.); and (3) by replacing the point multipoles with delocalized, diffuse charge distributions.

The PPD and shell models are nearly equivalent in this sense, because they model the electrostatic potential via static point charges and polarizable dipoles (of either zero or very small extent). Accuracy can be improved by extending the expansion to include polarizable quadrupoles or higher order terms.[193] The added computational expense and difficulty in parameterizing these higher order methods has prevented them from being used widely. The accuracy of the ESP for dipole-based methods can also be improved by adding off-atom dipolar sites.[96,166]

Because the EE-based methods truncate the series representation of the electrostatic potential one term earlier (i.e., by using only monopole charges), these methods would appear to sacrifice some accuracy in representing the electrostatic potential. It is becoming widely appreciated that models based solely on point charges may require the use of off-atom charge sites to successfully fit the electrostatic potential.[166,255] However, nearly all polarizable simulation methods based on charge-transfer methods have used some sort of delocalized charges, rather than point charges.[22,125,126,130,146,158,171] This approach has been shown to be successful at reproducing the electrostatic potential for most extramolecular sites, although the use of point dipoles can improve the performance for certain conformations (such as bifurcated hydrogen bonds) in which molecular symmetries prevent accurate charge distributions.[146] Indeed, it has been claimed that the better representation of intermolecular interactions due to diffuse charges is as important as the use of polarizability.[205] The chemical potential equalization (CPE) methods are

noteworthy in this regard because they use both diffuse monopoles and dipoles to represent the system's polarization.[130,131,171]

The difference between models having polarizable point dipoles and fixed point charges and those with fluctuating charges and fixed Lennard–Jones interactions reduces to considering which term is static and which is polarizable. For the PPD model, the charge–charge term is static and the induced dipole–induced dipole term is polarizable. For the EE model, the charge–charge term is polarizable and the induced dipole–induced dipole terms (included in the Lennard–Jones r^{-6} interaction) are static. Note that including a Lennard–Jones r^{-6} dispersion term is not redundant for polarizable models because this represents the interaction arising from correlated thermal fluctuations of the induced dipole. With a few exceptions,[22,57,202] most models—whether based on matrix, iterative, or extended Lagrangian algorithms—are adiabatic and do not allow for substantial fluctuations away from the minimum-energy polarization state.

SUMMARY AND CONCLUSIONS

There are a variety of different models used to treat polarizability in molecular simulations: polarizable point dipoles, shell models, fluctuating charge models, and semiempirical models, along with variations and combinations of these. There are advantages and disadvantages of each model, as discussed in detail in previous sections. These relative merits range from differing computational efficiencies and ease of implementation to different accuracies in representing the external electric field and transferability of parameters. Regardless of the differences in convenience and efficiency, the most important consideration when choosing a polarizable model for a particular problem should be the model's applicability to the system in question.

Despite the many differences between the various polarizable models, it is encouraging to note that the most recent models seem to be converging on the same set of necessary features. A variety of successful models based on different formalisms all share many of the same characteristics.[126,130,131,146,150,166,171,205] Regardless of the direction from which the models evolved, there is a growing consensus that accurate treatment of polarization requires (1) either diffuse charge distributions or some other type of electrostatic screening (2) a mixture of both monopoles and dipoles to represent the electrostatic charge distribution, and (3) only linear polarizability.

Although much work remains to be done before there is a truly accurate, transferable model for a wide range of conditions and systems, it is fair to say that polarizable models have matured considerably since their earliest implementations. Future developments will almost certainly include continued development and parameterization of the more mainstream models, along with their incorporation into commercial and academic simulation software

packages, thereby making these methods much more accessible to the nonspecialist. In particular, we expect polarizable models, and especially polarizable water models, to become more prevalent in biomolecular simulations involving heterogeneous solvent environments. Inclusion of polarizability in the potentials for proteins and other macromolecular systems is also likely to become more common, and hence a careful assessment of the importance of polarizability to these systems is needed. Until the importance of polarizability has been clearly demonstrated, the added computational cost of modeling the polarization makes it is unlikely that polarizable models will displace more traditional models for the bulk of routine simulation, particularly when applied to large systems.

Future directions in the development of polarizable models and simulation algorithms are sure to include the combination of classical or semiempirical polarizable models with fully quantum mechanical simulations, and with empirical reactive potentials. The increasingly frequent application of Car–Parrinello ab initio simulations methods[156] may also influence the development of potential models by providing additional data for the validation of models, perhaps most importantly in terms of the importance of various interactions (e.g., polarizability, charge transfer, partially covalent hydrogen bonds, lone-pair-type interactions). It is also likely that we will see continued work toward better coupling of charge-transfer models (i.e., EE and semiempirical models) with purely local models of polarization (polarizable dipole and shell models).

REFERENCES

1. J. A. McCammon and S. C. Harvey, *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge, UK, 1987.

2. M. Rigby, E. B. Smith, W. A. Wakeham, and G. C. Maitland, *The Forces Between Molecules*, Oxford Science Publications, Oxford, UK, 1986.

3. T. A. Halgren, *J. Am. Chem. Soc.*, **114**, 7827–7843 (1992). Representation of van der Waals (vdW) Interactions in Molecular Mechanics Force Fields: Potential Form, Combination Rules, and vdW Parameters.

4. H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, *J. Phys. Chem.*, **91**, 6269–6271 (1987). The Missing Term in Effective Pair Potentials.

5. K. Watanabe and M. L. Klein, *Chem. Phys.*, **131**, 157–167 (1989). Effective Pair Potentials and the Properties of Water.

6. L. Kuyper, D. Ashton, K. M. Merz Jr., and P. A. Kollman, *J. Phys. Chem.*, **95**, 6661–6666 (1991). Free Energy Calculations on the Relative Solvation Free Energies of Benzene, Anisole, and 1,2,3-Trimethoxybenzene: Theoretical and Experimental Analysis of Aromatic Methoxy Solvation.

7. W. L. Jorgensen and J. Gao, *J. Am. Chem. Soc.*, **110**, 4212–4216 (1988). Cis–Trans Energy Difference for the Peptide Bond in the Gas Phase and in Aqueous Solution.

8. D. E. Williams, *Biopolymers*, **29**, 1367–1386 (1990). Alanyl Dipeptide Potential-Derived Net Atomic Charges and Bond Dipoles, and Their Variation with Molecular Conformation.

9. P. Cieplak and P. Kollman, *J. Comput. Chem.*, **12**, 1232–1236 (1991). On the Use of Electrostatic Potential Derived Charges in Molecular Mechanics Force Field. The Relative Solvation Free Energy of *Cis-* and *Trans-N-Methyl-Acetamide.*

10. S. W. Rick and B. J. Berne, *J. Am. Chem. Soc.*, **118**, 672–679 (1996). Dynamical Fluctuating Charge Force Fields: The Aqueous Solvation of Amides.

11. J. J. Urban and G. R. Famini, *J. Comput. Chem.*, **14**, 353–362 (1993). Conformational Dependence of the Electrostatic Potential-Derived Charges of Dopamine: Ramifications in Molecular Mechanics Force Field Calculation in the Gas Phase and Aqueous Solution.

12. U. Dinur and A. T. Hagler, *J. Comput. Chem.*, **16**, 154–170 (1995). Geometry-Dependent Atomic Charges—Methodology and Application to Alkanes, Aldehydes, Ketones, and Amides.

13. U. Koch, P. L. A. Popelier, and A. J. Stone, *Chem. Phys. Lett.*, **238**, 253–260 (1995). Conformational Dependence of Atomic Multipole Moments.

14. C. J. F. Böttcher, *Theory of Electric Polarization,* 2nd ed., Elsevier, New York, 1973.

15. T. P. Lybrand and P. A. Kollman, *J. Chem. Phys.*, **83**, 2923–2933 (1985). Water–Water and Water–Ion Potential Functions Including Terms for Many Body Effects.

16. P. Cieplak, T. P. Lybrand, and P. A. Kollman, *J. Chem. Phys.*, **86**, 6393–6403 (1987). Calculations of Free Energy Changes in Ion–Water Clusters Using Nonadditive Potentials and the Monte Carlo Method.

17. B. J. Berne and A. Wallqvist, *J. Chem. Phys.*, **88**, 8016–8017 (1988). Comment on: "Water–Water and Water–Ion Potential Functions Including Terms for Many-Body Effects," T. P. Lybrand and P. Kollman, *J. Chem. Phys.* **83**, 2923 (1985), and on "Calculation of Free Energy Changes in Ion–Water Clusters Using Nonadditive Potentials and the Monte Carlo Method," P. Cieplak, T. P. Lybrand, and P. Kollman, *J. Chem. Phys.* **86**, 6393 (1987).

18. P. Kollman, T. Lybrand, and P. Cieplak, *J. Chem. Phys.*, **88**, 8017 (1988). Reply to "Comment on "Water–Water and Water–Ion Potential Functions Including Terms for Many-Body Effects," T. P. Lybrand and P. Kollman, *J. Chem. Phys.* **83**, 2923 (1985), and on "Calculation of Free Energy Changes in Ion–Water Clusters Using Nonadditive Potentials and the Monte Carlo Method," P. Cieplak, T. P. Lybrand, and P. Kollman, *J. Chem. Phys.* **86**, 6393 (1987)".

19. J. A. C. Rullman and P. T. van Duijnen, *Mol. Phys.*, **63**, 451–475 (1988). A Polarizable Water Model for Calculation of Hydration Energies.

20. F. H. Stillinger and C. W. David, *J. Chem. Phys.*, **69**, 1473–1484 (1978). Polarization Model for Water and Its Ionic Dissociation Products.

21. P. Barnes, J. L. Finney, J. D. Nicholas, and J. E. Quinn, *Nature (London)*, **282**, 459–464 (1979). Cooperative Effects in Simulated Water.

22. M. Sprik and M. L. Klein, *J. Chem. Phys.*, **89**, 7556–7560 (1988). A Polarizable Model for Water Using Distributed Charge Sites.

23. P. Ahlström, A. Wallqvist, S. Engström, and B. Jönsson, *Mol. Phys.*, **68**, 563–581 (1989). A Molecular Dynamics Study of Polarizable Water.

24. P. Cieplak, P. A. Kollman, and T. P. Lybrand, *J. Chem. Phys.*, **92**, 6755–6760 (1990). A New Water Potential Including Polarization: Application to Gas-Phase, Liquid, and Crystal Properties of Water.

25. U. Niesar, G. Corongiu, E. Clementi, G. R. Kneller, and D. K. Bhattacharya, *J. Phys. Chem.*, **94**, 7949–7956 (1990). Molecular Dynamics Simulations of Liquid Water Using the NCC Ab Initio Potential.

26. S. Kuwajima and A. Warshel, *J. Phys. Chem.*, **94**, 460–466 (1990). Incorporating Electric Polarizabilities in Water–Water Interaction Potentials.

27. J. Caldwell, L. X. Dang, and P. A. Kollman, *J. Am. Chem. Soc.*, **112**, 9144–9147 (1990). Implementation of Nonadditive Intermolecular Potentials by Use of Molecular Dynamics: Development of a Water–Water Potential and Water–Ion Cluster Interactions.

28. L. X. Dang, *J. Chem. Phys.*, **97**, 2659–2660 (1992). The Nonadditive Intermolecular Potential for Water Revised.

29. R. E. Kozack and P. C. Jordan, *J. Chem. Phys.,* **96,** 3120–3130 (1992). Polarizability Effects on a Four-Charge Model for Water.

30. A. Wallqvist and B. J. Berne, *J. Phys. Chem.,* **97,** 13841–13851 (1993). Effective Potentials for Liquid Water Using Polarizable and Nonpolarizable Models.

31. D. N. Bernardo, Y. Ding, K. Krogh-Jespersen, and R. M. Levy, *J. Phys. Chem.,* **98,** 4180–4187 (1994). An Anisotropic Polarizable Water Model: Incorporation of All-Atom Polarizabilities into Molecular Mechanics Force Fields.

32. J. W. Caldwell and P. A. Kollman, *J. Phys. Chem.,* **99,** 6208–6219 (1995). Structure and Properties of Neat Liquids Using Nonadditive Molecular Dynamics: Water, Methanol, and *N*-Methylacetamide.

33. J. Brodholt, M. Sampoli, and R. Vallauri, *Mol. Phys.,* **85,** 81–90 (1995). Parameterizing Polarizable Intermolecular Potentials for Water with the Ice 1h Phase.

34. T. Chang, K. A. Peterson, and L. X. Dang, *J. Chem. Phys.,* **103,** 7502–7513 (1995). Molecular Dynamics Simulations of Liquid, Interface, and Ionic Solvation of Polarizable Carbon Tetrachloride.

35. A. A. Chialvo and P. T. Cummings, *J. Chem. Phys.,* **105,** 8274–8281 (1996). Engineering a Simple Polarizable Model for the Molecular Simulation of Water Applicable over Wide Ranges of State Conditions.

36. L. X. Dang and T. Chang, *J. Chem. Phys.,* **106,** 8149–8159 (1997). Molecular Dynamics Study of Water Clusters, Liquid, and Liquid–Vapor Interface of Water with Many-Body Potentials.

37. L. Ojamäe, I. Shavitt, and S. J. Singer, *J. Chem. Phys.,* **109,** 5547–5564 (1998). Potential Models for Simulations of the Solvated Proton in Water.

38. Y. Ding, D. N. Bernardo, K. Krogh-Jespersen, and R. M. Levy, *J. Phys. Chem.,* **99,** 11575–11583 (1995). Solvation Free Energies of Small Amides and Amines from Molecular Dynamics/Free Energy Perturbation Simulations Using Pairwise Additive and Many-Body Polarizable Potentials.

39. J. Applequist, J. R. Carl, and K.-K. Fung, *J. Am. Chem. Soc.,* **94,** 2952–2960 (1972). An Atom Dipole Interaction Model for Molecular Polarizability. Application to Polyatomic Molecules and Determination of Atom Polarizabilities.

40. J. Applequist, *Acc. Chem. Res.,* **10,** 79–85 (1977). An Atom Dipole Interaction Model for Molecular Optical Properties.

41. B. T. Thole, *Chem. Phys.,* **59,** 341–350 (1981). Molecular Polarizabilities Calculated with a Modified Dipole Interaction.

42. R. C. Weast, Ed., *CRC Handbook of Chemistry and Physics,* CRC Press, Boca Raton, FL, Vol. 66, 1985.

43. F. H. Stillinger, *J. Chem. Phys.,* **71,** 1647–1651 (1979). Dynamics and Ensemble Averages for the Polarization Models of Molecular Interactions.

44. A. Warshel and M. Levitt, *J. Mol. Biol.,* **103,** 227–249 (1976). Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme.

45. S. T. Russell and A.Warshel, *J. Mol. Biol.,* **185,** 389–404 (1985). Calculations of Electrostatic Energies in Proteins. The Energetics of Ionized Groups in Bovine Pancreatic Trypsin Inhibitor.

46. D. Van Belle, I. Couplet, M. Prevost, and S. J. Wodak, *J. Mol. Biol.,* **198,** 721–735 (1987). Calculations of Electrostatic Properties in Proteins. Analysis of Contributions from Induced Protein Dipoles.

47. 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,** 5179–5197 (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules.

48. E. L. Pollock and B. J. Alder, *Phys. Rev. Lett.*, **39,** 299–302 (1977). Effective Field of a Dipole in Polarizable Fluids.

49. F. J. Vesely, *J. Comput. Phys.*, **24,** 361–371 (1977). *N*-Particle Dynamics of Polarizable Stockmayer-Type Molecules.

50. A. Toukmaji, C. Sagui, J. Board, and T. Darden, *J. Chem. Phys.*, **113,** 10913–10927 (2000). Efficient Particle-Mesh Ewald Based Approach to Fixed and Induced Dipolar Interactions.

51. R. Kutteh and J. B. Nicholas, *Comput. Phys. Commun.*, **86,** 227–235 (1995). Efficient Dipole Iteration in Polarizable Charged Systems Using the Cell Multipole Method and Application to Polarizable Water.

52. J. Brodholt, M. Sampoli, and R. Vallauri, *Mol. Phys.*, **86,** 149–158 (1995). Parameterizing a Polarizable Intermolecular Potential for Water.

53. I. M. Svishchev, P. G. Kusalik, P. G. Wang, and R. J. Boyd, *J. Chem. Phys.*, **105,** 4742–4750 (1996). Polarizable Point-Charge Model for Water: Results Under Normal and Extreme Conditions.

54. A. Wallqvist, P. Ahlström, and G. Karlstrom, *J. Phys. Chem.*, **94,** 1649–1656 (1990). A New Intermolecular Energy Calculation Scheme: Applications to Potential Surface and Liquid Properties of Water.

55. G. Ruocco and M. Sampoli, *Mol. Phys.*, **82,** 875–886 (1994). Computer Simulation of Polarizable Fluids: A Consistent and Fast Way for Dealing with Polarizability and Hyperpolarizability.

56. M.-L. Saboungi, A. Rahman, J. W. Halley, and M. Blander, *J. Chem. Phys.,* **88,** 5818–5823 (1988). Molecular Dynamics Studies of Complexing in Binary Molten Salts with Polarizable Anions: MAX_4.

57. D. van Belle, M. F. G. Lippens, and S. J. Wodak, *Mol. Phys.*, **77,** 239–255 (1992). Molecular Dynamics Simulation of Polarizable Water by an Extended Lagrangian Method.

58. J. W. Halley, J. R. Rustad, and A. Rahman, *J. Chem. Phys.*, **98,** 4110–4119 (1993). A Polarizable, Dissociating Molecular Dynamics Model for Liquid Water.

59. J. M. Goodfellow, *Proc. Natl. Acad. Sci. USA,* **79,** 4977–4979 (1982). Cooperative Effects in Water-Biomolecule Crystal Systems.

60. B. J. Costo, J. L. Rivail, and B. Bigot, *J. Chem. Phys.*, **86,** 1467–1473 (1987). A Monte Carlo Simulation Study of a Polarizable Liquid: Influence of the Electrostatic Induction on Its Thermodynamic and Structural Properties.

61. K. Kiyohara, K. E. Gubbins, and A. Z. Panagiotopoulos, *Mol. Phys.*, **94,** 803–808 (1998). Phase Coexistence Properties of Polarizable Water Models.

62. P. Jedlovszky and R. Vallauri, *Mol. Phys.*, **97,** 1157–1163 (1999). Temperature Dependence of Thermodynamic Properties of a Polarizable Potential Model of Water.

63. M. W. Mahoney and W. L. Jorgensen, *J. Chem. Phys.*, **114,** 9337–9349 (2001). Rapid Estimation of Electronic Degrees of Freedom in Monte Carlo Calculations for Polarizable Models of Liquid Water.

64. M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids,* Oxford University Press, Oxford, UK, 1987.

65. D. Frenkel and B. Smit, *Understanding Molecular Simulation: from Algorithms to Applications,* Academic Press, San Diego, 1996.

66. W. Smith, *Information Quarterly for Computer Simulation of Condensed Phases,* **4,** 13–25 (1982). Point Multipoles in the Ewald Summation.

67. T. M. Nymand and P. Linse, *J. Chem. Phys.*, **112,** 6386–6395 (2000). Molecular Dynamics Simulations of Polarizable Water at Different Boundary Conditions.

68. D. Fincham, *Information Quarterly for Computer Simulation of Condensed Phases,* **38,** 17–24 (1993). Optimization of the Ewald Sum.

69. R. Kutteh and J. B. Nicholas, *Comput. Phys. Commun.*, **86,** 236–254 (1995). Implementing the Cell Multipole Method for Dipolar and Charged Dipolar Systems.

70. M. Born and T. von Kármán, *Phys. Z.*, **13**, 297–309 (1912). Vibrations in Space Gratings (Molecular Frequencies).

71. M. Born and K. Huang, *Dynamical Theory of Crystal Lattices*, Oxford University Press, Oxford, UK, 1954.

72. B. G. Dick and A. W. Overhauser, *Phys. Rev.*, **112**, 90 (1958). Theory of the Dielectric Constants of Alkali Halide Crystals.

73. J. E. Hanlon and A. W. Lawson, *Phys. Rev.*, **113**, 472–478 (1959). Effective Ionic Charge in Alkali Halides.

74. R. A. Cowley, W. Cochran, B. N. Brockhouse, and A. D. B. Woods, *Phys. Rev.*, **131**, 1030–1039 (1963). Lattice Dynamics of Alkali Halide Crystals. III. Theoretical.

75. W. Cochran, *CRC Crit. Rev. Solid State Sci.*, **2**, 1–44 (1971). Lattice Dynamics of Ionic and Covalent Crystals.

76. A. N. Basu, D. Roy, and S. Sengupta, *Phys. Stat. Sol. A*, **23**, 11–32 (1974). Polarisable Models for Ionic Crystals and the Effective Many-Body Interaction.

77. M. J. L. Sangster and M. Dixon, *Adv. Phys.*, **25**, 247–342 (1976). Interionic Potentials in Alkali Halides and Their Use in Simulations of the Molten Salts.

78. J. Shanker and S. Dixit, *Phys. Status Solidi A*, **123**, 17–50 (1991). Dielectric Constants and Their Pressure and Temperature Derivatives for Ionic Crystals.

79. P. Drude, *The Theory of Optics*, Longmans, Green, and Co., New York, 1901.

80. F. London, *Trans. Faraday Soc.*, **33**, 8–26 (1937). The General Theory of Molecular Forces.

81. W. Cochran, in *Phonons in Perfect Lattices and in Lattices with Point Imperfections*, R. W. H. Stevenson, Ed., Plenum Press, New York, 1966, Vol. 6 of Scottish Universities' Summer School, Chapter 2, pp. 53–72. Theory of Phonon Dispersion Curves.

82. S. J. Stuart and B. J. Berne, *J. Phys. Chem.*, **100**, 11934–11943 (1996). Effects of Polarizability on the Hydration of the Chloride Ion.

83. C. R. A. Catlow, K. M. Diler, and M. J. Norgett, *J. Phys. C*, **10**, 1395–1412 (1977). Interionic Potentials for Alkali Halides.

84. M. Dixon and M. J. Gillan, *Philos. Mag. B*, **43**, 1099–1112 (1981). Structure of Molten Alkali Chlorides I. A Molecular Dynamics Study.

85. G. V. Lewis and C. R. A. Catlow, *J. Phys. C.*, **18**, 1149–1161 (1985). Potential Models for Ionic Oxides.

86. A. M. Stoneham and J. H. Harding, *Annu. Rev. Phys. Chem.*, **37**, 53 (1986). Interatomic Potentials in Solid State Chemistry.

87. C. R. A. Catlow and G. D. Price, *Nature (London)*, **347**, 243–248 (1990). Computer Modelling of Solid-State Inorganic Materials.

88. J. H. Harding, *Rep. Progr. Phys.*, **53**, 1403–1466 (1990). Computer Simulation of Defects in Ionic Solids.

89. S. M. Tomlinson, C. R. A. Catlow, and J. H. Harding, *J. Phys. Chem. Solids*, **51**, 477–506 (1990). Computer Modelling of the Defect Structure of Non-Stoichiometric Binary Transition Metal Oxides.

90. P. J. Mitchell and D. Fincham, *J. Phys.: Condens. Matter*, **5**, 1031–1038 (1993). Shell Model Simulations by Adiabatic Dynamics.

91. P. J. D. Lindan, *Mol. Simul.*, **14**, 303–312 (1995). Dynamics with the Shell Model.

92. M. S. Islam, *J. Mater. Chem.*, **10**, 1027–1038 (2000). Ionic Transport in ABO(3) Perovskite Oxides: A Computer Modelling Tour.

93. J.-R. Hill, A. R. Minihan, E. Wimmer, and C. J. Adams, *Phys. Chem. Chem. Phys.*, **2**, 4255–4264 (2000). Framework Dynamics Including Computer Simulations of the Water Adsorption Isotherm of Zeolite Na-MAP. See also J.-R. Hill, C. M. Freeman, and L. Subramanian, in *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., Wiley-VCH, New York, 2000, Vol. 16, pp. 141–216. Use of Force Fields in Materials Modeling. The shell model is also discussed by B. van de Graaf, S. L. Njo, and K. S. Smirnov, in *Reviews in*

Computational Chemistry, K. B. Lipkowitz and D. B. Boyd, Eds., Wiley-VCH, New York, 2000, Vol. 14, pp. 137–223. Introduction to Zeolite Modeling.

94. J. Sauer and M. Sierka, *J. Comput. Chem.*, **21**, 1470–1493 (2000). Combining Quantum Mechanics and Interatomic Potential Functions in Ab Initio Studies of Extended Systems.

95. H. Saint-Martin, C. Medina-Llanos, and I. Ortega-Blake, *J. Chem. Phys.*, **93**, 6448–6452 (1990). Nonadditivity in an Analytical Intermolecular Potential: The Water–Water Interaction.

96. P. C. Jordan, P. J. van Maaren, J. Mavri, D. van der Spoel, and H. J. C. Berendsen, *J. Chem. Phys.*, **103**, 2272–2285 (1995). Towards Phase-Transferable Potential Functions: Methodology and Application to Nitrogen.

97. N. H. de Leeuw and S. C. Parker, *Phys. Rev. B*, **58**, 13901–13908 (1998). Molecular-Dynamics Simulation of MgO Surfaces in Liquid Water Using a Shell-Model Potential for Water.

98. H. Saint-Martin, J. Hernández-Cobos, M. I. Bernal-Uruchurtu, I. Ortega-Blake, and H. J. C. Berendsen, *J. Chem. Phys.*, **113**, 10899–10912 (2000). A Mobile Charge Densities in Harmonic Oscillators (MCDHO) Molecular Model for Numerical Simulations: The Water–Water Interaction.

99. P. J. van Maaren and D. van der Spoel, *J. Phys. Chem. B*, **105**, 2618–2626 (2001). Molecular Dynamics Simulations of Water with Novel Shell-Model Potentials.

100. N. Karasawa and W. A. Goddard, *Macromolecules*, **25**, 7268–7281 (1992). Force-Fields, Structures, and Properties of Poly(vinylidene Fluoride) Crystals.

101. M. Dixon and M. J. L. Sangster, *J. Phys. C: Solid State Phys.*, **8**, L8–L11 (1975). Simulation of Molten NaI Including Polarization Effects.

102. T. Schlick, in *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1992, Vol. 3, pp. 1–71. Optimization Methods in Computational Chemistry.

103. P. J. D. Lindan and M. J. Gillan, *J. Phys.: Condens. Matter*, **5**, 1019–1030 (1993). Shell-Model Molecular Dynamics Simulation of Superionic Conduction in CaF_2.

104. S. J. Stuart and B. J. Berne, *J. Phys. Chem. A*, **103**, 10300–10307 (1999). Surface Curvature Effects in the Aqueous Ionic Solvation of the Chloride Ion.

105. K. F. O'Sullivan and P. A. Madden, *J. Phys.: Condens. Matter*, **3**, 8751–8756 (1991). Light Scattering by Alkali Halides Melts: A Comparison of Shell-Model and Rigid-Ion Computer Simulation Results.

106. J. A. Board and R. J. Elliott, *J. Phys.: Condens. Matter*, **1**, 2427–2440 (1989). Shell Model Molecular Dynamics Calculations of the Raman spectra of Molten NaI.

107. W. Cochran, *Philos. Mag.*, **4**, 1082–1086 (1959). Lattice Dynamics of Alkali Halides.

108. J. Cao and B. J. Berne, *J. Chem. Phys.*, **99**, 2213–2220 (1993). Theory of Polarizable Liquid Crystals: Optical Birefringence.

109. U. Schröder, *Solid State Commun.*, **4**, 347–349 (1966). A New Model for Lattice Dynamics ("Breathing Shell Model").

110. M. P. Verma and R. K. Singh, *Phys. Status Solidi*, **33**, 769 (1969). The Contribution of Three-Body Overlap Forces to the Dynamical Matrix of Alkali Halides.

111. D. K. Fisler, J. D. Gale, and R. T. Cygan, *Am. Mineral.*, **85**, 217–224 (2000). A Shell Model for the Simulation of Rhombohedral Carbonate Minerals and Their Point Defects.

112. R. Ferneyhough, D. Fincham, G. D. Price, and M. J. Gillan, *Model. Simul. Mater. Sci. Eng.*, **2**, 1101–1110 (1994). The Melting of MgO Studied by Molecular-Dynamics Simulation.

113. P. J. D. Lindan and M. J. Gillan, *Philos. Mag. B*, **69**, 535–548 (1994). The Dynamical Simulation of Superionic UO_2 Using Shell-Model Potentials.

114. C. Rambaut, H. Jobic, H. Jaffrezic, J. Kohanoff, and S. Fayeulle, *J. Phys.: Condens. Matter*, **10**, 4221–4229 (1998). Molecular Dynamics Simulation of the α-Al_2O_3 Lattice: Dynamic Properties.

115. M. Baudin and K. Hermansson, *Surf. Sci.*, **474**, 107–113 (2001). Metal Oxide Surface Dynamics from Molecular Dynamics Simulations: The α-Al_2O_3(0001) Surface.

116. G. Jacucci, I. R. McDonald, and A. Rahman, *Phys. Rev. A*, **13**, 1581–1592 (1976). Effects of Polarization on Equilibrium and Dynamic Properties of Ionic Systems.

117. M. Dixon and M. J. L. Sangster, *J. Phys. C: Solid State Phys.*, **8**, 909–925 (1976). Effects of Polarization on Some Static and Dynamic Properties of Molten NaI.

118. M. Dixon, *Philos. Mag. B*, **47**, 509–530 (1983). Molecular Dynamics Studies of Molten NaI. I. Quasi-Elastic Neutron Scattering.

119. M. Dixon, *Philos. Mag. B*, **47**, 531–554 (1983). Molecular Dynamics Studies of Molten NaI. II. Mass-, Charge- and Number-Density Fluctuations.

120. M. Dixon, *Philos. Mag. B*, **48**, 13–29 (1983). Molecular Dynamics Studies of Molten NaI. III. Longitudinal and Transverse Currents.

121. J. D. Carbeck, D. J. Lacks, and G. C. Rutledge, *J. Chem. Phys.*, **103**, 10347–10355 (1995). A Model of Crystal Polarization in β-Poly(Vinylidene Fluoride).

122. R. S. Mulliken, *J. Chem. Phys.*, **2**, 782–793 (1934). A New Electronegativity Scale: Together with Data on Valence States and an Ionization Potential and Electron Affinities.

123. R. G. Parr and R. G. Pearson, *J. Am. Chem. Soc.*, **105**, 7512–7516 (1983). Absolute Hardness: Companion Parameter to Absolute Electronegativity.

124. R. G. Pearson, *Inorg. Chem.*, **27**, 734–740 (1988). Absolute Electronegativity and Hardness: Application to Inorganic Chemistry.

125. A. K. Rappé and W. A. Goddard III, *J. Phys. Chem.*, **95**, 3358–3363 (1991). Charge Equilibration for Molecular Dynamics Simulations.

126. S. W. Rick, S. J. Stuart, and B. J. Berne, *J. Chem. Phys.*, **101**, 6141–6156 (1994). Dynamical Fluctuating Charge Force Fields: Application to Liquid Water.

127. F. H. Streitz and J. W. Mintmire, *Phys. Rev. B*, **50**, 11996–12003 (1994). Electrostatic Potentials for Metal-Oxide Surfaces and Interfaces.

128. S. Ogata, H. Iyetomi, K. Tsuruta, F. Shimojo, R. K. Kalia, A. Nakano, and P. Vashista, *J. Appl. Phys.*, **86**, 3036–3041 (1999). Variable-Charge Interatomic Potentials for Molecular-Dynamics Simulations of TiO_2.

129. M. Berkowitz, *J. Am. Chem. Soc.*, **109**, 4823–4825 (1987). Density Functional Approach to Frontier Controlled Reactions.

130. D. M. York and W. Yang, *J. Chem. Phys.*, **104**, 159–172 (1996). A Chemical Potential Equalization Method for Molecular Simulations.

131. C. Bret, M. J. Field, and L. Hemmingsen, *Mol. Phys.*, **98**, 751–763 (2000). A Chemical Potential Equilization Model for Treating Polarization in Molecular Mechanics Force Fields.

132. R. G. Parr, R. A. Donelly, M. Levy, and W. E. Palke, *J. Chem. Phys.*, **68**, 3801–3807 (1978). Electronegativity: The Density Functional Viewpoint.

133. L. C. Allen, *Acc. Chem. Res.*, **23**, 175–176 (1990). Electronegativity Scales.

134. R. T. Sanderson, *Science*, **114**, 670–672 (1951). An Interpretation of Bond Lengths and a Classification of Bonds.

135. W. J. Mortier, K. V. Genechten, and J. Gasteiger, *J. Am. Chem. Soc.*, **107**, 829–835 (1985). Electronegativity Equalization: Application and Parameterization.

136. W. J. Mortier, S. K. Ghosh, and S. Shankar, *J. Am. Chem. Soc.*, **108**, 4315–4320 (1986). Electronegativity Equalization: Method for the Calculation of Atomic Charges in Molecules.

137. K. T. No, J. A. Grant, and H. A. Scheraga, *J. Phys. Chem.*, **94**, 4732–4739 (1990). Determination of Net Ionic Charges Using a Modified Partial Equalization of Orbital Electronegativity Method. 1. Application to Neutral Molecules as Models for Polypeptides.

138. K. T. No, J. A. Grant, M. S. Jkon, and H. A. Scheraga, *J. Phys. Chem.*, **94**, 4740–4746 (1990). Determination of Net Atomic Charges Using a Modified Equalization of Orbital Electronegativity Method. 2. Application to Ionic an Aromatic Molecules as Models for Polypeptides.

139. J. Hinze and H. H. Jaffe, *J. Am. Chem. Soc.*, **84**, 540–546 (1962). Electronegativity. I. Orbital Electronegativity of Neutral Atoms.

140. U. Dinur, *J. Mol. Struct. (THEOCHEM)*, **303**, 227–237 (1994). A Relationship Between the Molecular Polarizability, Molecular Dipole Moment and Atomic Electronegativities in AB and AB_n Molecules.

141. In Ref. 125, the hardness, J, and $J_{\alpha\beta}$ ($r_{\alpha\beta}$) are both characterized by a Slater exponent parameter, ζ. For hydrogen atoms, this parameter is taken to depend on its charge ($\zeta = \zeta_0 + q_H$), introducing nonlinearities in the electronegativities. However, rather than using the Mulliken definition of χ as $\frac{\partial U}{\partial q}$, Rappé and Goddard use the expression for χ from Eq. [42] and thereby ignore the nonlinear terms. This means that the charges in this model do not minimize the energy. This fact is not clear from Ref. 125, but it is necessary to use Eq. [42] to reproduce their results.

142. H. Toufar, B. G. Baekelandt, G. O. A. Janssens, W. J. Mortier, and R. A. Schoonheydt, *J. Phys. Chem.*, **99**, 13876–13885 (1995). Investigation of Supramolecular Systems by a Combination of the Electronegativity Equalization Method and a Monte Carlo Simulation Technique.

143. B.-C. Perng, M. D. Newton, F. O. Raineri, and H. L. Friedman, *J. Chem. Phys.*, **104**, 7153–7176 (1996). Energetics of Charge Transfer Reactions in Solvents of Dipolar and Higher Order Multiplier Character. I. Theory.

144. M. J. Field, *Mol. Phys.*, **91**, 835–845 (1997). Hybrid Quantum Mechanical/Molecular Mechanical Fluctuating Charge Models for Condensed Phase Simulations.

145. Y.-P. Liu, K. Kim, B. J. Berne, R. A. Friesner, and S. W. Rick, *J. Chem. Phys.*, **108**, 4739 (1998). Constructing *Ab Initio* Force Fields for Molecular Dynamics Simulations.

146. J. L. Banks, G. A. Kaminski, R. Zhou, D. T. Mainz, B. J. Berne, and R. A. Friesner, *J. Chem. Phys.*, **110**, 741–754 (1999). Parametrizing a Polarizable Force Field from *Ab Initio* Data. I. The Fluctuating Point Charge Model.

147. B. Chen, J. Xing, and J. I. Siepmann, *J. Phys. Chem. B*, **104**, 2391–2401 (2000). Development of Polarizable Water Force Fields for Phase Equilibria Calculations.

148. M. C. C. Ribeiro and L. C. J. Almeida, *J. Chem. Phys.*, **110**, 11445–11448 (1999). Fluctuating Charge Model for Polyatomic Ionic Systems: A Test Case with Diatomic Anions.

149. R. Chelli, P. Procacci, R. Righini, and S. Califano, *J. Chem. Phys.*, **111**, 8569–8575 (1999). Electrical Response in Chemical Potential Equilization Schemes.

150. H. A. Stern, G. A. Kaminski, J. L. Banks, R. Zhou, B. J. Berne, and R. A. Friesner, *J. Phys. Chem. B*, **103**, 4730–4737 (1999). Fluctuating Charge, Polarizable Dipole, and Combined Models: Parameterization from Ab Initio Quantum Chemistry.

151. K. Kitaura and K. Morokuma, *Int. J. Quantum Chem.*, **10**, 325–340 (1976). A New Energy Decomposition Scheme for Molecular Interactions within the Hartee–Fock Approximation.

152. F. Weinhold, *J. Mol. Struct.*, **399**, 181–197 (1997). Nature of H-Bonding in Clusters, Liquids, and Enzymes: An Ab Initio, Natural Bond Orbital Perspective.

153. A. van der Vaart and K. M. Merz Jr., *J. Am. Chem. Soc.*, **121**, 9182–9190 (1999). The Role of Polarization and Charge Transfer in the Solvation of Biomolecules.

154. J. Korchowiec and T. Uchimaru, *J. Chem. Phys.*, **112**, 1623–1633 (2000). New Energy Partitioning Scheme Based on the Self-Consistent Charge and Configuration Method for Subsystems: Application to Water Dimer System.

155. J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz Jr., *Phys. Rev. Lett.*, **49**, 1691–1694 (1982). Density-Functional Theory for Fractional Particle Number: Derivative Discontinuities of the Energy.

156. R. Car and M. Parrinello, *Phys. Rev. Lett.*, **55**, 2471–2474 (1985). Unified Approach for Molecular Dynamics and Density-Functional Theory.

157. J. Morales and T. J. Martinez, *J. Chem. Phys.*, **105**, 2842–2850 (2001). Classical Fluctuating Charge Theories: An Entropy Valence Bond Formalism and Relationships to Previous Models.

158. S.-B. Zhu, S. Singh, and G. W. Robinson, *J. Chem. Phys.*, **95**, 2791–2799 (1991). A New Flexible/Polarizable Water Model.

159. M. Wilson and P. A. Madden, *J. Phys.: Condens. Matter,* **5**, 2687–2706 (1993). Polarization Effects in Ionic Systems from First Principles.

160. T. Campbell, R. K. Kalia, A. Nakano, P. Vashista, S. Ogata, and S. Rodgers, *Phys. Rev. Lett.,* **82**, 4866–4869 (1999). Dynamics of Oxidation of Aluminum Nanoclusters Using Variable Charge Molecular-Dynamics Simulation on Parallel Computers.

161. D. J. Keffer and J. W. Mintmire, *Int. J. Quantum Chem.*, **80**, 733–742 (2000). Efficient Parallel Algorithms for Molecular Dynamics Simulations Using Variable Charge Transfer Electrostatic Potentials.

162. M. Medeiros and M. E. Costas, *J. Chem. Phys.*, **107**, 2012–2019 (1997). Gibbs Ensemble Monte Carlo Simulation of the Properties of Water with a Fluctuating Charges Model.

163. A. Nakano, *Comput. Phys. Commun.*, **104**, 59–69 (1997). Parallel Multilevel Preconditioned Conjugate-Gradient Approach to Variable-Charge Molecular Dynamics.

164. S. W. Rick, in *Simulation and Theory of Electrostatic Interactions in Solution,* L. R. Pratt and G. Hummer, Eds., American Institute of Physics, Melville, NY, 1999, pp. 114–126. The Influence of Electrostatic Truncation on Simulations of Polarizable Systems.

165. B. Chen, J. J. Potoff, and J. I. Siepmann, *J. Phys. Chem. B*, **104**, 2378–2390 (2000). Adiabatic Nuclear and Electronic Sampling Monte Carlo Simulations in the Gibbs Ensemble: Application to Polarizable Force Fields for Water.

166. H. A. Stern, F. Rittner, B. J. Berne, and R. A. Friesner, *J. Chem. Phys.*, **115**, 2237–2251 (2001). Combined Fluctuating-Charge and Polarizable Dipole Models: Application to a Five-Site Water Potential Function.

167. U. Dinur, *J. Phys. Chem.*, **97**, 7894–7898 (1993). Molecular Polarizabilities from Electronegativity Equalization Models.

168. S. W. Rick and B. J. Berne, *J. Phys. Chem. B,* **101**, 10488 (1997). The Free Energy of the Hydrophobic Interaction from Molecular Dynamics Simulations: the Effects of Solute and Solvent Polarizability.

169. R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules,* Oxford University Press, Oxford, UK, 1989.

170. L. J. Bartolotti and K. Flurchick, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, Vol. 7, pp. 187–216 (1996). An Introduction to Density Functional Theory. A. St-Amant, in *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, Vol. 7, pp. 217–259. Density Functional Methods in Biomolecular Modeling.

171. P. Itskowitz and M. L. Berkowitz, *J. Phys. Chem. A,* **101**, 5687–5691 (1997). Chemical Potential Equalization Principle: Direct Approach from Density Functional Theory.

172. D. Borgis and A. Staib, *Chem. Phys. Lett.,* **238**, 187–192 (1995). A Semiempirical Quantum Polarization Model for Water.

173. J. Gao, *J. Phys. Chem. B*, **101**, 657–663 (1997). Toward a Molecular Orbital Derived Empirical Potential for Liquid Simulations.

174. J. Gao, *J. Chem. Phys.,* **109**, 2346–2354 (1998). A Molecular-Orbital Derived Polarization Potential for Liquid Water.

175. B. D. Bursulaya and H. J. Kim, *J. Chem. Phys.*, **108**, 3277–3285 (1998). Generalized Molecular Mechanics Including Quantum Electronic Structure Variation of Polar Solvents. I. Theoretical Formulation via a Truncated Adiabatic Basis Set Description.

176. B. D. Bursulaya, J. Jeon, D. A. Zichi, and H. J. Kim, *J. Chem. Phys.*, **108**, 3286–3295 (1998). Generalized Molecular Mechanics Including Quantum Electronic Structure Variation of Polar Solvents. II. A Molecular Dynamics Simulation Study of Water.

177. A. E. Lefohn, M. Ovchinnikov, and G. A. Voth, *J. Phys. Chem. B*, **105**, 6628–6637 (2001). A Multistate Empirical Valence Bond Approach to a Polarizable and Flexible Water Model.

178. J. J. P. Stewart, in *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1990, Vol. 1, pp. 45–81. Semiempirical Molecular Orbital Methods. M. C. Zerner, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1991, Vol. 2, pp. 313–365. Semiempirical Molecular Orbital Methods.

179. D. E. Williams, in *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1991, Vol. 2, pp. 219–271. Net Atomic Charge and Multipole Models for the Ab Initio Molecular Orbital Potential.

180. R. S. Mulliken, *J. Chem. Phys.,* **23,** 1833, 1841 (1955). Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I. and II. Overlap Populations, Bond Orders, and Covalent Bond Energies. See also: S. M. Bachrach, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1994, Vol. 5, pp. 171–227. Population Analysis and Electron Densities from Quantum Mechanics.

181. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, *J. Am. Chem. Soc.,* **107,** 3902–3909 (1985). AM1: A New General Purpose Quantum Mechanical Molecular Model.

182. W. J. Hehre, R. F. Stewart, and J. A. Pople, *J. Chem. Phys,* **51,** 2657–2664 (1969). Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals. See also: D. Feller and E. R. Davidson, in *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1990, Vol. 1, pp. 1–43. Basis Sets for Ab Initio Molecular Orbital Calculations and Intermolecular Interactions.

183. S. W. Rick and R. E. Cachau, *J. Chem. Phys.,* **112,** 5230–5241 (2000). The Non-Planarity of the Peptide Group: Molecular Dynamics Simulations with a Polarizable Two-State Model for the Peptide Bond.

184. G. Corongiu, *Int. J. Quantum Chem.,* **42,** 1209–1235 (1992). Molecular Dynamics Simulation for Liquid Water Using a Polarizable and Flexible Potential.

185. M. Sprik, *J. Chem. Phys.,* **95,** 6762–6769 (1991). Hydrogen Bonding and the Static Dielctric Constant in Liquid Water.

186. P. G. Kusalik, F. Liden, and I. M. Svishchev, *J. Chem. Phys.,* **103,** 10169–10175 (1995). Calculation of the Third Virial Coefficient for Water.

187. S.-B. Zhu, S. Singh, and G. W. Robinson, *Adv. Chem. Phys.,* **85,** 627–731 (1994). Field-Perturbed Water.

188. A. Wallqvist and R. D. Mountain, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., Wiley-VCH, New York, 1999, Vol. 13, pp. 183–247. Molecular Descriptions of Water: Derivation and Description.

189. J. C. Shelley and D. R. Bérard, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., Wiley-VCH, New York, 1998, Vol. 12, pp. 137–205. Computer Simulation of Water Physisorption at Metal–Water Interfaces.

190. A. A. Chialvo and P. T. Cummings, *J. Phys. Chem.,* **100,** 1309–1316 (1996). Microstructure of Ambient and Supercritical Water: Direct Comparison Between Simulation and Neutron Scattering Experiments.

191. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, in *Intermolecular Forces,* B. Pullman, Ed., Reidel, Dordrecht, The Netherlands, 1981, pp. 331–342. Interaction Models for Water in Relation to Protein Hydration.

192. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.,* **79,** 926–935 (1983). Comparison of Simple Potential Functions for Simulating Liquid Water.

193. C. Millot and A. J. Stone, *Mol. Phys.,* **77,** 439–462 (1992). Towards an Accurate Intermolecular Potential for Water.

194. U. Niesar, G. Corongiu, M.-J. Huang, M. Dupuis, and E. Clementi, *Int. J. Quantum. Chem. Symp.,* **23,** 421–443 (1989). Preliminary Observations on a New Water–Water Potential.

195. G. C. Lie, G. Corongiu, and E. Clementi, *J. Phys. Chem.*, **89**, 4131–4134 (1985). Calculation of the Third Virial Coefficient for Water Using Ab Initio Two-Body and Three-Body Potentials.

196. S. L. Carnie and G. N. Patey, *Mol. Phys.*, **47**, 1129–1151 (1982). Fluids of Polarizable Hard Spheres with Dipoles and Tetrahedral Quadrupoles. Integral Equation Results with Application to Liquid Water.

197. E. R. Batista, S. S. Xantheas, and H. Jónsson, *J. Chem. Phys.*, **109**, 6011–6015 (1999). Multipole Moments of Water Molecules in Clusters and Ice Ih from First Principles Calculations.

198. E. R. Batista, S. S. Xantheas, and H. Jónsson, *J. Chem. Phys.*, **112**, 3285–3292 (2000). Electric Fields in Ice and Near Water Clusters.

199. K. Laasonen, M. Sprik, M. Parrinello, and R. Car, *J. Chem. Phys.*, **99**, 9080–9089 (1993). "Ab Initio" Liquid Water.

200. L. Delle Site, A. Alavi, and R. M. Lynden-Bell, *Mol. Phys.*, **96**, 1683–1693 (1999). The Electrostatic Properties of Water Molecules in Condensed Phases: An *Ab Initio* Study.

201. P. L. Silvestrelli and M. Parrinello, *J. Chem. Phys.*, **111**, 3572–3580 (1999). Structural, Electronic, and Bonding Properties of Liquid Water from First Principles.

202. M. Sprik, *J. Phys. Chem.*, **95**, 2283–2291 (1991). Computer Simulations of the Dynamics of Induced Polarization Fluctuations.

203. J. R. Reimers, R. O. Watts, and M. L. Klein, *Chem. Phys.*, **64**, 95–114 (1982). Intermolecular Potential Functions and the Properties of Water.

204. N. Yoshii, S. Miura, and S. Okazaki, *Chem. Phys. Lett.*, **345**, 195–200 (2001). A Molecular Dynamics Study of Water from Ambient to Sub- and Supercritical Conditions Using a Fluctuating-Charge Potential Model.

205. B. Guillot and Y. Guissani, *J. Chem. Phys.*, **114**(15), 6720–6733 (2001). How To Build a Better Pair Potential for Water.

206. I. M. Svishchev and T. M. Hayward, *J. Chem. Phys.*, **111**, 9034–9038 (1999). Phase Coexistence Properties for the Polarizable Point Charge Model of Water and the Effects of Applied Electric Field.

207. S. W. Rick, *J. Chem. Phys.*, **114**, 2276–2283 (2001). Simulations of Ice and Liquid Water over a Range of Temperatures Using the Fluctuating Charge Model.

208. F. H. Stillinger and A. Rahman, *J. Chem. Phys.*, **60**, 1545–1557 (1974). Improved Simulation of Liquid Water by Molecular Dynamics.

209. P. H. Poole, F. Sciortino, U. Essman, and H. E. Stanley, *Nature (London)*, **360**, 324–328 (1992). Phase Behaviour of Liquid Water.

210. S. R. Billeter, P. M. King, and W. F. van Gunsteren, *J. Chem. Phys.*, **100**, 6692–6699 (1994). Can the Density Maximum of Water Be Found by Computer Simulation?

211. L. A. Báez and P. Clancy, *J. Chem. Phys.*, **101**, 9837–9840 (1994). Existence of a Density Maximum in Extended Simple Point Charge Water.

212. A. Wallqvist and P.-O. Åstrand, *J. Chem. Phys.*, **102**, 6559–6565 (1995). Liquid Densities and Structural Properites of Molecular Models of Water.

213. S. Harrington, P. H. Poole, F. Sciortino, and H. E. Stanley, *J. Chem. Phys.*, **107**, 7443–7450 (1997). Equation of State of Supercooled Water Simulated Using the Extended Simple Point Charge Intermolecular Potential.

214. K. Bagchi, S. Balasubramanian, and M. L. Klein, *J. Chem. Phys.*, **107**, 8561–8567 (1997). The Effects of Pressure on Structural and Dynamical Properties of Associated Liquids: Molecular Dynamics Calculations for the Extended Simple Point Charge Model of Water.

215. W. L. Jorgensen and C. Jensen, *J. Comput. Chem.*, **19**, 1179–1186 (1998). Temperature Dependence of TIP3P, SPC, and TIP4P Water from NPT Monte Carlo Simulations: Seeking Temperatures of Maximum Density.

216. M. W. Mahoney and W. L. Jorgensen, *J. Chem. Phys.,* **112,** 8910–8922 (2000). A Five-Site Model for Liquid Water and the Reproduction of the Density Anomaly by Rigid, Non-polarizable Potential Functions.

217. M. Sprik, M. L. Klein, and K. Watanabe, *J. Phys. Chem.,* **94,** 6483–6488 (1990). Solvent Polarization and Hydration of the Chlorine Anion.

218. L. Perera and M. L. Berkowitz, *J. Chem. Phys.,* **100,** 3085–3093 (1994). Structures of $Cl^-(H_2O)_n$ and $F^-(H_2O)_n$ $(n = 2, 3, \ldots, 15)$ Clusters. Molecular Dynamics Computer Simulations.

219. P. Jungwirth and D. J. Tobias, *J. Phys. Chem. B,* **104,** 7702–7706 (2000). Surface Effects on Aqueous Ionic Solvation: A Molecular Dynamics Study of NaCl at the Air/Water Interface from Infinite Dilution to Saturation.

220. I.-C. Yeh and M. L. Berkowitz, *J. Chem. Phys.,* **112,** 10491–10495 (2000). Effects of the Polarizability and Water Density Constraint on the Structure of Water Near Charged Surfaces: Molecular Dynamics Simulations.

221. G. Gilli, F. Belluci, V. Ferretti, and V. Berolasi, *J. Am. Chem. Soc.,* **111,** 1023–1128 (1989). Evidence for Resonance-Assisted Hydrogen Bonding from Crystal Structure Correlations on the Enol Form of the β-Diketone Fragment.

222. G. A. Jeffrey and W. Saenger, *Hydrogen Bonding in Biological Structures,* Springer-Verlag, Heidelberg, 1991.

223. H. Guo and M. Karplus, *J. Phys. Chem.,* **98,** 7104–7105 (1994). Solvent Influence on the Stability of the Peptide Hydrogen Bond: A Supramolecular Cooperative Effect.

224. R. Ludwig, F. Weinhold, and T. C. Farrar, *J. Chem. Phys.,* **107,** 499–507 (1997). Theoretical Study of Hydrogen Bonding in Liquid and Gaseous *N*-Methylformamide.

225. H. Guo, N. Gresh, B. P. Rogues, and D. R. Salahub, *J. Phys. Chem. B,* **104,** 9746–9754 (2000). Many-Body Effects in Systems of Peptide Hydrogen- Bonded Networks and Their Contributions to Ligand Binding: A Comparison of DFT and Polarizable Molecular Mechanics.

226. I. M. Klotz and J. S. Franzen, *J. Am. Chem. Soc.,* **84,** 3461–3466 (1962). Hydrogen Bonds Between Model Peptide Groups in Solution.

227. A. Wada, *Adv. Biophys.,* **9,** 1–63 (1976). The α-Helix as an Electric MacroDipole.

228. A. van der Vaart, B. D. Bursulaya, C. L. Brooks III, and K. M. Merz Jr., *J. Phys. Chem. B,* **104,** 9554–9563 (2000). Are Many-Body Effects Important in Protein Folding?

229. H. Guo and D. R. Salahub, *Angew. Chem. Int. Ed.,* **37,** 2985–2990 (1998). Cooperative Hydrogen Bonding and Enzyme Catalysis.

230. J. Šponer, H. A. Gabb, J. Leszczynski, and P. Hobza, *Biophys. J.,* **73,** 76–87 (1997). Base-Base and Deoxyribose-Base Stacking Interactions in B-DNA and Z-DNA: A Quantum-Chemical Study.

231. N. Gresh and J. Šponer, *J. Phys. Chem. B,* **103,** 11415–11427 (1999). Complexes of Pentahy-drated Zn^{2+} with Guanine, Adenine, and the Guanine–Cytosine and Adenine–Thymine Base Pairs. Structures and Energies Characterized by Polarizable Molecular Mechanics and Ab Initio Calculations.

232. P. Cieplak, J. Caldwell, and P. Kollman, *J. Comput. Chem.,* **22,** 1048–1057 (2001). Molecular Mechanical Models for Organic and Biological Systems Going Beyond the Atom Centered Two Body Additive Approximation: Aqueous Solution Free Energies of Methanol and *N*-Methyl Acetamide, Nucleic Acid Base, and Amide Hydrogen Hydrogen Bonding and Chloroform/Water Partition Coefficients of the Nucleic Acid Bases.

233. E. D. Stevens, *Acta Crystallogr., Sect. B,* **34,** 544–551 (1978). Low Temperature Experimental Electron Density Distribution in Formamide.

234. G. A. Jeffrey, *An Introduction to Hydrogen Bonding,* Oxford University Press, New York and Oxford, 1997.

235. M. W. MacArthur and J. M. Thornton, *J. Mol. Biol.,* **264,** 1180–1195 (1996). Deviations from Planarity of the Peptide Bond in Peptides and Proteins.

236. E. J. Dodson, G. J. Davis, V. S. Lamzin, G. N. Murshudov, and K. S. Wilson, *Structure*, **6,** 685–690 (1998). Validation Tools: Can They Indicate the Information Content of Macro-molecular Crystal Structures?

237. P. Hobza and J. Šponer, *Chem. Rev.*, **99,** 3247–3276 (1999). Structure, Energics, and Dynamics of the Nucleic Acid Base Pairs: Nonempirical *Ab Initio* Calculations.

238. J. N. Wilson and R. M. Curtis, *J. Phys. Chem.*, **74,** 187–196 (1970). Dipole Polarizabilities of Ions in Alkali Halide Crystals.

239. G. D. Mahan, *Solid State Ionics*, **1,** 29–45 (1980). Polarizability of Ions in Crystals.

240. P. W. Fowler and P. A. Madden, *Phys. Rev. B: Condens. Matter*, **29,** 1035–1042 (1984). In-crystal Polarizabilities of Alkali and Halide Ions.

241. J. K. Badenhoop and F. Weinhold, *J. Chem. Phys.*, **107,** 5422–5432 (1997). Natural Steric Analysis: *Ab Initio* van der Waals Radii of Atoms and Ions.

242. A. Warshel and R. M. Weiss, *J. Am. Chem. Soc.*, **102,** 6218–6226 (1980). An Emprical Valence Bond Approach for Comparing Reactions in Solutions and in Enzymes.

243. M. J. L. Sangster, *Solid State Commun.*, **15,** 471–474 (1974). Properties of Diatomic Molecules from Dynamical Models for Alkali Halide Crystals.

244. L. Greengard, and V. Rokhlin, *J. Comput. Phys.*, **73,** 325–348 (1987). A Fast Algorithm for Particle Simulations.

245. H.-Q. Ding, N. Karasawa, and W. A. Goddard III, *J. Chem. Phys.*, **97,** 4309–4315 (1992). Atom Level Simulations of a Million Particles: The Cell Multipole Method for Coulomb and London Nonbond Interactions.

246. R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles,* Institute of Physics Publishing, Bristol, UK, 1988.

247. T. Darden, D. York, and L. Pedersen, *J. Chem. Phys.*, **93,** 10089–10092 (1993). Particle Mesh Ewald: An $N \log(N)$ Method for Ewald Sums in Large Systems.

248. P. Ewald, *Ann. Phys.*, **64,** 253–287 (1921). Die Berechnung optischer und elektrostatischer Gitterpotentiale.

249. D. M. Heyes, *J. Chem. Phys.*, **74,** 1924–1929 (1981). Electrostatic Potentials and Fields in Infinite Point Charge Lattices.

250. M. Tuckerman, B. J. Berne, and G. J. Martyna, *J. Chem. Phys.*, **97,** 1990–2001 (1992). Reversible Multiple Time Scale Molecular Dynamics.

251. S. J. Stuart, R. Zhou, and B. J. Berne, *J. Chem. Phys.*, **105,** 1426–1436 (1996). Molecular Dynamics with Multiple Timescales: The Selection of Efficient Reference System Propaga-tors.

252. P. Kaatz, E. A. Donley, and D. P. Shelton, *J. Chem. Phys.*, **180,** 849–856 (1998). A Comparison of Molecular Hyperpolarizabilities from Gas and Liquid Phase Measurements.

253. G. Maroulis, *J. Chem. Phys.*, **94,** 1182–1190 (1991). Hyperpolarizability of H_2O.

254. The linear increase in polarization with molecule size assumes that the characteristic polarizability α is small compared to the characteristic volume per polarizable unit: $\alpha \ll r^3$. See Eqs. [15] and [16]. In cases where this approximation does not hold, the polarizability will increase faster than linearly with system size, leading to a polarization catastrophe.

255. B. L. Bush, C. I. Bayly, and T. A. Halgren, *J. Comput. Chem.*, **20,** 1495–1516 (1999). Consensus Bond-Charge Increments Fitted to Electrostatic Potential or Field of Many Compounds: Application to MMFF94 Training Set.

CHAPTER 4

New Developments in the Theoretical Description of Charge-Transfer Reactions in Condensed Phases

Dmitry V. Matyushov* and Gregory A. Voth[†]

*Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287-1604, and [†]Department of Chemistry and Henry Eyring Center for Theoretical Chemistry, University of Utah, 315 South 1400 East, Salt Lake City, Utah 84112

INTRODUCTION

Nearly half a century of intense research in the field of electron transfer (ET) reactions in condensed phases has produced remarkable progress in the experimental and theoretical understanding of the key factors influencing the kinetics and thermodynamics of these reactions. The field evolved in order to describe many important processes in chemistry and is actively expanding into biological and materials science applications.[1] Due to its significant experimental background and relative simplicity of the reaction mechanism, the problem of electron transitions in condensed solvents turned out to be a benchmark for testing fundamental theoretical approaches to chemical activation. A number of excellent reviews dealing with various aspects of the field have been written. Two volumes of *Advances in Chemical Physics* (Vols. 106 and 107, 1999) covered much of the progress in the field achieved in recent decades. Therefore, the aim of this chapter is not to replicate these

reviews, but rather to highlight some very recent developments in the field that have not been reviewed. This chapter will provide the reader with a step-by-step statistical mechanical buildup of the theoretical machinery currently employed in ET research. By virtue of the "frontier" nature of this material, many traditional subjects of ET studies are not covered here. The reader will be referred to previous reviews whenever possible, but many excellent contributions are not directly cited.

This chapter concerns the energetics of charge-transfer (CT) reactions. We will not discuss subjects dealing with nuclear dynamical effects on CT kinetics.[2–4] The more specialized topic of employing the liquid-state theories to calculate the solvation component of the reorganization parameters[5] is not considered here. We concentrate instead on the general procedure of the statistical mechanical analysis of the activation barrier to CT, as well as on its connection to optical spectroscopy. Since the very beginning of ET research,[6] steady-state optical spectroscopy has been the major source of reliable information about the activation barrier and preexponential factor for the ET rate. The main focus in this chapter is therefore on the connection between the statistical analysis of the reaction activation barrier to the steady-state optical band shape.

The ET reaction is usually referred to as a process of underbarrier tunneling and subsequent localization of an electron from the potential well of the donor to the potential well of the acceptor (Figure 1). This phenomenon occurs in a broad variety of systems and reactions (see Ref. 1 for a list of
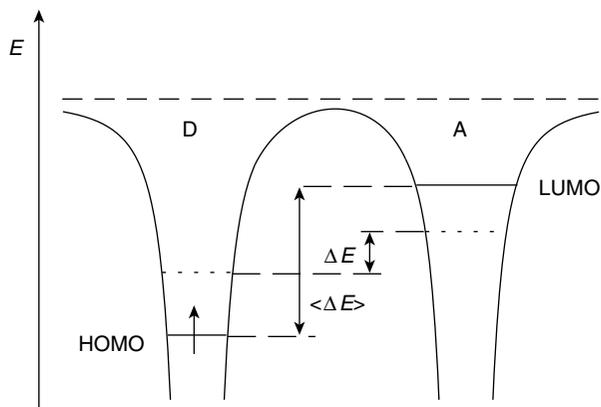

Figure 1 Potential energy wells for the electron localized on the donor (D) and acceptor (A) sites. The parameter $\langle \Delta E \rangle$ indicates the average energy gap for an instantaneous (Franck–Condon) transfer of the electron from the donor HOMO to the acceptor LUMO. The dotted lines show the electronic energies on the donor and acceptor at a nonequilibrium nuclear configuration with a nonequilibrium energy gap ΔE. The upper dashed horizontal line indicates the bottom of the conduction band of the electrons in the solvent.

applications). For electron tunneling to occur, the electronic states of the donor and acceptor sites must come into resonance (degeneracy) with each other. Degeneracy occurs as a result of thermal nuclear motions of the donor–acceptor complex and the condensed-phase medium. The condition of zero energy gap, $\Delta E = 0$, between the donor and acceptor electronic levels determines the position of the transition state for an ET reaction. The ET rate constant is proportional to the probability of such a configuration

$$k_{\mathrm{ET}} \propto \mathrm{FCWD}(0) \tag{1}$$

where the *Franck–Condon weighted density* (FCWD), $\mathrm{FCWD}(\Delta E)$, determines the probability of creating a configuration with energy gap ΔE.

Electron transfer refers to the situation when essentially all the electronic density is transferred from the donor to the acceptor. The process of CT, in the present context, refers to basically the same event, but the electron density is not completely relocalized and is distributed between the two potential wells. The key factor discriminating between ET and CT reactions is the ET matrix element,[7] H_{ab}, often called the hopping element in solid-state applications. The ET matrix element is the off-diagonal matrix element of the system Hamiltonian taken on the localized diabatic states of the donor and acceptor sites (see below). [The term *diabatic* refers to localized states which do not diagonalize the system Hamiltonian. These localized states are the true states of the donor and acceptor fragments when these fragments are infinitely separated. For covalently bound complexes, diabatic states become just some basis states that allow reasonable localization of the electronic density on the donor and acceptor fragments of the molecule. *Adiabatic* states, in contrast, are actual states of the molecule between which electronic (including optical) transitions occur.]

For long-range electron transitions, the direct electronic overlap, exponentially decaying with distance between the donor and acceptor units, is weak, leading to a small magnitude of the expectation value of H_{ab}. Such processes, especially important in biological applications,[8] can be characterized as nonadiabatic ET reactions. The small magnitude of the ET matrix element can be employed to find the transition rate using quantum mechanical perturbation theory. In this theory, the rate constant found by the Golden Rule approximation[9,10] is called the nonadiabatic ET rate constant, and the ET reaction is classified as nonadiabatic ET.[11] (The *Golden Rule formula* is the first-order perturbation solution for the rate of quantum mechanical transitions caused by that perturbation.) The ET rate constant is then proportional to $|H_{ab}|^2$

$$k_{\mathrm{NA}} \propto |H_{ab}|^2 \, \mathrm{FCWD}(0) \tag{2}$$

Creation of the resonance electronic configuration of the ET transition state, $\Delta E = 0$, is by necessity a many-body event, including complex interactions of the transferred electron with many nuclear degrees of freedom. The
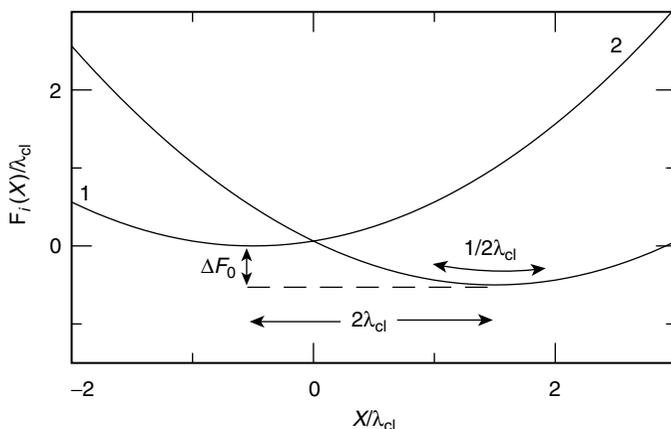
Figure 2 Two parameters defining the Marcus–Hush model of two intersecting parabolas: the equilibrium free energy gap ΔF_0 and the classical reorganization energy λ_{cl}. The parabolas curvature is $1/(2\lambda_{cl})$.

great achievement of the Marcus–Hush (MH) model[6,12–14] of ET was to reduce the many-body problem to a one-dimensional (1D) picture of intersecting ET free energy surfaces, $F_i(X)$ ($i = 1$ for the initial ET state, $i = 2$ for the final ET state, Figure 2). Each point on the free energy surface represents the reversible work invested to create a nonequilibrium fluctuation of the nuclei resulting in a particular value of the donor–acceptor electronic energy gap

$$X = \Delta E \qquad [3]$$

The electronic energy gap thus serves as a collective reaction coordinate X reflecting the strength of coupling of the nuclear modes to the electronic states of the donor and acceptor. The point of intersection of $F_1(X)$ and $F_2(X)$ sets up the ET transition state, $X = 0$.

    The definition of the ET reaction coordinates according to Eq. [3] allows a direct connection between the activated ET kinetics and steady-state optical spectroscopy. In a spectroscopic experiment, the energy of the incident light with the frequency v ($\bar{v}$ is used for the wavenumber) is equal to the donor–acceptor energy gap

$$hv = X \qquad [4]$$

and monitoring the light frequency directly probes the distribution of donor–acceptor energy gaps. The intensity of optical transitions $I(v)$ is then proportional to FCWD(hv)[15]

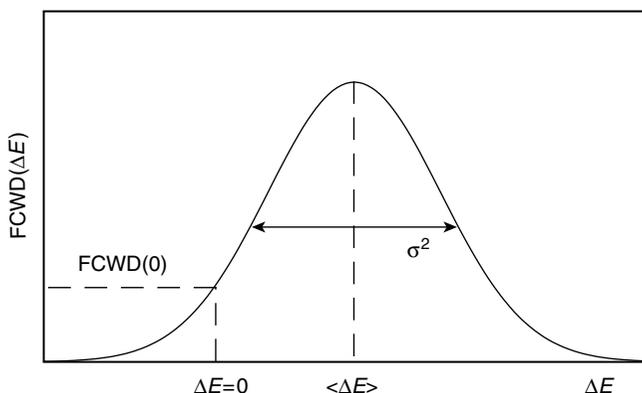$$I(v) \propto |m_{12}|^2 \, \text{FCWD}(hv) \qquad [5]$$

Figure 3 Franck–Condon weighted density of energy gaps between the donor and acceptor electronic energy levels. The parameters $\langle \Delta E \rangle$ and σ^2 indicate the first and second spectral moments, respectively. FCWD(0) shows the probability of zero energy gap entering the ET rate (Eq. [2]).

where m_{12} is the adiabatic transition dipole moment. Knowledge of the spectral band shape can in principle provide the activation barrier through FCWD(0) (Figure 3), and the Mulliken-Hush relation connects $|H_{ab}|$ to $|m_{12}|$.[6] (In contrast to Marcus-Hush which refers to the theory of electron transfer activation, the Mulliken-Hush equation describes the preexponential factor of the rate constant. We spell out Mulliken–Hush each place it occurs in this chapter and use the acronym MH to refer to only Marcus–Hush.) In practice, however, FCWD(0) cannot be extracted from experimental spectra, and one needs a theoretical model to calculate FCWD(0) from experimental band shapes measured at the frequencies of the corresponding electronic transitions. This purpose is achieved by a band shape analysis of optical lines.

The two main nuclear modes affecting electronic energies of the donor and acceptor are intramolecular vibrations of the molecular skeleton of the donor–acceptor complex and molecular motions of the solvent. If these two nuclear modes are uncoupled, one can arrive at a set of simple relations between the two spectral moments of absorption and/or emission transitions and the activation parameters of ET. The most transparent representation is achieved when the quantum intramolecular vibrations are represented by a single, effective vibrational mode with the frequency ν_v (Einstein model).[15–17] If both the forward (absorption) and backward (emission) optical transitions are available, their first spectral moments determine the reorganization energies of quantum vibrations, λ_v, and of the classical nuclear motions of the donor–acceptor skeleton and the solvent, λ_{cl}:

$$h(\nu_{abs} - \nu_{em}) = 2(\lambda_{cl} + \lambda_v) \qquad [6]$$

where ν_{abs} and ν_{em} are the first spectral moments for absorption and emission, respectively:

$$\nu_{abs/em} = \frac{\int \nu I_{abs/em}(\nu)d\nu}{\int I_{abs/em}(\nu)d\nu} \qquad [7]$$

Here $I_{abs/em}(\nu)$ is the transition intensity. The vibrational reorganization energy λ_v is defined in terms of force constants, k_α, and displacements, ΔQ_α, of the vibrational normal coordinates Q_α as $\lambda_v = \frac{1}{2}\sum_\alpha k_\alpha \Delta Q_\alpha^2$.[15–17] In this chapter, we use λ for the solvent component of the classical reorganization energy λ_{cl}. The subscripts 1 and 2 are used to distinguish between the reorganization energy of the initial ($i = 1$) and final ($i = 2$) ET states when the reorganization energies in these states are different.

The mean of the first two moments gives the equilibrium free energy difference between the final and initial states of the ET reaction

$$h\nu_m = \tfrac{1}{2}h(\nu_{abs} + \nu_{em}) = \Delta F_0 = F_{02} - F_{01} \qquad [8]$$

The two parameters, λ_{cl} and ΔF_0, actually fully define the parabolic ET free energy surfaces $F_i(X)$ in the MH formulation (Figure 2). Calculation of these two parameters has become the main historical focus of the ET models addressing the thermodynamics of the ET activation barrier. The latter, according to MH theory, can be written in terms of ΔF_0 and λ_{cl} as

$$F_i^{act} = \frac{(\lambda_{cl} \pm \Delta F_0)^2}{4\lambda_{cl}} \qquad [9]$$

where $i = 1$ and "+" refer to the forward transition, and $i = 2$ and "−" refer to the backward transition.

The second spectral moments of absorption and emission lines

$$\sigma_{abs/em}^2 = \frac{\int \nu^2 I_{abs/em}(\nu)d\nu}{\int I_{abs/em}(\nu)d\nu} - \left(\nu_{abs/em}\right)^2 \qquad [10]$$

are equal in the MH formulation

$$\sigma_{abs}^2 = \sigma_{em}^2 \qquad [11]$$

They are related to the classical and vibrational reorganization energies as follows[18]

$$\sigma_{abs/em}^2 = 2k_B T \lambda_{cl} + h\nu_v \lambda_v \qquad [12]$$

where k_B is the Boltzmann constant and T is temperature.

Equations [6]–[12] establish a theoretical basis for calculating the activation barrier of ET from spectroscopic observables. This formalism rests on a set of fundamental assumptions of the MH picture that can be summarized as follows: (1) The electronic coupling between the donor and acceptor states is neglected in the calculation of the Franck–Condon weighted density. The latter depends only on electronic energies of localized electronic states and their coupling to the nuclear modes of the solvent and the donor–acceptor complex. (2) A two-state solute is considered. The manifold of the donor and acceptor electronic levels is limited to only two states between which the electron is transferred. (3) The intramolecular vibrations and solvent molecular motions are decoupled. (4) The linear response approximation is used for the interaction of the donor–acceptor complex with the solvent. The linear response approximation assumes that the free energy of solvation of an electric charges localized on the donor–acceptor complex is a quadratic function of this charge.

The neglect of the electronic coupling in the calculation of the FCWD (assumption 1) was adopted in the original Marcus and Hush formulation.[6,12] Within this framework, the ET matrix element does not strongly affect the nuclear fluctuations, although a nonzero value of $|H_{ab}|$ is required for electronic transitions to occur. In other words, the transferred electron is assumed to be fully localized in the calculation of the FCWD. To classify electronic delocalization, Robin and Day distinguished between three classes of symmetrical ($\Delta F_0 = 0$) systems.[19]

- In Class I systems, the coupling is very weak, and there are essentially no electronic transitions.
- Class II systems remain valence-trapped (localized), and $0 < 2|H_{ab}| \leq \lambda_{cl}$.
- In Class III systems, $2|H_{ab}| > \lambda_{cl}$, and the electron is fully delocalized between the donor and acceptor.

The MH formulation is designed to describe the case of intermediate couplings (weak-coupling limit of Class II) when $|H_{ab}|$ can be neglected in the FCWD(0) for activated transitions and the transition moment m_{12} can be neglected in the FCWD(v) for optical transitions. In the absence of a theory incorporating $|H_{ab}|$ and m_{12} into the FCWD, there is no general understanding when this approximation is applicable to particular ET systems or how the relations between optical and activation observables are affected by inclusion of electronic delocalization into the FCWD.[20]

The limitations of the MH picture considerably narrow the range of systems covered by the theory. A considerable range of processes in which the donor–acceptor coupling is strong enough to change the molecular charge distribution under the influence of nuclear fluctuations cannot be treated theoretically. All such processes can be characterized as CT reactions. Weak electronic coupling characteristic of ET exists for intermolecular and

long-distance intramolecular reactions. Many systems with intramolecular electronic transitions over a relatively short distance between the initial and final centers of electron localization have been synthesized in recent years.[21,22] They commonly incorporate the same basic design in which the donor and acceptor units are linked in one molecule through a bridge moiety. In a case of closely separated donor and acceptor units, electronic states on these two sites are strongly coupled, resulting in a substantial delocalization of the electronic density. The electronic density is only partially transferred, and the process can be classified as a CT transition.

The MH formulation for the activation barrier and the related connection between activation ET parameters and optical observables generally do not apply to CT reactions. Hence the researcher is left without a procedure of calculating the activation barrier from spectroscopy. Not being able to calculate the barrier is a deficiency, and this chapter discusses some emerging approaches to develop a theory of CT processes with an explicit account for electronic delocalization effects. In application to optical transitions, this theory should lead to the development of a band shape analysis broadly applicable to Class II and III systems. The effect of electronic delocalization on the solvent component of the FCWD is emphasized here. The previously reviewed problem of delocalization effects on intramolecular vibrations[23] is not included. We also review some new approaches going beyond the two-state approximation in terms of incorporating polarizability of the donor–acceptor complex (assumption 2), and discuss some recent studies on nonlinear solvation effects (assumption 4). There are some very recent indications in the literature pointing to a possibility of an effective coupling between vibrational modes of the donor–acceptor complex and solvent fluctuations (assumption 3), but no consensus on when and why these effects are significant has yet been reached. We briefly discuss the available experimental and theoretical findings.

The first part of this chapter contains an introduction to the statistical mechanical formulation of the CT free energy surfaces. Importantly, it shows how to extend the traditional MH picture of two ET parabolas to a more general case of two CT free energy surfaces of a two-state donor–acceptor complex. The notation we utilize below distinguishes between these two cases in the following fashion: we use the indices 1 and 2 to denote the two ET free energy surfaces, as in Figure 2, and refer to the lower and upper CT free energy surfaces with "−" and "+", respectively. The parameters entering the activation barrier of CT transitions depend on the choice of the basis set of wave functions of the initial and final states of the donor–acceptor complex. The standard MH formulation is based on the choice of a localized, diabatic basis set. When this choice is adopted, we use the superscript "d" to refer to diabatic wave functions. An alternative description is possible in terms of adiabatic wave functions, and this situation is distinguished by the superscript "ad". We also provide a basis-invariant formulation of the theory for a two-state

donor–acceptor complex. A description of CT activation and spectroscopy in terms of two crossing, free energy surfaces (Figure 2) is in fact possible for any choice of the basis set as long as the off-diagonal matrix elements of the solute quantum mechanical operators can be neglected. In cases when a description in both diabatic and adiabatic representations is possible (as it is for the Q-model discussed below), we will not specify the basis by dropping the "d" and "ad" superscripts.

The statistical mechanical analysis of ET and CT free energy surfaces developed in the first part of this chapter is applied to the calculation of optical absorption and emission profiles in the second part. This application of the theory, related to the band shape analysis of optical line shapes, has been a central issue in understanding CT energetics for several decades.[16] The chapter is designed to demonstrate how the extension of the basic models used to describe the thermodynamics of CT is reflected in asymmetry of the energy gap law (dependence of the CT activation barrier on the equilibrium free energy gap) and more complex and structured optical band shapes. The development of a corresponding band shape analysis incorporating these new features is in its infancy, and we will certainly see more activity in this field in the future.

PARADIGM OF FREE ENERGY SURFACES

The CT/ET free energy surface is the central concept in the theory of CT/ET reactions. The surface's main purpose is to reduce the many-body problem of a localized electron in a condensed-phase environment to a few collective reaction coordinates affecting the electronic energy levels. This idea is based on the Born–Oppenheimer (BO) separation[24] of the electronic and nuclear time scales, which in turn makes the nuclear dynamics responsible for fluctuations of electronic energy levels (Figure 1). The choice of a particular collective mode is dictated by the problem considered. One reaction coordinate stands out above all others, however, and is the energy gap between the two CT states as probed by optical spectroscopy (i.e., an experimental observable).

Our discussion of the CT free energy surfaces involves a hierarchy of reaction coordinates (Figure 4). It starts from the instantaneous free energy surfaces obtained from tracing out (statistical averaging) the electronic degrees of freedom in the system density matrix (i.e., solving the electronic problem for fixed nuclear coordinates). In the case when the direction of electron transfer sets up the only preferential direction in the CT system, one can define a scalar reaction coordinate as the projection of the nuclear solvent polarization on the differential electrical field of the solute. Depending on the basis set employed, this gives the diabatic or adiabatic scalar reaction coordinates, Y^d and Y^{ad} (Figure 4). At this step, a reaction coordinate depends on the basis set of solute wave functions employed. This dependence is eliminated when a scalar reaction coordinate is projected on the energy gap between the CT surfaces.
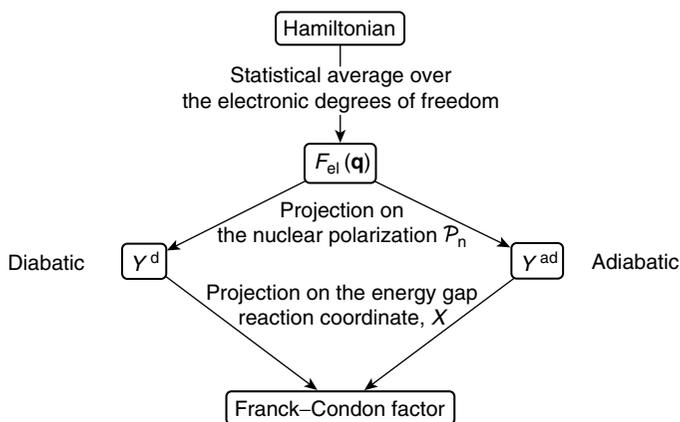
Figure 4 Hierarchy of reaction coordinates in deriving the Franck–Condon factor from the system Hamiltonian.

The free energy gap, equal to the energy of the incident light, is basis independent. It defines the Franck–Condon factor entering the optical band shapes. The analysis below follows this general scheme (Figure 4).

Formulation

Electron transfer and, more broadly, CT reactions belong to a general class of problems having a quantum subsystem interacting with a condensed-phase thermal bath. The main challenge in describing such systems is the necessity to treat the quantum subsystem coupled to a wide spectrum of classical and quantum modes of the condensed environment. It implies that the calculation of some property of interest F involves taking a restricted statistical average (trace, Tr) over both the electronic and nuclear modes

$$F(\mathbf{Q}, t) = \mathrm{Tr}'_n \mathrm{Tr}_{el}[\hat{\rho}(t)] \tag{13}$$

where

$$\hat{\rho}(t) = e^{iHt}\hat{\rho}(0)e^{-iHt} \tag{14}$$

is the density matrix of the system defined by the Hamiltonian H; $\hat{\rho}(0) = \exp(-\beta H)$ and $\beta = 1/k_B T$. The quantity Tr_{el} denotes the trace over the electronic degrees of freedom, and Tr'_n refers to an incomplete or restricted trace over the nuclear degrees of freedom, excluding a manifold of modes $\mathbf{Q}$ that are of interest for some particular problem.

Depending on the order of the statistical average in Eq. [13], there are two basic approaches to calculate $F(\mathbf{Q}, t)$. Considerable progress has been

achieved in treating the quantum dynamics in the framework of the functional integral representation of the quantum subsystem.[25,26] In this approach, the electron trace is taken out $(\mathrm{Tr_{el}Tr'_n})$ and is represented by a functional integral over the quantum trajectories of the system. The inner trace over the nuclear coordinates is then taken for each point of the quantum path by statistical mechanics methods or by computer simulations of the many-particle system.

The more traditional approach to treat the problem outlined by Eq. [13] goes back to the theory of polarons in dielectric crystals.[27,28] It employs the two-step procedure corresponding to two traces in Eq. [13]: first, the trace over the electronic subsystem is taken with the subsequent restricted trace over the nuclear coordinates. This approach, basic to the MH theory of ET, turns out to be very convenient for a general description of several quantum dynamical problems in condensed phases. It is currently widely used in steady-state[29] and time-resolved[2] spectroscopies and in theories of proton transfer,[30] dissociation reactions,[31] and other types of reactions in condensed media. The central feature of the approach is the intuitively appealing and pictorially convenient representation of the activated electron transition as dynamics on the free energy surface of the reaction. Here, we start with outlining the basic steps and concepts leading to the *paradigm of the free energy surfaces*. In this section, we confine the discussion only to classical modes of the solvent. The results obtained here are then used to discuss the construction of the Franck–Condon factor of optical transitions, including quantum intramolecular excitations of the donor–acceptor complex.

The first step of the derivation involves the BO approximation separating the characteristic timescales of the electronic and nuclear motions in the system. In this step, the instantaneous free energy depending on the system nuclear coordinates $\mathbf{q}$ is defined by

$$e^{-\beta F_{el}(\mathbf{q})} = \mathrm{Tr_{el}}\left[e^{-\beta H}\right] \qquad [15]$$

For many homogeneous ET reactions, the energies of electronic excitations are much higher than the energy of the thermal motion, which is of the order of $k_B T$. In such cases, the free energy $F_{el}(\mathbf{q})$ in Eq. [15] can be replaced by the energy, independent of the bath temperature. This does not, however, happen for electrochemical discharge where states of conduction electrons form a continuum with thermal excitations between them. Entropic effects then gain importance, and the free energy $F_{el}(\mathbf{q})$ should be considered in Eq. [15] (see below).

The instantaneous free energy $F_{el}(\mathbf{q})$ is the equilibrium free energy, implying equilibrium populations of the electronic states in the system. It is not suitable for describing nonequilibrium processes with nonequilibrium populations of the ground and excited states of the donor–acceptor complex.

In such cases, the instantaneous eigenvalues $E_i(\mathbf{q})$ of the solute electronic Hamiltonian that form the free energy $F_{\text{el}}(\mathbf{q})$

$$\sum_i e^{-\beta E_i(\mathbf{q})} = e^{-\beta F_{\text{el}}(\mathbf{q})} \qquad [16]$$

should be considered as the basis for building the ET free energy surfaces. The energies $E_i(\mathbf{q})$ can be used for nonequilibrium dynamics, since the population of each surface is not limited by the condition of equilibrium as it is the case in Eq. [16].

An electron is transferred between its centers of localization as a result of underbarrier tunneling when the instantaneous electronic energies $E_i(\mathbf{q})$ come into resonance due to thermal fluctuation or radiation of the medium (Figure 1). The difference between the energies $E_i(\mathbf{q})$ thus makes a natural choice for the ET reaction coordinate (cf. to Eq. [3])

$$X = \Delta E(\mathbf{q}) = E_2(\mathbf{q}) - E_1(\mathbf{q}) \qquad [17]$$

as first suggested by Lax[15] and then utilized in many ET studies.[5,17,32,33] The reversible work necessary to achieve a particular magnitude of the energy gap X defines the free energy profile of CT in terms of a Dirac delta function

$$e^{-\beta F_i(X) + \beta F_{0i}} = \beta^{-1} \text{Tr}_n[\delta(X - \Delta E(\mathbf{q})) e^{-\beta E_i(\mathbf{q})}] / \text{Tr}_n[e^{-\beta E_i(\mathbf{q})}] \qquad [18]$$

The partial trace in nuclear degrees of freedom in Eq. [13] is replaced in Eq. [18] by the constraint imposed on the collective reaction coordinate X representing the energy gap between the two levels involved in the transition. This reduces the many-body problem of calculating the activation dynamics in the coordinate space $\mathbf{q}$ to the dynamics over just one coordinate X. As we show in the discussion of optical transition below, the same Boltzmann factor as in Eq. [18] comes into expressions for optical profiles of CT bands. The solvent component of the FCWD then becomes

$$\text{FCWD}_i^{\text{s}}(X) = \beta e^{-\beta F_i(X) + \beta F_{0i}} \qquad [19]$$

A more general definition of the FCWD includes overlap integrals of quantum nuclear modes.[15,17] The definition given by Eq. [19] includes only classical solvent modes (superscript "s") for which these overlap integrals are identically equal to unity. An extension of Eq. [19] to the case of quantum intramolecular excitations of the donor–acceptor complex is given below in the section discussing optical Franck–Condon factors.

In Eqs. [18] and [19], F_{0i} is the equilibrium free energy of the system in each CT state

$$e^{-\beta F_{0i}} = \text{Tr}_n[e^{-\beta E_i(\mathbf{q})}] \qquad [20]$$

Although the free energy profile $F_i(X)$ and the free energy F_{0i} are combined in one equation (Eq. [18]), they have a somewhat different physical meaning. The free energy F_{0i} is the total, equilibrium free energy of the system calculated for all its configurations. The difference of F_{02} and F_{01} makes the free energy gap ΔF_0 entering the MH theory of ET (Figure 2). Thus

$$\Delta F_0 = F_{02} - F_{01} \qquad [21]$$

On the other hand, $F_i(X)$ is the *constrained*, incomplete free energy implying that some of the configurations of the system separated by the δ-function in Eq. [18] are not included in the calculation of $F_i(X)$.[33] The phase space of the system is not completely sampled in defining $F_i(X)$, in contrast to the complete sampling for F_{0i}. Using molecular dynamics simulations and explicit atomistic models, the free energy in Eq. [18] can be explicitly mapped out. This kind of calculation has become fairly routine (see, e.g., Refs. 32 and 33). It should be noted, however, that such simulations usually neglect the electronic polarizability of both the CT complex and the solvent. These effects may be large (cf. Ref. 32 and the later discussion in this chapter).

When the number of electronic states can be limited to two (two-state model), the analytic properties of the generating function for the two CT free energy surfaces can be used to establish a linear relation between them.[32] The δ-function in Eq. [18] can be represented as a Fourier integral that allows one to rewrite the CT free energy in the integral form

$$e^{-\beta F_i(X) + \beta F_{0i}} = \int_{-\infty}^{\infty} \frac{d\xi}{2\pi} \mathcal{G}_i(\xi, X) \qquad [22]$$

The integral is taken over one of the variables of the generating function

$$\mathcal{G}_i(\xi, X) = e^{i\xi\beta X} \text{Tr}_n\left(e^{-i\xi\beta\Delta E - \beta E_i}\right) / \text{Tr}_n\left[e^{-\beta E_i}\right] \qquad [23]$$

Analytic properties of $\mathcal{G}_i(\xi, X)$ in the complex ξ-plane then allow one to obtain a linear connection between the free energy surfaces

$$F_2(X) = F_1(X) + X \qquad [24]$$

as first established by Warshel.[32,34] This relation is based on the transformation of the integral

$$\int_{-\infty}^{\infty} \frac{d\xi}{2\pi} \mathcal{G}_2(\xi, X) = e^{\beta(\Delta F_0 - X)} \int_{-i-\infty}^{-i+\infty} \frac{d\xi}{2\pi} \mathcal{G}_1(\xi, X) \qquad [25]$$

that leads to Eq. [24] provided the integrals over the segments $(-i - \infty,$ $-i + \infty)$ and $(-\infty, +\infty)$ are equal. This happens when $\mathcal{G}_1(\xi, X)$ is analytic in ξ inside the closed contour with the two segments as its boundaries. The linear relation between $F_2(X)$ and $F_1(X)$ breaks down when the generating function is not analytic inside this contour.

Two-State Model

The two-state model (TSM) provides a very basic description of quantum transitions in condensed-phase media. It limits the manifold of the electronic states of the donor–acceptor complex to only two states participating in the transition. In this section, the TSM will be explored analytically in order to reveal several important properties of ET and CT reactions. The gas-phase Hamiltonian of the TSM reads

$$H_0 = \sum_{i=a,b} I_i a_i^+ a_i + H_{ab}\left(a_a^+ a_b + a_b^+ a_a\right) \qquad [26]$$

where I_i are diagonal gas-phase energies, and H_{ab} is the off-diagonal Hamiltonian matrix element usually called the ET matrix element.[7] In Eq. [26], a_i^+, a_i are the fermionic creation and annihilation operators in the states $i = a, b$.

The Hamiltonian in Eq. [26] is usually referred to as the diabatic representation, employing the diabatic basis set $\{\phi_a, \phi_b\}$ in which the Hamiltonian matrix is not diagonal. There is, of course, no unique diabatic basis as any pair $\{\tilde{\phi}_a, \tilde{\phi}_b\}$ obtained from $\{\phi_a, \phi_b\}$ by a unitary transformation can define a new basis. A unitary transformation defines a linear combination of ϕ_a and ϕ_b which, for a two-state system, can be represented as a rotation of the $\{\phi_a, \phi_b\}$ basis on the angle ψ

$$\begin{aligned} \tilde{\phi}_a &= \cos\psi\,\phi_a + \sin\psi\,\phi_b \\ \tilde{\phi}_a &= -\sin\psi\,\phi_a + \cos\psi\,\phi_b \end{aligned} \qquad [27]$$

One such rotation is usually singled out. A unitary transformation $\{\phi_a, \phi_b\} \rightarrow \{\phi_1, \phi_2\}$ diagonalizing the Hamiltonian matrix

$$H_0 = \sum_{i=1,2} E_i a_i^+ a_i \qquad [28]$$

generates the adiabatic basis set $\{\phi_1, \phi_2\}$. The adiabatic gas-phase energies are then given as

$$E_i = \tfrac{1}{2}(I_a + I_b) \pm \tfrac{1}{2}\sqrt{(I_b - I_a)^2 + 4H_{ab}^2}, \qquad \Delta E_{12} = E_2 - E_1 \qquad [29]$$

where "+" and "−" correspond to $i = 1$ and $i = 2$, respectively. Here, we outline the procedure of building the CT free energy surfaces in the diabatic representation and then discuss advantages of using the adiabatic representation.

When the donor–acceptor complex is placed in a solvent, its Hamiltonian changes due to the solute–solvent interaction

$$H_{\text{int}} = -\hat{\mathcal{E}} \cdot \mathcal{P} \qquad [30]$$

Here, the dot product of two calligraphic letters stands for an integral over the solvent volume V

$$\hat{\mathcal{E}} \cdot \mathcal{P} = \int_V \hat{\mathbf{E}} \cdot \mathbf{P} \, d\mathbf{r} \qquad [31]$$

and $\hat{\mathcal{E}}$ is the electric field operator of the transferred electron coupled to the polarizability of the solvent $\mathcal{P}$. The system Hamiltonian then becomes

$$H = H_{\text{B}} + \sum_{i=a,b} (I_i - \mathcal{E}_i \cdot \mathcal{P}) a_i^+ a_i + (H_{ab} - \mathcal{E}_{ab} \cdot \mathcal{P})(a_b^+ a_a + a_a^+ a_b) \qquad [32]$$

where H_{B} refers to the Hamiltonian of the solvent (thermal bath); $\mathcal{E}_i = \langle \phi_i | \hat{\mathcal{E}} | \phi_i \rangle$ and $\mathcal{E}_{ab} = \langle \phi_a | \hat{\mathcal{E}} | \phi_b \rangle$.

The solvent Hamiltonian H_{B} includes two components. The first one is an intrinsically quantum part that describes polarization of the electronic clouds of the solvent molecules. This polarization is given by the electronic solvent polarization, $\mathcal{P}_{\text{e}}$. The second part is due to thermal nuclear motions that can be classical or quantum in character. Here, to simplify the discussion, we consider only the classical spectrum of nuclear fluctuations resulting in the classical field of nuclear polarization, $\mathcal{P}_{\text{n}}$. Fluctuations of the solvent polarization field are usually well described within the Gaussian approximation,[35] leading to the quadratic solvent Hamiltonian

$$H_{\text{B}} = H_{\text{B}}[\mathcal{P}_{\text{n}}] + H_{\text{B}}[\mathcal{P}_{\text{e}}] = \tfrac{1}{2} \mathcal{P}_{\text{n}} \cdot \chi_{\text{n}}^{-1} \cdot \mathcal{P}_{\text{n}} + \tfrac{1}{2} \left(\omega_{\text{e}}^{-2} \dot{\mathcal{P}}_{\text{e}} \cdot \dot{\mathcal{P}}_{\text{e}} + \mathcal{P}_{\text{e}} \cdot \chi_{\text{e}}^{-1} \cdot \mathcal{P}_{\text{e}} \right) \quad [33]$$

Here, χ_{e} and χ_{n} are the Gaussian response functions of the electronic and nuclear solvent polarization, respectively; $\dot{\mathcal{P}}_{\text{e}}$ is the time derivative of the electronic polarization field entering the corresponding kinetic energy term. In terms of the Gaussian solvent model,[35] the nuclear response function is defined through the correlator of corresponding polarization fluctuations (high-temperature limit of the fluctuation–dissipation theorem[36])

$$\chi_{\text{n}}(\mathbf{r} - \mathbf{r}') = \beta \langle \delta P_{\text{n}}(\mathbf{r}) \, \delta P_{\text{n}}(\mathbf{r}') \rangle \qquad [34]$$

In Eq. [33], ω_{e} denotes a characteristic frequency of the optical excitations of the solvent. The kinetic energy of the nuclear polarization $\mathcal{P}_{\text{n}}$ is left out in

Eq. [33] according to the assumption of the classical character of this collective mode. Depending on the form of the coupling of the electron donor–acceptor subsystem to the solvent field, one may consider linear or nonlinear solvation models. The coupling term $-\mathcal{E}_i \cdot \mathcal{P}$ in Eq. [32] represents the linear coupling model (L model) that results in a widely used linear response approximation.[37] Some general properties of the bilinear coupling (Q model) are discussed below.

Equations [32] and [33] represent the system Hamiltonian that can be used to build the CT free energy surfaces. According to the general scheme outlined above, the first step in this procedure is to take the average over the electronic degrees of freedom of the system. This implies integrating over the electronic polarization $\mathcal{P}_e$ and the fermionic populations $a_i^+ a_i$. The trace Tr_{el} can be taken exactly, resulting in two instantaneous energies[38]

$$E_{\pm}[\mathcal{P}_n] = \tilde{I}_{av}[\mathcal{P}_n] \pm \tfrac{1}{2}\sqrt{(\Delta\tilde{I}[\mathcal{P}_n])^2 + 4(H_{ab}^{eff}[\mathcal{P}_n])^2} \qquad [35]$$

where $\tilde{I}_{av} = (\tilde{I}_a + \tilde{I}_b)/2$ and $\Delta\tilde{I} = \tilde{I}_b - \tilde{I}_a$. For $i = a, b$

$$\tilde{I}_i[\mathcal{P}_n] = I_i - \mathcal{E}_i \cdot \mathcal{P}_n - \tfrac{1}{2}(\mathcal{E}_i \cdot \chi_e \cdot \mathcal{E}_i + \mathcal{E}_{12} \cdot \chi_e \cdot \mathcal{E}_{12}) \qquad [36]$$

The effective ET matrix element has the form

$$\tilde{H}_{ab}^{eff}[\mathcal{P}_n] = e^{-S_e/2}[H_{ab} - \mathcal{E}_{ab} \cdot \mathcal{P}_n - \mathcal{E}_{av} \cdot \chi_e \cdot \mathcal{E}_{ab}] \qquad [37]$$

with $\mathcal{E}_{av} = (\mathcal{E}_a + \mathcal{E}_b)/2$. The matrix element $\tilde{H}_{ab}^{eff}[\mathcal{P}_n]$ depends on the solvent through two components: (1) interaction of the off-diagonal solute electric field with the nuclear solvent polarization (second term) and (2) solvation of the off-diagonal field by the electronic polarization of the solvent (third term). The former component leads to solvent-induced fluctuations of the ET matrix element, which represent a non-Condon effect[39] of the dependence of electron coupling on nuclear degrees of freedom of the system. This effect is commonly neglected in the Condon approximation employed in treating nonadiabatic ET rates.[11]

Equation [37] is derived within the assumption that both the electronic polarization and the donor–acceptor complex are characterized by quantum excitation frequencies,[38] $\beta\hbar\omega_e \gg 1$, $\beta\Delta E_{12} \gg 1$, where $\Delta E_{12} = E_2 - E_1$ is the gas-phase adiabatic energy gap in Eq. [29]. The derivation does not assume any particular separation of these two characteristic time scales. The traditional formulation[27] assumes $\Delta E_{12} \ll \hbar\omega_e$ that eliminates the electronic Franck–Condon factor $\exp(-S_e/2)$ in Eq. [37]. The parameter[38,40]

$$S_e = \Delta\mathcal{E}_{ab} \cdot \chi_e \cdot \Delta\mathcal{E}_{ab}/2\hbar\omega_e \qquad \Delta\mathcal{E}_{ab} = \mathcal{E}_b - \mathcal{E}_a \qquad [38]$$

is, however, small for the usual conditions of CT reactions and will be neglected throughout the discussion below.

The energies $E_{\pm}[\mathcal{P}_n]$ in Eq. [35] depend on the nuclear solvent polarization that serves as a three-dimensional (3D) nuclear reaction coordinate driving electronic transitions. The two-state model actually sets up two directions: the vector of the differential field $\Delta\mathcal{E}_{ab}$ and the off-diagonal field $\mathcal{E}_{ab}$. Therefore, only two projections of $\mathcal{P}_n$ need to be considered: the longitudinal field parallel to $\Delta\mathcal{E}_{ab}$ and the transverse field perpendicular to $\Delta\mathcal{E}_{ab}$. In the case when the directions of the differential and off-diagonal fields coincide, one needs to consider only the longitudinal field, and the theory can be formulated in terms of the scalar reaction coordinate

$$Y^d = \Delta\mathcal{E}_{ab} \cdot \mathcal{P}_n \qquad [39]$$

The superscript "d" in the above equations refers to "diabatic" since the diabatic basis set is used to define the electric field difference $\Delta\mathcal{E}_{ab}$. The corresponding free energy profile is obtained by projecting the nuclear polarization $\mathcal{P}_n$ on the direction of the solute field difference

$$e^{-\beta F_{\pm}(Y^d)} = \int \mathcal{D}\mathcal{P}_n \delta(Y^d - \Delta\mathcal{E}_{ab} \cdot \mathcal{P}_n) e^{-\beta E_{\pm}[\mathcal{P}_n]} \qquad [40]$$

where $\mathcal{D}\mathcal{P}_n$ denotes a functional integral[41] over the field $\mathbf{P}_n(\mathbf{r})$.

The integration in Eq. [40] generates the upper and lower CT free energy surfaces that, after the shift in the reaction coordinate $Y^d \rightarrow Y^d + \Delta\mathcal{E}_{ab} \cdot \chi_n \cdot \mathcal{E}_{av}$, take the following form[42]

$$F_{\pm}(Y^d) = \frac{(Y^d)^2}{4\lambda^d} \pm \frac{\Delta E(Y^d)}{2} + C \qquad [41]$$

with

$$\Delta E(Y^d) = [(\Delta F_0^d - Y^d)^2 + 4(H_{ab} + \alpha_{ab}(\Delta F_s^d - Y^d))^2]^{1/2} \qquad [42]$$

and

$$C = \frac{F_{0a}^d + F_{0b}^d}{2} + \frac{\lambda^d}{4} \qquad [43]$$

The constant α_{ab} in Eq. [42] represents the ratio of the collinear difference and off-diagonal fields of the solute

$$\alpha_{ab} = \mathcal{E}_{ab}/\Delta\mathcal{E}_{ab} \qquad [44]$$

The diabatic solvent reorganization energy is defined by the nuclear response function χ_n and by the diabatic field difference

$$\lambda^d = \tfrac{1}{2}\Delta\mathcal{E}_{ab} \cdot \chi_n \cdot \Delta\mathcal{E}_{ab} \tag{45}$$

The free energy gap

$$\Delta F_0^d = F_{0b}^d - F_{0a}^d = \Delta I_{ab} + \Delta F_s^d \tag{46}$$

is composed of the gas-phase splitting $\Delta I_{ab} = I_b - I_a$ and the solvation free energy

$$\Delta F_s^d = -\mathcal{E}_{av} \cdot \chi \cdot \Delta\mathcal{E}_{ab} \tag{47}$$

where $\chi = \chi_e + \chi_n$ is the total response function of the solvent.

Projection on the energy gap reaction coordinate in Eq. [18] is simple to perform for the scalar reaction coordinate Y^d

$$\mathrm{FCWD}_{\pm}^s(X) = \sum_k (\beta Q_{\pm}\Delta E'[Y^{(k)}])^{-1} e^{-\beta E_{\pm}[Y^{(k)}]} \tag{48}$$

where

$$Q_{\pm} = \int e^{-\beta E_{\pm}(Y^d)} dY^d \tag{49}$$

and $Y^{(k)}$ are all the roots of the equation

$$X = \Delta E[Y^d] \tag{50}$$

In Eq. [48], $\Delta E'[Y^{(k)}]$ denotes the derivative

$$\Delta E'\left[Y^{(k)}\right] = \left.\frac{d\Delta E(Y^d)}{dY^d}\right|_{Y^d=Y^{(k)}} \tag{51}$$

where $Y^d = Y^{(k)}$ indicates that the derivative is taken at the coordinate $Y^{(k)}$ obtained as a solution of Eq. [50].

Equations [41]–[50] provide an exact solution for the CT free energy surfaces and Franck–Condon factors of a two-state system in a condensed medium with quantum electronic and classical nuclear polarization fields. The derivation does not make any specific assumptions about the off-diagonal matrix elements of the Hamiltonian. It, therefore, includes the off-diagonal

solute–solvent coupling through the off-diagonal matrix element of the electric field of the solute.[40] This coupling represents a non-Condon dependence of the ET matrix element on the nuclear solvent polarization (this contribution is commonly neglected in MH theory[13]). In the case of weak electronic overlap, all off-diagonal matrix elements are neglected in the free energy surfaces, and the above equations are transformed to the well-known case of two intersecting parabolas (Figure 2) representing the diabatic ET free energy surfaces

$$F_i(Y^d) = F_{0i} + \frac{(Y^d \pm \lambda^d)^2}{4\lambda^d} \qquad [52]$$

The reaction rate constant is then given by the Golden Rule perturbation expansion in the solvent-dependent ET matrix element $H_{ab}^{eff}[\mathcal{P}_n]$.[43] Careful account for non-Condon solvent dependence of the ET matrix element generates the Mulliken-Hush matrix element in the rate preexponent (see below). In the opposite case of strong electronic overlap, the off-diagonal matrix elements cannot be neglected, and one should consider the CT free energy surfaces, instead of ET free energy surfaces, with partial transfer of the electronic density. The free energy surfaces are then substantially nonparabolic; we discuss this case in the section on Electron Delocalization Effect.

Heterogeneous Discharge

The diabatic two-state representation for homogeneous CT can be extended to heterogeneous CT processes between a reactant in a condensed-phase solvent and a metal electrode. The system Hamiltonian is then given by the Fano–Anderson model[44,45]

$$H = H_B + [E - \mathcal{D}_e \cdot \mathcal{P}_n]c^+ c + \sum_{\mathbf{k}} \epsilon_{\mathbf{k}} c_{\mathbf{k}}^+ c_{\mathbf{k}} + \sum_{\mathbf{k}}(H_{\mathbf{k}} c_{\mathbf{k}}^+ c + \text{h.c.}) \qquad [53]$$

where $\mathbf{k}$ is the lattice reciprocal vector, the two summations are over the wave vectors of the electrons of a metal, $\epsilon_{\mathbf{k}}$ is the kinetic energy of the conduction electrons (hence $\epsilon_{\mathbf{k}} = \mathbf{k}^2/2m_e$, with m_e being the electron mass), and "h.c." designates the corresponding Hermetian conjugate. In Eq. [53], c^+ and c are the Fermionic creation and annihilation operators of the localized reactant state. $c_{\mathbf{k}}^+$ and $c_{\mathbf{k}}$ are the creation and annihilation operators, respectively, for a conduction electron with momentum $\mathbf{k}$, and $H_{\mathbf{k}}$ is the coupling of this metal state to the localized electron state on the reactant. The energy of the localized reactant state includes solvation by the solvent electronic polarization (included in E) and the interaction of the electron electric field $\mathcal{D}_e$ with the nuclear solvent polarization $\mathcal{P}_n$. The transferred electron is much faster than the ions dissolved in the electrolyte. Therefore, on the time scale of charge

redistribution, no screening of the electron field by rearrangement of the electrolyte ions occurs, and the electron field includes the field of the image charge on the metal surface

$$\mathbf{D}_e(\mathbf{r}) = e \int |\Psi_e(\mathbf{r}')|^2 \nabla \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} - \frac{1}{|\mathbf{r} - \mathbf{r}'_{im}|} \right) d\mathbf{r}' \qquad [54]$$

where $\mathbf{r}'_{im}$ is the mirror image of the electron at the point $\mathbf{r}'$ relative to the electrode plane, $\Psi_e(\mathbf{r})$ is the wave function of the localized electron, and e is the electron charge. (In Eq. [54], e appears because we are not using atomic units. Thoughout this chapter, the energies are generally in electron volts.) The off-diagonal solute–solvent coupling is dropped in the off-diagonal part of the system Hamiltonian in Eq. [53] as no experimental or theoretical information is currently available about the strength of the off-diagonal solute field in the near-to-electrode region.

The free energy surface for the electron heterogeneous discharge can be directly written as

$$e^{-\beta F(Y^d)} = (\beta Q_B)^{-1} \mathrm{Tr}_n \mathrm{Tr}_{el} [\delta(Y^d - \mathcal{D}_e \cdot \mathcal{P}_n) \hat{\rho}] \qquad [55]$$

where Q_B refers to the partition function of the pure solvent and the Dirac delta function is invoked. In electrochemical discharge, the reactant is coupled to a macroscopic bath of metal electrons. The total number of electrons in the system is thus not conserved, and the grand canonical ensemble should be considered for the electronic subsystem. The density matrix in Eq. [55] then reads

$$\hat{\rho} = e^{\beta(\mu_e N - H)} \qquad [56]$$

Here, μ_e is the chemical potential of the electronic subsystem containing

$$N = c^+ c + \sum_{\mathbf{k}} c_{\mathbf{k}}^+ c_{\mathbf{k}} \qquad [57]$$

electrons.

The path-integral formulation of the trace in Eq. [55] allows us to take it exactly. This leads to the following expression for the free energy surface[46]

$$F(Y^d) = \frac{(Y^d)^2}{4\lambda^d} + \frac{\epsilon(Y^d)}{2} + \beta^{-1} \ln \left[\left| \Gamma \left(\frac{\beta \tilde{\Delta}}{2\pi} - i \frac{\beta \epsilon(Y^d)}{2\pi} \right) \right|^2 \right] \qquad [58]$$

Here, Y^d is the classical reaction coordinate, $\Gamma(x)$ is the gamma function, and

$$\epsilon(Y^d) = E - \mu_e - Y^d = \lambda^d + e\eta - Y^d \qquad [59]$$

where η is the electrode overpotential. Equation [58] presents the exact solution for the free energy surface of an electrochemical system along the classical reaction coordinate Y^d. It includes the free energy of a classical Gaussian solvent fluctuation (the first term) and the free energy of charge redistribution between the localized reactant state and the continuum of delocalized conduction states of the metal (the second and the third terms). Delocalization effectively proceeds on the range of reaction coordinates given by the effective width

$$\tilde{\Delta} = \Delta + \pi\beta^{-1} \qquad [60]$$

built on the direct electron overlap

$$\Delta = \pi \sum_{\mathbf{k}} \rho_F |H_{\mathbf{k}}|^2 \qquad [61]$$

and the width of the thermal distribution of the conductance electrons on the metal Fermi level ($\pi\beta^{-1}$); ρ_F is the electron density of states of the metal on its Fermi level. In the limit

$$\beta\tilde{\Delta} \gg 1 \qquad [62]$$

Eq. [58] reduces to the free energy

$$F(Y^d) = \frac{(Y^d)^2}{4\lambda^d} + \frac{\epsilon(Y^d)}{\pi}\cot^{-1}\frac{\epsilon(Y^d)}{\tilde{\Delta}} + \frac{\tilde{\Delta}}{2\pi}\ln\left[(\beta\tilde{\Delta})^2 + (\beta\epsilon(Y^d))^2\right] \qquad [63]$$

The overlap $\tilde{\Delta}$ can be replaced by Δ when $\Delta \gg \pi\beta^{-1}$. Equation [63] then leads to the ground-state energy $E(Y^d)$ (zero temperature for the electronic subsystem) often used to describe adiabatic heterogeneous CT.[45]

Equations [58] and [63] indicate an important point concerning the instantaneous energies obtained by tracing out (integrating) the electronic degrees of freedom of the system (Eq. [15]). When the separation of electronic states is much higher than the thermal energy $k_B T$, the free energies can be replaced by energies. This does not happen for heterogeneous discharge where thermal excitations of the conductance electrons lead to entropic effects embodied in the temperature-dependent summand in $\tilde{\Delta}$ (Eq. [60]).

BEYOND THE PARABOLAS

The paradigm of free energy surfaces provides a very convenient and productive conceptual framework to analyze the thermodynamics and dynamics of electronic transitions in condensed phases. It, in fact, replaces the complex dynamics of a quantum subsystem interacting with a many-body

thermal bath with the motion of a classical representative particle over the activation barrier.[47] The MH solution gives the barrier as the vertical gap between the bottom of the initial free energy surface and the intersection point. The problem of finding the activation barrier then reduces to two parameters: the free energy equilibrium gap, ΔF_0, and the classical nuclear reorganization energy, λ_{cl} (Figure 2). From a broader perspective, as surprising as it seems, the MH model for classical nuclear modes and its extension to quantum intramolecular skeletal vibrations[17] presents the only exact, closed-form solution for $F_i(X)$ available currently in the field of ET.

The success of the MH theory can also, to a large degree, be attributed to the fact that the parameters of the model are connected to spectroscopic observables. The first spectral moments for absorption and emission transitions $\nu_{abs/em}$ fully define the classical reorganization energy λ_{cl} and the equilibrium free energy gap ΔF_0 through the mean energy and the Stokes shift (Eqs. [6] and [8])

$$h\Delta\nu_{st} = h(\nu_{abs} - \nu_{em}) = 2\lambda_{cl} \qquad [64]$$

Clearly, the MH description does not capture all possible complicated mechanisms of ET activation in condensed phases. The general question that arises in this connection is whether we are able to formulate an extension of the mathematical MH framework that would (1) exactly derive from the system Hamiltonian, (2) comply with the fundamental linear constraint in Eq. [24], (3) give nonparabolic free energy surfaces and more flexibility to include nonlinear electronic or solvation effects, and (4) provide an unambiguous connection between the model parameters and spectroscopic observables. In the next section, we present the bilinear coupling model (Q model), which satisfies the above requirements and provides a generalization of the MH model.

It has in fact been anticipated for many years that the CT free energy surfaces may deviate from parabolas. A part of this interest is provoked by experimental evidence from kinetics and spectroscopy. First, the dependence of the activation free energy, F_i^{act}, for the forward ($i = 1$) and backward ($i = 2$) reactions on the equilibrium free energy gap ΔF_0 (ET energy gap law) is rarely a symmetric parabola as is suggested by the Marcus equation,[48] Eq. [9]. Second, optical spectra are asymmetric in most cases[17] and in some cases do not show the mirror symmetry between absorption and emission.[49] In both types of experiments, however, the observed effect is an ill-defined mixture of the intramolecular vibrational excitations of the solute and thermal fluctuations of the solvent. The band shape analysis of optical lines does not currently allow an unambiguous separation of these two effects, and there is insufficient information about the solvent-induced free energy profiles of ET.

Nonlinear solvation (breakdown of assumption 4 in the Introduction) has long been considered as the main possible origin of nonparabolic free

energy surfaces of ET.[33,50–56] It turns out, however, that equilibrium solvation of fixed solute charges in dense liquids is well described within the linear response approximation,[37] which leads to parabolic free energy surfaces. When the distribution of fixed molecular charges changes with excitation, the equilibrium solvation is still linear and deviations from the linear dynamic response are well described by linear solvation with a time-varying force constant of the Gaussian fluctuations of the medium.[57] The situation changes, however, when the model of fixed charges is replaced by a more realistic model of a distributed electronic density that can be polarized by an external field. The solute free energy then gains the energy of self-polarization that is generally quadratic in the field of the condensed environment.[58] When this self-polarization energy changes with electronic transition, the solute–solvent coupling becomes a bilinear function of solvent nuclear modes instead of a linear function incorporated in the MH model of parabolic ET surfaces. This bilinear coupling model (Q model) produces some very generic types of behavior that are substantially different from what is predicted by the MH model. We thus start our discussion of nonparabolic CT surfaces with a general analysis of the Q model.

Bilinear Coupling Model

The MH description is isomorphic to the two-state (TS) model with a linear coupling of the solute to a classical harmonic oscillator (L model). Since the earliest days of the theory of radiationless transitions, a possibility of a bilinear solute–solvent coupling (Q model) has been anticipated.[38,59,60] This problem can be interpreted as a TS solute linearly coupled to a harmonic solvent mode with force constants different in the initial and final electronic states (Duschinsky rotation of normal modes[38]). Although a general quantum solution of the Q model exists,[59] no closed-form, analytical representation for $F_i(X)$ was given. The model hence has not received wide application to ET reactions. Instead, nonlinear solute–solvent coupling has been modeled by two displaced free energy parabolic surfaces $F_i(X)$ with different curvatures.[50,53] This approach, advanced by Kakitani and Mataga,[50] was designed to represent nonlinear solvation effects on the ET energy gap law. However, the approximation of the ET energy surfaces by two displaced parabolas with different curvatures suffers from a general drawback of not complying with the exact linear relationship between the free energy surfaces in Eq. [24].

The Q model allows an exact formulation for $F_i(X)$ for classical solvent modes.[61] The instantaneous energy in this case is given by the bilinear form

$$E_i(q) = I_i - C_i q + \tfrac{1}{2}\kappa_i q^2 \qquad [65]$$

where q is a collective nuclear mode driving electron transitions (the longitudinal projection of the nuclear polarization $\mathcal{P}_n$ is an example of such a

collective mode). In Eq. [65], both the linear coupling constant, C_i, and the *harmonic force constant*, κ_i, change with the transition. The MH L model is recovered when $\kappa_1 = \kappa_2$. Note, that since the off-diagonal matrix elements of the Hamiltonian are excluded from consideration, the formalism described here may apply to any choice of wave functions for which such an approximation is warranted. We therefore do not specify the basis set here, and the indices $i = 1, 2$ refer to any basis set in which the energies $E_i(q)$ are obtained.

The calculations of the diabatic (no off-diagonal matrix elements) free energy surfaces in Eq. [18] can be performed exactly for $E_i(q)$ given by Eq. [65]. This procedure yields the closed-form, analytical expressions for the free energies $F_i(X)$. It turns out that the solution exists only in a limited, one-sided band of the energy gaps X.[61] Specifically, an asymptotic expansion of the exact solution leads to a simple expression for the free energy

$$F_i(X) = F_{0i} + \left(\sqrt{|\alpha_i||X - X_0|} - |\alpha_i|\sqrt{\lambda_i} \right)^2 \qquad [66]$$

within a one-sided band of reaction coordinate X and

$$F_i(X) = \infty \qquad [67]$$

outside the band.

The parameter X_0 establishes the boundary of the energy gaps for which a finite solution $F_i(X)$ exists. The band definition and its boundary

$$X_0 = \Delta I - \frac{\Delta C^2}{2\Delta\kappa} \qquad [68]$$

both depend on the sign of the variation of the force constant $\Delta\kappa$. The one-sided band is defined as (Figure 5):

$$\text{fluctuation band} = \begin{cases} X < X_0 & \text{at} \quad \Delta\kappa < 0 \\ X > X_0 & \text{at} \quad \Delta\kappa > 0 \end{cases} \qquad [69]$$

Figure 5 Upper energy ($\Delta\kappa > 0$) and lower energy ($\Delta\kappa < 0$) fluctuations boundaries in the Q model.

This result indicates a fundamental distinction between the Q and L models. In the latter, the band of the energy gap fluctuations is not limited, leading to a finite, even small, probability to find a fluctuation of any magnitude of the energy gap. On the contrary, the Q model suggests a limited band for the energy gap fluctuations. The gap magnitudes achievable due to the nuclear fluctuations are limited by a low-energy boundary for $\Delta\kappa > 0$ and by a high-energy boundary for $\Delta\kappa < 0$. The probability of finding an energy gap fluctuation outside these boundaries is identically zero because there is no real solution of the equation

$$X = \Delta E(q) = E_2(q) - E_1(q) \tag{70}$$

The absence of a solution is the result of a bilinear dependence of the energy gap $\Delta E(q)$ on the driving nuclear mode q (Figure 6).

The other model parameters entering Eq. [66] are the nuclear reorganization energies defined through the second cumulants of the reaction coordinate

$$\lambda_i = \tfrac{1}{2}\beta\langle(\delta X)^2\rangle_i = \tfrac{1}{2\kappa_i}(C_i/\alpha_i - \Delta C)^2 \tag{71}$$

and the relative changes in the force constants

$$\alpha_i = \frac{\kappa_i}{\Delta\kappa} \tag{72}$$

The two sets of parameters defined for each state are not independent because of the following connections between them
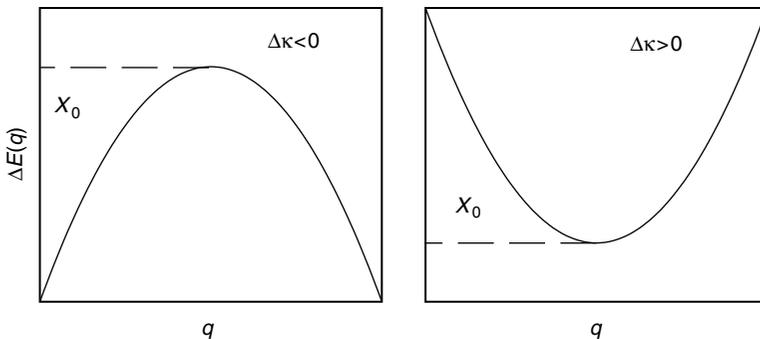
$$\alpha_1^3\lambda_1 = \alpha_2^3\lambda_2 \tag{73}$$

Figure 6 The origin of the upper energy ($\Delta\kappa < 0$) and lower energy ($\Delta\kappa > 0$) fluctuation boundaries due to a bilinear dependence of ΔE on q.

and

$$\alpha_2 = 1 + \alpha_1 \qquad [74]$$

An additional constraint on the magnitudes of the parameter α_1 comes from the condition of the thermodynamic stability of the collective solvent mode in both states, $\kappa_i > 0$, resulting in two inequalities

$$\alpha_1 > 0 \qquad \text{or} \qquad \alpha_1 < -1 \qquad [75]$$

The inequalities in Eq. [75] also define the condition for the generating function (Eq. [23]) to be analytic in the integration contour in Eq. [25]. This condition is equivalent to the linear connection between the diabatic free energy surfaces, Eq. [24]. The Q model solution thus explicitly indicates that the linear relation between the diabatic free energy surfaces is equivalent to the condition of thermodynamic stability of the collective nuclear mode driving ET.

Equations [73] and [74] reduce the number of independent parameters of the Q model to three: ΔF_0, λ_1, and α_1. Here, ΔF_0 (Eq. [21]) is the free energy gap between equilibrium configurations of the system (Figure 2). The fluctuation boundary X_0 is connected to ΔF_0 by the relation

$$X_0 = \Delta F_0 + \lambda_1 \alpha_1^2 / \alpha_2 \qquad [76]$$

Compared to the two-parameter MH theory (λ and ΔF_0),[12] the Q model introduces an additional flexibility in terms of the relative variation of the fluctuation force constant through α_1. The MH theory is recovered in the limit $\alpha_1 \to \infty$.

Importantly, the new free energy surfaces lead to qualitatively new features for the activated ET kinetics. The standard high-temperature limit of two diabatic ET free energy surfaces

$$F_i(X) = F_{0i} + \frac{(X - \Delta F_0 \mp \lambda_i)^2}{4\lambda_i} \qquad [77]$$

is reproduced when $\alpha_i \gg 1$ (the driving mode force constants κ_i in the two states are similar) and, additionally, $|X - \Delta F_0 \mp \lambda_i| \ll |\alpha_i|\lambda_i$. Here, "$-$" and "$+$" correspond to $i = 1$ and $i = 2$, respectively. The second requirement implies that the reaction coordinate should be not too far from the free energy minimum to preserve its parabolic form. By contrast, in the limit $|X - X_0| \gg \lambda_i |\alpha_i|$, the linear dependence wins over the parabolic law

$$F_i(X) = F_{0i} + |\alpha_i| \left| X - \Delta F_0 + \lambda_1 \frac{\alpha_1^2}{\alpha_2} \right| \qquad [78]$$
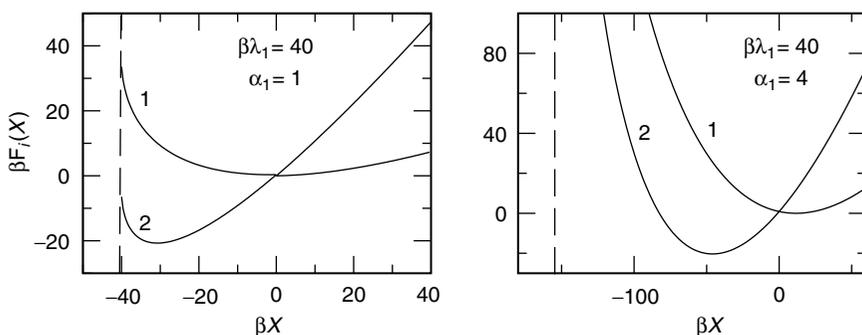
Figure 7 The free energy surfaces $F_1(X)$ (1) and $F_2(X)$ (2) at various α_1; $\Delta I = 0$. The dashed line indicates the position of the fluctuation boundary X_0.

As a combination of these two effects, plus the existence of the fluctuation boundary, the free energy surfaces are asymmetric with a steeper branch on the side of the fluctuation boundary X_0. The other branch is less steep tending to a linear dependence at large X (Figure 7). The minima of the initial and final free energy surfaces get closer to each other and to the band boundary with decreasing α_1 and λ_1. The crossing point then moves to the inverted ET region where the free energies are nearly linear functions of the reaction coordinate.

The ET activation energy follows from Eq. [66]

$$F_i^{\text{act}} = F_i(0) - F_{0i}$$

$$= |\alpha_i| \left(\sqrt{|\Delta F_0 - \lambda_1 \alpha_1^2 / \alpha_2|} - \sqrt{|\alpha_i| \lambda_i} \right)^2 \qquad [79]$$

Equation [79] produces the MH quadratic energy gap law at small $|\Delta F_0| \ll |\alpha_1 \lambda_1|$ and yields a linear dependence of the activation energy on the equilibrium free energy gap at $|\Delta F_0 - \lambda_1 \alpha_1^2 / \alpha_2| \gg |\alpha_i| \lambda_i$.

A linear energy gap law is by no means unusual in ET kinetics. It is quite often encountered at large equilibrium energy gaps. Experimental observations of the linear energy gap law are made for intermolecular[62] as well as intramolecular[63] organic donor–acceptor complexes, in binuclear metal–metal CT complexes,[16] and in CT crystals.[64] It is commonly explained in terms of the weak coupling limit of the theory of vibronic band shapes yielding the linear-logarithmic dependence proportional to $\Delta F_i \ln \Delta F_i$ on the vertical energy gap ΔF_i.[17] On the contrary, a strictly linear dependence proportional to ΔF_i arises from the Q model.

To complete the Q model, one needs to relate the model parameters to spectral observables. Already, the reorganization energies λ_i are directly related to the solvent-induced inhomogeneous widths of absorption ($i = 1$)

and emission ($i = 2$)

$$\lambda_i = \tfrac{1}{2}\beta h^2 \langle \delta v^2 \rangle_i = \tfrac{1}{2}\beta \sigma_i^2 \qquad [80]$$

where the Gaussian spectral width σ_i is experimentally defined through the half-intensity width Γ_i as

$$\sigma_i^2 = \Gamma_i^2/(8\ln 2) \qquad [81]$$

As is easy to see from Eq. [80] and Figure 7, the Q model predicts the breaking of the symmetry between the absorption and emission widths (Eq. [11]) generated by a statistical distribution of solvent configurations around a donor–acceptor complex (inhomogeneous broadening). This fact may have a significant application to the band shape analysis of optical transitions since unequal absorption and emission width are often observed experimentally.[65,66]

    The parameter α_1 is defined through the Stokes shift and two reorganization energies from optical widths

$$\alpha_1 = -\Delta\lambda^{-1}(h\Delta v_{\text{st}} + \lambda_2) \qquad \Delta\lambda = \lambda_2 - \lambda_1 \qquad [82]$$

Similarly, the equilibrium energy gap is (cf. to Eq. [8])

$$\Delta F_0 = h v_{\text{m}} - \frac{\lambda_1}{2}\frac{\alpha_1}{\alpha_2^2} \qquad [83]$$

which is equivalent to

$$\Delta F_0 = h v_{\text{m}} + \frac{\lambda_1 \Delta\lambda}{2}\frac{h\Delta v_{\text{st}} + \lambda_2}{(h\Delta v_{\text{st}} + \lambda_1)^2} \qquad [84]$$

The Stokes shift and two second spectral moments fully define the parameters of the model. In addition, they should satisfy Eqs. [73] and [74]. The latter feature establishes the condition of model consistency that is important for mapping the model onto condensed-phase simulations that we discuss below.

    The connection of the model parameters to the first and second spectral cumulants enables one to build global, nonequilibrium free energy surfaces of ET based on two cumulants obtained at equilibrium configuration of the system. This allows one to apply the model to equilibrium computer simulations data or to spectral modeling. Compared to the MH picture of intersecting parabolas, the Q model predicts a more diverse pattern of possible system regimes including (1) an existence of a one-sided band restricting the range of permissible reaction coordinates, (2) singular free energies outside the fluctuation band, and (3) a linear energy gap law at large activation barriers. The main features of the Q and L models are compared in Table 1.

Table 1 Main Features of the Two-Parameter L Model (MH) and the Three-Parameter Q Model

	L Model	Q Model
Parameters	ΔF_0, λ	ΔF_0, λ_1, α_1
Reaction coordinate	$-\infty < X < \infty$	$X > X_0$ at $\alpha_1 > 0$
		$X < X_0$ at $\alpha_1 < 0$
Spectral moments	$\Delta F_0 = h\nu_m$	$\Delta F_0 = h\nu_m - [\lambda_1\alpha_1/2(1+\alpha_1)^2]$
	$\lambda = \frac{1}{2}h\Delta\nu_{st}$	$\lambda_1 = \frac{1}{2}\beta h^2\langle(\delta\nu)^2\rangle_1$
		$\alpha_1 = (h\Delta\nu_{st} + \lambda_2)/(\lambda_1 - \lambda_2)$
Energy gap law		
$\Delta F_0 + \lambda_1 \ll \lambda_1$	$F_1^{act} \propto (\Delta F_0 + \lambda)^2$	$F_1^{act} \propto (\Delta F_0 + \lambda_1)^2$
$\lvert\Delta F_0\rvert \gg \lambda_1$	$F_1^{act} \propto \Delta F_0^2$	$F_1^{act} \propto \lvert\Delta F_0\rvert$

Electron Transfer in Polarizable Donor–Acceptor Complexes

The mathematical model incorporating the bilinear solute–solvent coupling considered above can be realized in various situations involving non-linear interactions of the CT electronic subsystem with the condensed-phase environment. The most obvious reason for such effects is the coupling of the two states participating in the transition to other excited states of the donor–acceptor complex. These effects bring about polarizability and electronic delocalization in CT systems. The instantaneous energies obtained for a two-state donor–acceptor complex contain a highly nonlinear dependence on the solvent field through the instantaneous adiabatic energy gap. Expansion of the energy gap in the solvent field truncated after the second term generates a state-dependent bilinear solute–solvent coupling characteristic of the Q model. The second derivative of the energy in the external field is the system's polarizability. It is therefore hardly surprising that models incorporating the polarizability of the solute[67] turn out to be isomorphic to the Q model.[61] Here, we focus on some specific features of polarizable CT systems.

The common starting point to build a theoretical description of the thermodynamics and dynamics of the condensed environment response to an electronic transition is to assume that the transition alters the long-range solute–solvent electrostatic forces. This change comes about due to the variation of the electronic density distribution caused by the transition. The combined electron and nuclear charge distributions are represented by a set of partial charges that are assumed to change when the transition occurs. Actually, a change in the electronic state of a molecule changes not only the electronic charge distribution, but also the ability of the electron cloud to polarize in the external field. In other words, the set of transition dipoles to other electronic states is individual for each state of the molecule, and the dipolar (and higher order) polarizability changes with the transition.

Optical excitations quite often generate considerable changes in fixed partial charges, usually described in terms of the difference solute dipole Δm_0 ("0" refers here to the solute). Chromophores with high magnitudes of the ratio $\Delta m_0/R_0^3$, where R_0 is the effective solute radius, are often used as optical probes of the local solvent structure and solvation power.[68] High polarizability changes are also quite common for optical chromophores,[69] as is illustrated in Table 2. Naturally, the theory of ET reactions and optical transitions needs extension for the case when the dipole moment and polarizability both vary with electronic transition:

$$\mathbf{m}_{01} \rightarrow \mathbf{m}_{02} \qquad \alpha_{01} \rightarrow \alpha_{02} \qquad [85]$$

To derive the instantaneous free energies E_i, one needs an explicit model for a dipolar polarizable solute in a dipolar polarizable solvent. This need is addressed by the Drude model for induced solute and solvent dipole moments.[70] The Drude model represents the induced dipoles as fluctuating vectors: $\mathbf{p}_j$ for the solvent molecules and $\mathbf{p}_0$ for the solute. The potential energy of creating a fluctuating induced dipole $\mathbf{p}$ is given by that of a harmonic oscillator, $\mathbf{p}^2/2\alpha$, with the polarizability α appearing as the oscillator mass. The system Hamiltonian H_i is the sum of the solvent–solvent, H_{ss}, and solute–solvent, $H_{0s}^{(i)}$, parts, giving

$$H_i = H_{0s}^{(i)} + H_{ss} \qquad [86]$$

In H_i, the permanent and induced dipoles add up resulting in the solute–solvent and solvent–solvent Hamiltonians in the form

$$H_{0s}^{(i)} = I_i + U_{0s}^{\text{rep}} - \sum_j (\mathbf{m}_{0i} + \mathbf{p}_0) \cdot \mathbf{T}_{0j} \cdot (\mathbf{m}_j + \mathbf{p}_j) + (1/2\alpha_{0i})[\omega_0^{-2}\dot{\mathbf{p}}_0^2 + \mathbf{p}_0^2] \qquad [87]$$

Table 2 Ground-State Polarizability (α_1) and Trace of the Tensor of Polarizability Variation $(1/3)\text{Tr}[\Delta\alpha]$ for Several Optical Dyes and Charge Transfer Complexes

Chromophore	$\alpha_1/\text{Å}^3$	$(1/3)\text{Tr}[\Delta\alpha]/\text{Å}^3$
Anthracene	25	17
2,2′-Bipyridine-3,3′diol	21	11
Bis(adamantylidene)	42	29
1-Dimethylamino-2,6-dicyano-4-methylbenzene	22	35
Tetraphenylethylene	50	38
$[(NC)_5Fe^{II}CNOs^{III}(NH_3)_5]^-$		57
$[(NC)_5Os^{II}CNRu^{III}(NH_3)_5]^-$		(190) 317[a]

[a]For two different CT transitions.

and

$$H_{ss} = U_{ss}^{\text{rep}} - \frac{1}{2}\sum_{j,k}(\mathbf{m}_j + \mathbf{p}_j) \cdot \tilde{\mathbf{T}}_{jk} \cdot (\mathbf{m}_k + \mathbf{p}_k) + \left(\frac{1}{2\alpha}\right)\sum_j [\omega_e^{-2}\dot{\mathbf{p}}_j^2 + \mathbf{p}_j^2] \qquad [88]$$

Here, $\mathbf{T}_{jk}$ is the dipole–dipole interaction tensor, and $\tilde{\mathbf{T}}_{jk} = \mathbf{T}_{jk}(1 - \delta_{jk})$; U_{0s}^{rep} and U_{ss}^{rep} stand for repulsion potentials, and $\omega_0 = \Delta E_{12}/\hbar$, where ΔE_{12} is the adiabatic gas-phase energy gap (Eq. [29]).

   The statistical average over the electronic degrees of freedom in Eq. [15] is equivalent, in the Drude model, to integration over the induced dipole moments $\mathbf{p}_0$ and $\mathbf{p}_j$. The Hamiltonian H_i is quadratic in the induced dipoles, and the trace can be calculated exactly as a functional integral over the fluctuating fields $\mathbf{p}_0$ and $\mathbf{p}_j$.[39,67] The resulting solute–solvent interaction energy is[67]

$$E_{0s,i} = I_i + U_{0s}^{\text{rep}} + U_{0s,i}^{\text{disp}} - a_e f_{ei}\mathbf{m}_{0i}^2 - f_{ei}\mathbf{m}_{0i} \cdot \mathbf{R}_p - \frac{1}{2}\alpha_{0i}f_{ei}\mathbf{R}_p^2 \qquad [89]$$

Here, $\mathbf{R}_p$ is the reaction field of the solvent nuclear subsystem, and the factor

$$f_{ei} = [1 - 2a_e\alpha_{0i}]^{-1} \qquad [90]$$

describes an enhancement of the condensed-phase solute dipole and polarizability by the self-consistent field of the electronic polarization of the solvent.

   For the statistical average over the nuclear configurations, generating the distribution over the solute energy gaps (Eq. [18]), one needs to specify the fluctuation statistics of the nuclear reaction field $\mathbf{R}_p$. A Gaussian statistics of the field fluctuations[35] implies using the distribution function

$$P(\mathbf{R}_p) = \left(4\pi a_p k_B T\right)^{-1/2}\exp[-\beta\mathbf{R}_p^2/4a_p] \qquad [91]$$

where a_p is the response coefficient of the nuclear solvent response. Combined with the Gaussian function $P(\mathbf{R}_p)$, Eq. [89] is essentially equivalent to the Q model (Eq. [65]). The vector of the nuclear reaction field plays the role of the nuclear collective mode driving activated transitions (q). One can then directly employ the results of the Q model to produce the diabatic free energy surfaces of polarizable donor–acceptor complexes or to calculate the spectroscopic observables.

   The reorganization energies follow from Eq. [71] and take the following form for polarizable chromophores:

$$\lambda_i = (a_p f_i/f_{ei})\left(\Delta\tilde{\mathbf{m}}_0 + 2a_p f_i\Delta\tilde{\alpha}_0\mathbf{m}_{0i}\right)^2 \qquad [92]$$

The parameter f_{ei} is defined by Eq. [90]. It scales the solute dipole moment and the polarizability yielding the effective difference values

$$\Delta \tilde{\mathbf{m}}_0 = f_{e2}\mathbf{m}_{02} - f_{e1}\mathbf{m}_{01} \qquad \Delta \tilde{\alpha}_0 = f_{e2}\alpha_{02} - f_{e1}\alpha_{01} \qquad [93]$$

The parameter

$$f_i = [1 - 2a\,\alpha_{0i}]^{-1} \qquad [94]$$

represents the self-consistent reaction field of the solvent including both the electronic and nuclear polarization components; $a = a_e + a_p$, where a_e is the solvent response coefficient of the solvent electronic polarization. The electronic and total solvent response coefficients can be evaluated from the dielectric cavity or explicit solvent models.[5,71,72] The dielectric continuum estimate for a spherical solute yields

$$a_e = \frac{1}{R_0^3}\frac{\epsilon_\infty - 1}{2\epsilon_\infty + 1} \qquad a = \frac{1}{R_0^3}\frac{\epsilon_s - 1}{2\epsilon_s + 1} \qquad [95]$$

where ϵ_∞ and ϵ_s are the high frequency and static dielectric constants of the solvent. When the solute polarizability is constant, the reorganization energy is the same in both reaction states ($f = f_1 = f_2$; $f_e = f_{e1} = f_{e2}$) and is given by the well-known relation[73]

$$\lambda = (af - a_e f_e)\Delta \mathbf{m}_0^2 \qquad [96]$$

A polarizability change leads to a significant variation of the reorganization energy, which is illustrated in Figure 8, where λ_i are plotted against α_{02}. As can be seen, the reorganization energy approximately doubles with excitation when the excited-state polarizability is about 50% higher than the ground-state value. Such polarizability differences are not uncommon for optical chromophores (Table 2). The effect of the negative polarizability variation is much weaker, and λ_2 is only slightly smaller than λ_1.

From the Q model, the solvent-induced shift of the equilibrium free energy gap $F_{0i} = I_i + \Delta F_{s,i}$ is given by the following relation:

$$\Delta F_{s,i} = -2a_p f_i \left[\Delta \tilde{\mathbf{m}}_0 \cdot \mathbf{m}_{0i} + a_p f_i \Delta \tilde{\alpha}_0\,\mathbf{m}_{0i}^2 \right] \qquad [97]$$

Also, the solvent-induced Stokes shift between the absorption and emission first spectral moments is

$$\begin{aligned} h\Delta\nu_{st} &= h\nu_{abs} - h\nu_{em} \\ &= 2a_p\Delta\tilde{\mathbf{m}}_0 \cdot [f_2\mathbf{m}_{02} - f_1\mathbf{m}_{01}] + 2a_p^2\Delta\tilde{\alpha}_0[(f_2 m_{02})^2 - (f_1 m_{01})^2] \end{aligned} \qquad [98]$$
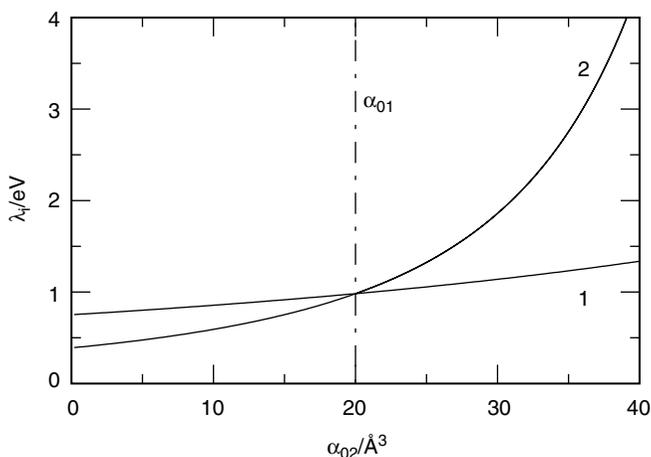
Figure 8 Dependence of the solvent reorganization energy in the neutral (1, m_{01}) and charge-separated (2, m_{02}) states on the polarizability of the final state α_{02}. The solvent response coefficients are estimated from the continuum dielectric model (Eq. [95]). Solute and solvent parameters are $m_{01} = 0$, $m_{02} = 15\,\text{D}$, $\alpha_{01} = 20\,\text{Å}^3$, $R_0 = 4\,\text{Å}$, $\epsilon_\infty = 2$, $\epsilon_s = 30$. In this and subsequent figures, some of the axes are labeled as the ratios shown in order to make the quantities dimensionless. For example, the ordinate in this plot is in units of electron volts, and the abscissa is in units of cubic angstroms.

Both the free energy gap and the Stokes shift include two contributions: one arising from the variation of the solute dipole (the first term) and one due to the polarizability change (the second term). The Stokes shift is hence nonzero even if the charge distribution does not change in the course of the transition ($m_{02} = m_{01}$).

The polarizability difference determines the relative change in the frequency of the solvent driving mode given by the parameter α_1 of the Q model

$$\alpha_1 = -\frac{f_{e1}}{2a_p f_1 \Delta\tilde{\alpha}_0} \qquad [99]$$

The fact that the parameter α_1 is connected to spectroscopic moments for absorption and emission transitions opens an interesting opportunity to derive the polarizability change of optical chromophores from spectroscopic first and second moments. The equation for the polarizability change is as follows:

$$\Delta\tilde{\alpha}_0 = \frac{1}{2\lambda_1} \frac{\Delta\lambda}{h\Delta\nu_{\text{st}} + \lambda_2} \left(\tilde{\mathbf{m}}_{02} - \tilde{\mathbf{m}}_{01} \frac{h\Delta\nu_{\text{st}} + \lambda_1}{h\Delta\nu_{\text{st}} + \lambda_2} \right)^2 \qquad [100]$$

In many practical cases, the factors f_{ei} are very close to unity and can be omitted. The parameters $\tilde{\alpha}_{0i}$ and $\tilde{\mathbf{m}}_{0i}$ are then equal to their gas-phase values α_{0i} and $\mathbf{m}_{0i}$. Equation [100] then gives the polarizability change in terms of spectroscopic moments and gas-phase solute dipoles. Experimental measurement and theoretical calculation of $\Delta\alpha_0 = \alpha_{02} - \alpha_{01}$ is still challenging. Perhaps the most accurate way to measure $\Delta\alpha_0$ presently available is that by Stark spectroscopy,[74–76] which also gives Δm_0. Equation [100] can therefore be used as an independent source of $\Delta\alpha_0$, provided all other parameters are available, or as a consistency test for the band shape analysis.

One of the consequences of a nonzero $\Delta\alpha_0$ is that the relation between the solvent-induced Stokes shift and the corresponding spectral width ($\lambda_v = 0$)

$$h\Delta\nu_{st} = \beta\sigma^2 \tag{101}$$

which is valid for linear solvation response and $\Delta\alpha_0 = 0$, does not hold any more. In Figure 9, the widths $\beta\sigma_i^2$ are plotted versus the Stokes shift obtained by varying the static dielectric constant of the solvent in the range $\epsilon_s = 3-65$. The aborption width deviates downward from the unity slope line predicted by Eq. [101], and the emission width goes upward. The opposite behavior follows from nonlinear solvation effects:[77] the absorption width deviates upward from Eq. [101], and the emission width goes downward. This situation arises because nonlinear solvation results in narrowing of emission lines in contrast to the broadening effect of $\Delta\alpha_0 > 0$. The two effects, therefore, tend to compensate each other for $\Delta\alpha_0 > 0$ and to enforce each other for $\Delta\alpha_0 < 0$.

Both the inequality of the charge separation (CS) and charge recombination (CR) reorganization energies ($\lambda_1 \neq \lambda_2$, Figure 8) and the deviation from
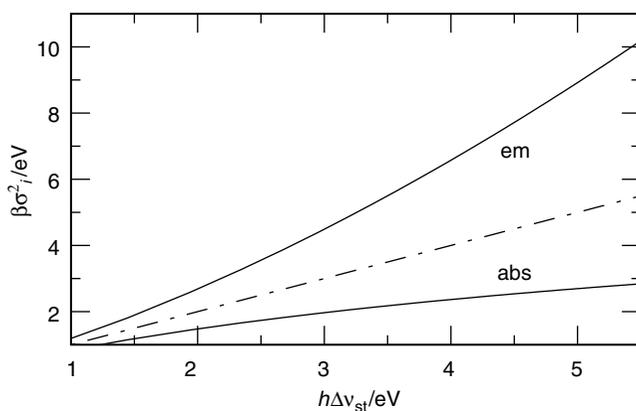

Figure 9 Absorption (abs.) and emission (em.) widths obtained by changing the static solvent dielectric constant in the range $\epsilon_s = 3 - 65$ versus the Stokes shift; $\epsilon_\infty = 2.0$. The dash–dotted line indicates the equality $h\Delta\nu_{st} = \beta\sigma^2$ is valid for $\Delta\alpha_0 = 0$.

the width/Stokes shift relation (Eq. [101], Figure 9) are indicators of a non-parabolic form of the CS and CR free energy surfaces. Another indication of this effect is the energy gap law. The *energy gap law* refers to the dependence of the activation energy of a reaction on the difference in the Gibbs energy between the products and reactants.[34,38,50] The Marcus equation, Eq. [9], is an example of the energy gap law. Experimentally, the energy gap law is monitored by changing the gas-phase component of ΔF_0 through chemical substitution of the donor and/or acceptor units.[48] The solvent component of ΔF_i is usually assumed to be reasonably constant. Figure 10 shows the activation energy of the forward (charge separation, CS) reaction plotted against $\Delta F_{CS} = \Delta F_0$ and backward (charge recombination, CR) reaction plotted against $\Delta F_{CR} = -\Delta F_0$ for the transition $m_{01} = 0 \rightarrow m_{02} = 15$ D and $\alpha_{01} = 20$ Å$^3 \rightarrow \alpha_{02} = 40$ Å^3. Two important effects of nonzero $\Delta\alpha_0$ manifest themselves in Figure 10. First, in contrast to the case of zero $\Delta\alpha_0$, the maxima of the CS and CR curves do not coincide, as is suggested by Eq. [9]. Second, the CR curve is broader and shallower from the side of negative energy gaps compared to the CS curve.

The energy gap law for thermally activated ET reactions is often obtained by superimposing CS and CR data on a common scale of ΔF_0.[78] For such a procedure, depending on the energy range studied, two outcomes can be predicted. For a narrow range of ΔF_{CS} and ΔF_{CR} values close to zero,
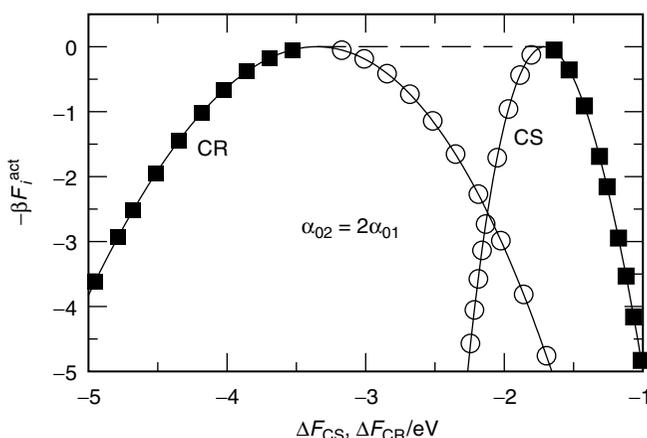

Figure 10 ET energy gap law for the charge separation (CS, $m_{01} \rightarrow m_{02}$, $\Delta F_{CS} = \Delta F_0$) and charge recombination (CR, $m_{02} \rightarrow m_{01}$, $\Delta F_{CR} = -\Delta F_0$) reactions at $\alpha_{01} = 20$ Å^3 and $\alpha_{02} = 40$ Å^3. Parameters are as in Figure 8. The points and dashed line are drawn to illustrate two possible outcomes of combining CS and CR experimental data in one plot with a common energy gap scale (see the text). The open circles correspond to crossing curves, whereas the solid squares correspond to a single curve bridged by the dashed line.

intersection of the two curves (illustrated by circles in Figure 10) may occur. Such a behavior was indeed observed in Ref. 78 for a series of porphyrin–quinone diads in tetrahydrofuran. Maxima of the CS and CR curves get closer to each other with decreasing solvent polarity, and, in fact, no curve crossing was seen for the same systems in benzene as a solvent.[78] When the normal region of CS is combined with the inverted region for CR, another scenario is possible. The two branches (shown by squares in Figure 10) fitted by a single curve (the dashed line in Figure 10) result in a plateau in the energy gap law (a picture reminiscent of this behavior can be seen in Figure 4 of Ref. 79).

Nonlinear Solvation Effects

Experiment provides very limited evidence whether the free energy surfaces of ET should be calculated invoking the linear or nonlinear solvent response. In the absence of direct experimental evidence, the problem of nonlinear solvation effects on the ET free energy surfaces has been approached by computer simulations[33,51–53] and liquid-state solvation theories (integral equations[80] and perturbation techniques[81]). In computer simulations, the free energy surfaces are calculated either directly by umbrella sampling techniques[82] or indirectly by generating a few equilibrium cumulants. In both cases, the lack of a general analytical framework to generate global free energy surfaces from limited data available from simulations considerably impedes the application of the simulation results to generate optical band shapes or to make predictions concerning the ET energy gap law.

The Q model considered above may provide enough flexibility to be used as an analytical background to analyze condensed-phase simulations of the ET energetics. The great advantage of the model is that it requires only two first equilibrium cumulants of the energy gap fluctuations for each electronic state to generate $F_i(X)$ in the whole range of X values in the permissible fluctuation band. The applicability of the model to mapping the simulations can be tested on the consistency requirement given by Eq. [73]. Rewritten in terms of the moments of the reaction coordinate X, this requirement implies that the factor

$$\gamma = \frac{\langle(\delta X)^2\rangle_1}{\langle(\delta X)^2\rangle_2} \left(\frac{\langle(\delta X)^2\rangle_2 + 2k_\mathrm{B}T\Delta\langle X\rangle}{\langle(\delta X)^2\rangle_1 + 2k_\mathrm{B}T\Delta\langle X\rangle} \right)^3 \qquad [102]$$

$(\Delta\langle X\rangle = \langle X\rangle_2 - \langle X\rangle_1)$ should obey the condition

$$\gamma = 1 \qquad [103]$$

Table 3 lists the parameters γ extracted from simulations available in the literature. The condition of Eq. [103] holds very well indeed, which allows one

Table 3 Mapping of the Q Model on Simulation Data for Charge Separation Reactions (Energies are in kcal/mole)

Solvent	$h\Delta\nu_{st}{}^a$	λ_1	λ_2	α_1	γ	Reference
Lattice of point dipoles	157	121.1	48.3	2.82	1.01	54
Lattice of point dipoles	14.3	9.1	5.6	5.6	1.00	61
Dipolar liquid	20.3	10.5	8.7	14	1.01	61
Polar liquid	50	27.0	17.4	7.04	1.04	53
Polar liquid	267	231.5	67.1	2.03	1.04	55
Water	421	164.4	181.2	−35.8	0.99	56

$^a h\Delta\nu_{st} = \langle X \rangle_1 - \langle X \rangle_2.$

to apply the Q model to generating $F_i(X)$ from equilibrium simulations. Figure 11 (left panel) compares the results of the analytical Q model with simulated free energy surfaces for a dipolar solute in a lattice of dipolar hard spheres (DHS) (the two sets of curves coincide on the plot scale). A dipolar lattice as a solvent is chosen because it generates a far larger nonlinear solvation effect than nonpolarizable and polarizable liquids of the same polarity (Figure 11). The parameter

$$\alpha_1 = -\frac{2k_B T \Delta \langle X \rangle + \langle (\delta X)^2 \rangle_2}{\langle (\delta X)^2 \rangle_2 - \langle (\delta X)^2 \rangle_1} \qquad [104]$$

of the Q model serves as an indicator of the strength of nonlinear solvation effects (the linear response is recovered in the limit $\alpha_1 \to \infty$). The right panel
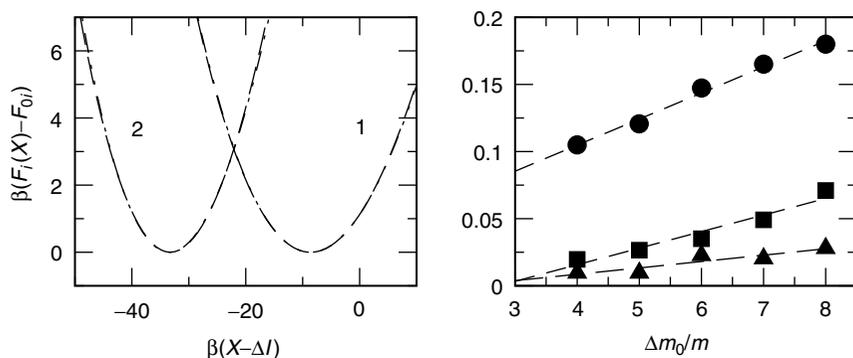

Figure 11 Left panel: $F_1(X)$ (1) and $F_2(X)$ (2) from the analytical Q model (dashed lines) and from simulations (dash–dotted lines) at $m_{01}/m = 2$ and $m_{02}/m = 10$ in the dipolar lattice with $\beta m^2/\sigma^3 = 1.0$; $R_0/\sigma = 0.9$; σ is the hard-sphere diameter of the solvent molecules; m is the solvent dipole moment. The dashed and dash–dotted curves essentially superimpose. Right panel: $1/\alpha_1$ versus $\Delta m_0/m$. Circles indicate the lattice DHS solvent, squares correspond to a liquid DHS solvent, and triangles indicate a nonpolarizable solute in a polarizable DHS liquid; $\beta m^2/\sigma^3 = 1$, $\alpha/\sigma^3 = 0.05$, α is the solvent polarizability.

in Figure 11 shows the dependence of α_1 on the magnitude of solute's dipole. The dipolar lattice demonstrates a considerably higher extent of nonlinear solvation compared to dipolar liquids. The reason for this effect is that the lattice dipoles are immobilized and the orientational saturation is not compensated by local density compression as happens in liquid solvents.[83]

Electron Delocalization Effects

Equations [41]–[42] give a general, exact solution for the free energy surfaces of a two-level system characterized by two collinear vectors: differential, $\Delta\mathcal{E}_{ab}$, and off-diagonal, $\mathcal{E}_{ab}$, electric fields of the donor–acceptor complex. When the off-diagonal matrix elements are nonnegligible, the free energy surfaces are substantially nonparabolic. They are defined by five parameters: λ^d, ΔF_s^d, ΔI_{ab}, H_{ab}, and α_{ab}. A careful choice of the basis set allows the elimination of one parameter. Two approaches can be employed. In the adiabatic basis set, $\{\phi_1, \phi_2\}$, the gas-phase ET matrix element is zero, $H_{12} = 0$. Alternatively, one can define the basis set by demanding the off-diagonal matrix element of the solute electric field be zero, $\alpha_{ab} = 0$. This choice sets up the generalized Mulliken–Hush (GMH) basis.[7] These two approaches are essentially equivalent in terms of building the CT free energy surfaces,[42] but the adiabatic basis may be more convenient for practical applications. The reason is that most quantum chemical software packages are designed to diagonalize the gas-phase Hamiltonian matrix, thus generating the adiabatic basis and corresponding adiabatic matrix elements of the solute electric field.

There are several fundamental reasons why the GMH and adiabatic formulations are to be preferred over the traditionally employed diabatic formulation. The definition of the diabatic basis set is straightforward for *intermolecular* ET reactions when the donor and acceptor units are separated before the reaction and form a donor–acceptor complex in the course of diffusion in a liquid solvent. The diabatic states are then defined as those of separate donor and acceptor units. The current trend in experimental design of donor–acceptor systems, however, has focused more attention on intramolecular reactions where the donor and acceptor units are coupled in one molecule by a bridge.[22] The direct donor–acceptor overlap and the mixing to bridge states both lead to electronic delocalization,[75,76] with the result that the centers of electronic localization and localized diabatic states are ill-defined. It is then more appropriate to use either the GMH or adiabatic formulation.

There is an additional, more fundamental, issue involved in applying the standard diabatic formalism. The solvent reorganization energy and the solvent component of the equilibrium free energy gap are bilinear forms of $\Delta\mathcal{E}_{ab}$ and $\mathcal{E}_{av}$ (Eqs. [45] and [47]). A unitary transformation of the diabatic basis (Eq. [27]), which should not affect any physical observables, then changes $\Delta\mathcal{E}_{ab}$ and $\mathcal{E}_{av}$, affecting the reorganization parameters. The activation parameters of ET consequently depend on transformations of the basis set!

This situation is of course not satisfactory as observable quantities should be invariant with respect to unitary basis transformations.[84] Here, we outline the adiabatic route to a basis-invariant formulation of the theory.[42]

In the adiabatic gas-phase basis, the number of independent parameters drops to four: λ^{ad}, ΔF_s^{ad}, ΔE_{12}, and α_{12}, where the superscript "ad" refers to the adiabatic representation in which ΔE_{12} is the gas-phase gap between the eigenenergies, Eq. [29]. The equation for the free energy surfaces can then be rewritten in the basis-invariant form

$$F_{\pm}(Y^{ad}) = \frac{(Y^{ad})^2}{4\Delta e^2 \lambda^I} \pm \tfrac{1}{2}\Delta E(Y^{ad}) + C \qquad [105]$$

with

$$\Delta E(Y^{ad}) = \left[\Delta E_{12}^2 + 2\Delta E_{12}(\Delta e \Delta F_s^I - Y^{ad}) + [\Delta F_s^I - (Y^{ad}/\Delta e)]^2\right]^{1/2} \qquad [106]$$

and

$$\Delta e = \left[1 + 4\alpha_{12}^2\right]^{-1/2} \qquad [107]$$

The reaction coordinate is now a projection of the nuclear solvent polarization on the adiabatic differential solute field

$$Y^{ad} = \Delta\mathcal{E}_{12} \cdot \mathcal{P}_n \qquad [108]$$

Both the solvent reorganization energy

$$\lambda^I = \tfrac{1}{2}\left(\Delta\mathcal{E}_{12}^2 + 4\mathcal{E}_{12}^2\right)^{1/2} \cdot \chi_n \cdot \left(\Delta\mathcal{E}_{12}^2 + 4\mathcal{E}_{12}^2\right)^{1/2} \qquad [109]$$

and the solvent component of the free energy gap

$$\Delta F_s^I = -\tfrac{1}{2}\left(\Delta\mathcal{E}_{12}^2 + 4\mathcal{E}_{12}^2\right)^{1/2} \cdot \chi \cdot (\mathcal{E}_1 + \mathcal{E}_2) \qquad [110]$$

are invariants of unitary basis transformations (Eq. [27]) and have the same magnitude in the GMH and any diabatic basis set. This follows from the invariance property of the matrix trace

$$\sum_i A_{ii} = \text{inv} \qquad [111]$$

and the expression

$$\Delta A_{12}^2 + 4A_{12}^2 = \Delta A_{ab}^2 + 4A_{ab}^2 = \text{inv} \qquad [112]$$

Here "inv" stands for an invariant in respect to transformation consistent with the symmetry of the system. For quantum mechanical operators, this means unitary transformations. The parameter Δe in Eq. [107] quantifies the extent of mixing between two adiabatic gas-phase states induced by the interaction with the solvent. For a dipolar solute, it is determined through the adiabatic differential and the transition dipole moments

$$\Delta e = \left[1 + \frac{4m_{12}^2}{\Delta m_{12}^2}\right]^{-1/2} \qquad [113]$$

The differential and transition dipoles can be determined from experiment: the former from the Stark spectroscopy[75,76] and the latter from absorption or emission intensities (see below).

The parameter Δe should not be confused with the actual difference in electronic occupation numbers of the two CT states. When the eigenfunctions $\{\tilde{\phi}_+(Y^{ad}), \tilde{\phi}_-(Y^{ad})\}$ corresponding to the eigenstates $F_\pm(Y^{ad})$ are represented as a linear combination of the wave functions of the adiabatic basis, $\{\phi_1, \phi_2\}$,

$$
\begin{aligned}
\tilde{\phi}_+(Y^{ad}) &= \sqrt{1 - f(Y^{ad})}\phi_1 + \sqrt{f(Y^{ad})}\phi_2 \\
\tilde{\phi}_-(Y^{ad}) &= -\sqrt{f(Y^{ad})}\phi_1 + \sqrt{1 - f(Y^{ad})}\phi_2
\end{aligned}
\qquad [114]
$$

then the parameter $f(Y^{ad})$ defines the occupation number of the adiabatic state 1 on the lower CT free energy surface at the reaction coordinate Y^{ad}. For CT transitions in the normal region, two equilibrium minima are located on the lower CT free energy surface. The occupation number difference in the final and initial states can thus be defined as

$$\Delta z = |1 - f(Y_1^-) - f(Y_2^-)| \qquad [115]$$

where Y_1^- and Y_2^- are two minima positioned on the lower CT surface (Figure 12). In contrast, when transitions between the lower and upper CT surfaces occur in the inverted CT region, the occupation number difference becomes

$$\Delta z = |f(Y^+) - f(Y^-)| \qquad [116]$$

where now Y^+ and Y^- define the positions of equilibrium on the upper and lower CT surfaces, respectively (Figure 13). Figure 14 illustrates the difference in the dependence of the occupation number difference on Δe in the normal and inverted CT regions. The parameter Δz is indeed close to Δe for reactions with $|\Delta F_s^I| \ll \lambda^I$. As the absolute value of the equilibrium energy gap increases,
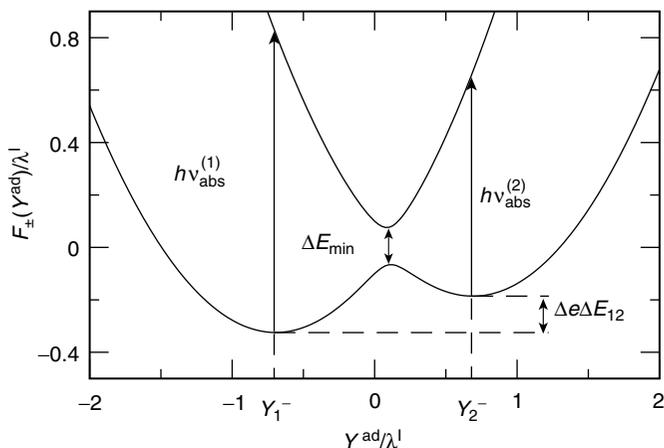
Figure 12 The CT adiabatic free energy surfaces in the normal CT region. The labels $h\nu_{abs}^{(1)}$ and $h\nu_{abs}^{(2)}$ indicate two adiabatically split absorption transitions corresponding to two minima of the lower surface with the coordinates Y_1^- and Y_2^-; $\Delta e = 0.7$, $\Delta F_s^I = 0$, $\Delta E_{12}/\lambda^I = 0.2$. The gap ΔE_{min} is the minimum splitting between the upper and lower CT surfaces (Eq. [149]).
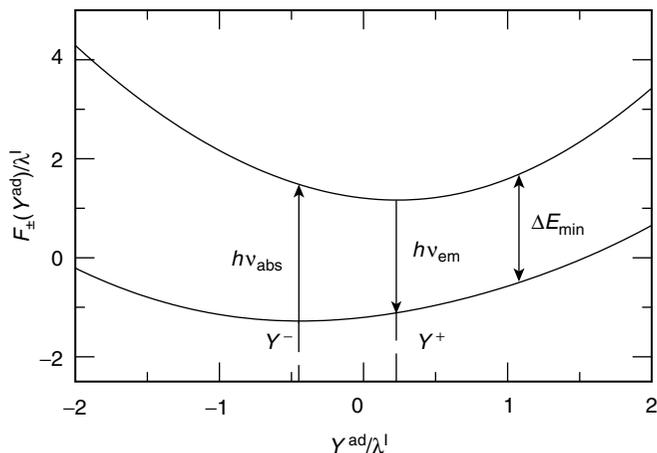

Figure 13 The CT adiabatic free energy surfaces in the CT inverted region; $\Delta e = 0.7$, $\Delta F_s^I/\lambda^I = -1.0$, $\Delta E_{12}/\lambda^I = 3.0$. The points Y^- and Y^+ indicate the minima of the lower and upper adiabatic surfaces, respectively. The labels $h\nu_{abs/em}$ are absorption and emission energies, and ΔE_{min} is the minimum energy gap between the free energy surfaces (Eq. [149]).
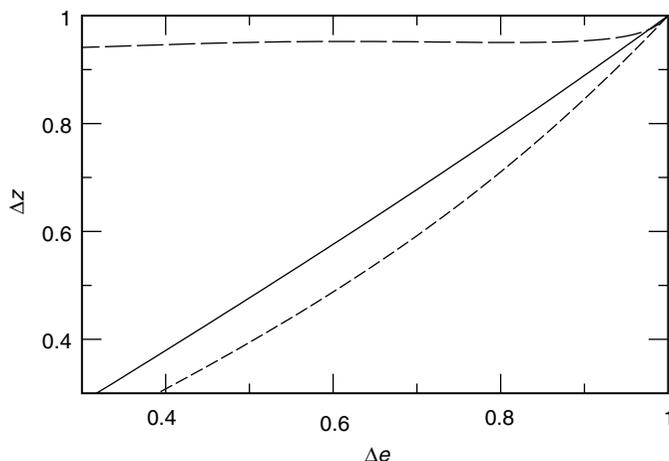
Figure 14 Dependence of the occupation number difference Δz on the mixing parameter Δe at $\Delta E_{12}/\lambda^I = 0.2$, $\Delta F_s^I = 0$ (solid line); $\Delta E_{12}/\lambda^I = 0.5$, $\Delta F_s^I = 0$ (dot–dashed line); $\Delta E_{12}/\lambda^I = 3.0$, $\Delta F_s^I/\lambda^I = -1.0$ (dashed line).

Δz increasingly deviates from Δe. In the inverted region, Δz is nearly 1 and is almost independent of Δe.

The establishment of the invariant reorganization energy λ^I allows one to use electrostatic models for the reorganization energy based on solvation of fixed charges located at molecular sites[5] instead of using a more complicated algorithm through the delocalized electronic density.[84] This ability to use electrostatic fixed charge models instead of distributed density of quantum mechanics is permitted because the invariant reorganization energy sets up the characteristic length between centers of charge localization to be used in electrostatic models of solvent reorganization[7]

$$r_{CT} = e^{-1}\left[\Delta m_{12}^2 + 4m_{12}^2\right]^{1/2} \qquad [117]$$

For self-exchange transitions, due to the relation $2m_{12} = \Delta m_{ab}$, one gets

$$r_{CT} = \left[r_{12}^2 + r_{ab}^2\right]^{1/2} \qquad [118]$$

where r_{ab} is the distance between the centers of electron localization in the diabatic representation.

The mixing parameter Δe makes the CT free energy surfaces dependent on the gas-phase, adiabatic transition dipole moment. The standard extension of the MH theory on the case of strong electronic overlap[85] assumes a nonzero ET matrix element H_{ab}, but neglects the diabatic transition dipole (or eliminates

it by choosing the GMH basis set[7]). In this case, the CT free energy surface is defined by Eq. [41] with the following energy gap:

$$\Delta E^{\mathrm{d}}(Y^{\mathrm{d}}) = [\Delta E_{12}^2 + 2\Delta I_{ab}(\Delta F_{\mathrm{s}}^{\mathrm{d}} - Y^{\mathrm{d}}) + (\Delta F_{\mathrm{s}}^{\mathrm{d}} - Y^{\mathrm{d}})^2]^{1/2} \qquad [119]$$

The diabatic and adiabatic formulations can be compared when the condition $\mathcal{E}_{ab} = 0$ is imposed. Then, one obtains $Y^{\mathrm{ad}} = \Delta e Y^{\mathrm{d}}$, $\lambda^I = \lambda^{\mathrm{d}}$, $\Delta F_{\mathrm{s}}^I = \Delta F_{\mathrm{s}}^{\mathrm{d}}$.

Figure 15 compares the free energy surfaces given by Eqs. [105] and [106] to those from Eqs. [41] and [119] for self-exchange CT ($\Delta I_{ab} = 0$, $\Delta F_{\mathrm{s}}^{\mathrm{d}} = \Delta F_{\mathrm{s}}^I = 0$). Several important distinctions between the two formulations can be emphasized. (1) The positions of transition points do not coincide. The maximum of $F_-(Y^{\mathrm{ad}})$ in the present formulation deviates from the position of the resonance of the diagonal elements of the two-state Hamiltonian matrix, $Y^{\ddagger} = 0$, and is approximately equal to $Y^{\ddagger} = (\Delta e)^2 \Delta E_{12}$ when $\Delta E_{12}/\lambda^I \ll 1$ and $\Delta F_{\mathrm{s}}^I = 0$. (2) The splitting of the lower and upper adiabatic surfaces is larger in the MH formulation than in the basis-invariant formulation. For self-exchange CT, the splitting is $2|H_{ab}| = \Delta E_{12}$ in the former case and $\Delta E_{12}\sqrt{1 - \Delta e^2}$ in the latter case. (3) The MH formula involves the diabatic equilibrium free energies F_{0i}^{d} without donor–acceptor overlap. The gap ΔF_0^{d} is therefore zero for self-exchange reactions. The adiabatic representation includes explicitly the donor–acceptor overlap that results in a symmetry-breaking splitting of the gas-phase electronic states to the energy ΔE_{12}.
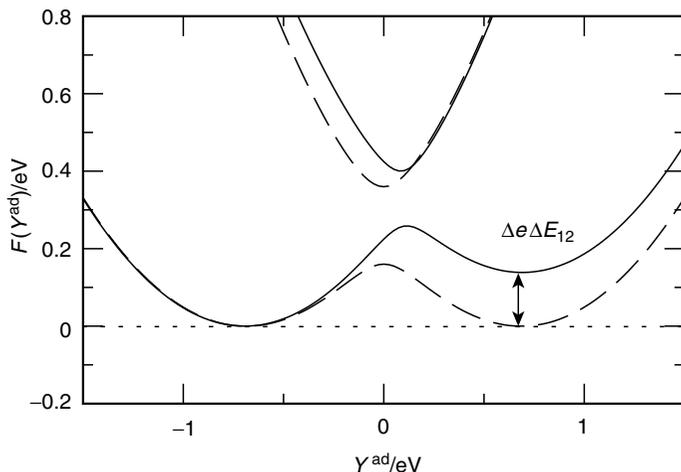

Figure 15 Adiabatic free energy surfaces $F_{\pm}(Y^{\mathrm{ad}})$ in the present model (solid lines, Eqs. [105] and [106]) and in the Marcus–Hush formulation (long-dashed lines, Eqs. [41] and [119]) for self-exchange CT with $\Delta F_{\mathrm{s}}^I = \Delta F_{\mathrm{s}}^{\mathrm{d}} = 0$, $\lambda^I = \lambda^{\mathrm{d}} = 1$ eV, $\Delta E_{12} = 0.2$ eV, and $\Delta e = 0.7$. All free energy surfaces are vertically shifted to have zero value (dotted line) at the position of the left minimum.

Electronic transitions in the gas phase thus proceed from the lower state E_1 to the upper state E_2. In condensed phases, these states are of course "dressed" by a solvating environment, but at $\Delta F_s^I = 0$ one gets a nonzero equilibrium driving force approximately equal to $\Delta e \Delta E_{12}$ when $\Delta E_{12}/\lambda^I \ll 1$. The factor Δe in the *free energy* driving force appears because the free energy represents the work done to transfer the charge Δe ($\Delta z \approx \Delta e$ at $\Delta E_{12}/\lambda^I \ll 1$, see Figure 14) over the energy barrier ΔE_{12} that results in $\Delta e \Delta E_{12}$ for small splittings ΔE_{12}.

Note above that the GMH[7] and adiabatic formulations are equivalent in terms of building the CT free energy surfaces. The distinctions seen in Figure 15 may seem to contradict to this statement. The problem is resolved by noting that the requirement $\mathcal{E}_{ab} = 0$ imposed by the GMH formulation makes the diabatic energy gap nonzero for self-exchange transitions:

$$\Delta I_{ab}^{GMH} = \Delta e \Delta E_{12} \qquad [120]$$

which is indeed the gap shown in Figure 15. The standard MH formulation[85] is then recovered when $m_{12} = 0$ for symmetry reasons and thus $\Delta e = 1$.

Nonlinear Solvation versus Intramolecular Effects

The origin of nonparabolic free energy surfaces of ET can be divided into two broad categories: (1) intramolecular electronic effects and (2) nonlinear solvation effects. Although these two origins can, at some instances, be treated within the same mathematical framework (Q model), there are substantial differences between them at both the quantitative and qualitative levels. From the quantitative viewpoint, nonlinear solvation produces a much weaker distortion of ET parabolas than do the polarizability change and electronic delocalization. From a qualitative viewpoint, the two categories of effects produce a nonzero nonparabolic distortion in different orders of the expansion of the system Hamiltonian in the driving solvent mode.

The free energy $F(P)$ invested in creation of a nonequilibrium solvent polarization P can be expressed as a series in even powers of P with the two first terms as follows:

$$F(P) = a_1 P^2 + a_2 P^4 \qquad [121]$$

where a_1, $a_2 > 0$. The interaction energy of the solute field with the solvent polarization, U_{0s}, is linear in P

$$U_{0s} = -bP \qquad b > 0 \qquad [122]$$

For weak solute–solvent interactions, deviations from zero polarization of the solvent are small, and one can keep only the first, harmonic, term in Eq. [121].

Anharmonic higher order terms gain importance for stronger solute-solvent couplings requiring $a_2 \neq 0$ in Eq. [121]. The nonequilibrium solvent polarization can be considered as an ET reaction coordinate. The curvature of the corresponding free energy surface is

$$F''(P_0) = 2a_1 + 12a_2 P_0^2 \qquad [123]$$

at the minimum point P_0 defined by the condition $F'(P_0) = b$. Equation [123] indicates that nonlinear solvation effects, usually associated with dielectric saturation, enhance the curvature compared to the linear response result $F'' = 2a_1$. This enhancement of curvature leads to a decrease in the solvent reorganization energy. The effect is, however, relatively small as it arises from anharmonic expansion terms.

When the electron is partially delocalized, one should switch to the adiabatic representation in which the upper and lower CT surface are split by an energy gap depending on P. If this energy gap is expanded in P with truncation after the second-order term, we come to the model of a donor–acceptor complex whose dipolar polarizabilities are different in the ground and excited states. The solute–solvent interaction energy then attains the energy of solute polarization that is quadratic in P

$$U_{0s} = -bP - cP^2 \qquad c > 0 \qquad [124]$$

The total system energy $F(P) + U_{0s}$ includes, therefore, a quadratic in P term with the coefficient $(a_1 - c)$. This quadratic term initiates a revision of the frequency of solvent fluctuations driving CT. The curvature of harmonic surfaces decreases producing higher reorganization energies. Since the solute polarizability contributes already to the harmonic term, its effect on the reorganization energy is stronger than that of nonlinear solvation.

The revision of characteristic frequencies of nuclear modes is a general result of electronic delocalization holding for both the intramolecular vibrational modes[65] and the solvent modes. The fact that this effect shows up already in the harmonic expansion term makes it much stronger compared to nonlinear solvation in respect to nonparabolic distortion of the free energy surfaces.

OPTICAL BAND SHAPE

Spectral measurements open a door to access the rate constant parameters of ET. The connection between optical observables and ET parameters can be divided into two broad categories: (1) analysis of the optical band profile (band shape analysis) and (2) the use of integrated spectral intensities (see

below). The former route connects the spectral moments to ET activation parameters (Table 1). The latter is applied to extract off-diagonal matrix elements, most often the ET matrix element and the transition dipole. Band shape analysis of optical spectra has been successfully used in ET research for many years, and our present knowledge about mechanisms and energetics of ET originates largely from spectroscopic measurements.[16] The understanding of electronic and solvent effects on the ET kinetics has been recently supplemented by extensive information about the intramolecular, vibronic envelope from resonance Raman spectroscopy.[86]

The fast growth of the field of ET research and, especially, the design of new bridge-coupled donor–acceptor pairs imposes new demands on the theory of optical spectra. Several major challenges are currently faced by the field. They may be summarized as follows: (1) The presently existing band-shape analysis has been created for ET transitions.[17] It has not anticipated strong electronic coupling and thus fails when applied to transitions with high magnitudes of the ET matrix element.[87] (2) The model is limited to two states only. Mixing to higher excited states, resulting in intensity borrowing, is commonly neglected. Extension to more then two states is especially important for photoinduced CT where a CT state is formed from and is strongly coupled to a locally exited state of either donor or acceptor unit.[17,88] (3) There are indications in the literature that the common assumption of complete decoupling between the intramolecular vibrational modes and solvent thermal motions may fail for some systems.[89,90] Understanding the origin of and full account for these effects should be incorporated into new models of optical bands.

The challenges outlined above still await a solution. In this section, we show how some of the theoretical limitations employed in traditional formulations of the band shape analysis can be lifted. We discuss two extensions of the present-day band shape analysis. First, the two-state model of CT transitions is applied to build the Franck–Condon optical envelopes. Second, the restriction of only two electronic states is lifted within the band shape analysis of polarizable chromophores that takes higher lying excited states into account through the solute dipolar polarizability. Finally, we show how a hybrid model incorporating the electronic delocalization and chromophore's polarizability effects can be successfully applied to the calculation of steady-state optical band shapes of the optical dye coumarin 153 (C153). We first start with a general theory and outline the connection between optical intensities and the ET matrix element and transition dipole.

Optical Franck–Condon Factors

Absorption of light by molecules, resulting in electronic excitations, is caused by the interaction of the bound molecular electrons with the electric field of the radiation. In the dipolar approximation, the interaction of the dipole operator of the solute $\hat{\mathbf{m}}_0$ with the time-dependent electric field $\mathbf{E}(t)$

of the radiation is the perturbation that drives the electronic excitation. The time-dependent interaction Hamiltonian is

$$-f(n_D)\,\hat{\mathbf{m}}_0 \cdot \mathbf{E}(t) \qquad [125]$$

where the parameter $f(n_D)$ accounts for the deviation of the local field acting on the solute dipole from the external field $\mathbf{E}(t)$; n_D is the solvent refractive index. Dielectric theories[91] predict for spherical cavities

$$f(n_D) = \frac{3n_D^2}{2n_D^2 + 1} \qquad [126]$$

The perturbation given by Eq. [125] mixes the electronic states for which the off-diagonal matrix element of the dipole operator, m_{jk}, is nonzero. The latter is called the transition dipole.[49] Mixing of electronic states by a time-dependent external field leads to the dependence of the corresponding electronic state populations on time. The rate constant of the population kinetics is given by the transition probability. Quantum mechanical perturbation theory, limited to the first order in the interaction perturbation, is commonly used to calculate the one-photon transition probability and absorption intensity.[15,92] This formalism, combined with the Einstein relation between absorption intensity and the probability of spontaneous emission,[49,92] leads to experimental observables, the extinction coefficient of absorption, $\epsilon(v)$ $(\mathrm{cm}^{-1}\,\mathrm{M}^{-1})$, and the emission rate, $I_{em}(v)$ (number of photons per unit frequency), as functions of the light frequency v. They are given by the following relations:

$$\frac{\epsilon(v)}{v} = \frac{8\pi^3 N_A}{3000\,(\ln 10)\,c}\frac{f^2(n_D)}{n_D}G_-(v) \qquad [127]$$

and

$$I_{em}(v) = \frac{64\pi^4 v^3}{3c^3}n_D f^2(n_D)G_+(v) \qquad [128]$$

In Eq. [127], N_A is the Avogadro number, and c in Eqs. [127] and [128] is the speed of light in vacuum.

The extinction coefficient and emission rate are defined through the spectral density function $G_\pm(v)$ that combines the effects of solvent-induced inhomogeneous broadening and vibrational excitations of the donor–acceptor complex. A substantial simplification of the description can be achieved if the two types of nuclear motions are not coupled to each other. The spectral density $G_\pm(v)$ is then given by the convolution[17]

$$G_\pm(v) = |\tilde{m}_{12}(hv)|^2 \int \mathrm{FCWD}_\pm^s(x)\mathrm{FCWD}_\pm^v(v - x)dx \qquad [129]$$

of the gas-phase vibronic envelope $\text{FCWD}^v_\pm(v)$ with the normalized solvent-induced band shape

$$\text{FCWD}^s_\pm(v) = \langle \delta(\Delta E(\mathbf{q}) - hv) \rangle_\pm \qquad [130]$$

where the average is taken over the solvent configurations statistically weighted with the Boltzmann factor $\exp(-\beta F_\pm)$ with "$-$" for absorption and "$+$" for emission.

In Eqs. [129] and [130], $\text{FCWD}^v_\pm(v)$ and $\text{FCWD}^s_\pm(v)$ refer to the normalized Franck–Condon weighted density of the vibrational excitations of the solute (including quantum overlap integrals of the vibrational normal modes of the solute coupled to the transferred electron[17]) and the normalized solvent-induced spectral distribution function, respectively. The gap, $\Delta E(\mathbf{q}) = E_+(\mathbf{q}) - E_-(\mathbf{q})$, in Eq. [130] is defined between the upper adiabatic surface $E_+(\mathbf{q})$ and the lower adiabatic surface $E_-(\mathbf{q})$ depending on a set of nuclear solvent modes $\mathbf{q}$. Because the transitions occur between the adiabatic free energy surfaces $E_\pm(\mathbf{q})$, the unperturbed basis set in the quantum mechanical perturbation theory is built on the wave functions $\{\tilde{\phi}_1(\mathbf{q}), \tilde{\phi}_2(\mathbf{q})\}$ diagonalizing the corresponding two-state Hamiltonian matrix (Eq. [114]). The dependence on the nuclear solvent configuration comes into the transition dipole moment (as calculated within the two-state model, TSM)

$$\begin{aligned} |\tilde{m}_{12}(\mathbf{q})| &= |\langle \tilde{\phi}_1(\mathbf{q})|\hat{\mathbf{m}}_0|\tilde{\phi}_2(\mathbf{q})\rangle| \\ &= |m_{12}| \frac{\Delta E_{12}}{\Delta E(\mathbf{q})} \end{aligned} \qquad [131]$$

only through the energy gap $\Delta E(\mathbf{q})$, which is equal to hv according to Eq. [130]. This relationship is the reason for the dependence of the transition dipole on the light frequency in Eq. [129]. Coupling to higher lying excited states modifies Eq. [131], but if the dependence on the solvent field comes into $\tilde{m}_{12}(\mathbf{q})$ only through the instantaneous energy gap, the transition dipole can still be taken out of the solvent average with, however, a more complicated dependence on the frequency of the incident light.[17,93] In the TSM, one has, according to Eq. [131]

$$\tilde{m}_{12}(v) = m_{12}\Delta E_{12}/hv \qquad [132]$$

where m_{12} is the gas-phase adiabatic transition dipole moment.

The vibronic envelope $\text{FCWD}^v_\pm(v)$ in Eq. [129] can be an arbitrary gas-phase spectral profile. In condensed-phase spectral modeling, one often simplifies the analysis by adopting the approximation of a single effective vibrational mode (Einstein model) with the frequency v_v and the vibrational reorganization energy λ_v. The vibronic envelope is then a Poisson distribution of

individual vibrational excitations[44]

$$\text{FCWD}_\pm^v(\nu) = e^{-S_v} \sum_{m=0}^{\infty} \frac{S_v^m}{m!} \delta(h\nu \pm mh\nu_v) \qquad [133]$$

where S_v the Huang–Rhys factor $S_v = \lambda_v/h\nu_v$ (cf. to Eq. [38]). The whole inhomogeneous line shape then takes the form of a weighed sum over the solvent-induced bands, each shifted relative to the other by ν_v

$$G_\pm(\nu) = |\tilde{m}_{12}(h\nu)|^2 e^{-S_v} \sum_{m=0}^{\infty} \frac{S_v^m}{m!} \text{FCWD}_\pm^s(\nu \pm mh\nu_v) \qquad [134]$$

Equation [134], given in the form of a weighted sum of individual solvent-induced line shapes, provides an important connection between optical band shapes and CT free energy surfaces. Before turning to specific models for the Franck–Condon factor in Eq. [134], we present some useful relations, following from integrated spectral intensities, that do not depend on specific features of a particular optical line shape.

Absorption Intensity and Radiative Rates

Extraction of activation CT parameters requires an analysis of spectral band shapes. One parameter, however, can be obtained from the integrated absorption and emission intensities. Since mixing of the electronic states in the external electric field of radiation is governed by the magnitude of the transition dipole, the transition dipole also defines the intensity of the corresponding optical line. The extinction coefficient or emission rate integrated over light frequencies then allows one to obtain the transition dipole, provided its frequency dependence is known. [Traditionally, the transition dipole is assumed to be frequency independent.[49] This leads, however, to systematic errors in estimates of transition dipoles from optical spectra, see below.] For the TSM, this procedure leads to the gas-phase transition dipole. The transition dipole is important as a parameter quantifying the extent of CT delocalization and to generate CT free energy surfaces in electronically delocalized donor–acceptor complexes. It also has an important implication due to its connection to the ET matrix element (through the Mulliken-Hush relation),[7] which enters the rate constant of nonadiabatic ET reaction rates (Eq. [2]; see below).

Integration of absorption extinction coefficient (Eq. [127]) and emission rate (Eq. [128]) gives two alternative estimates for the adiabatic gas-phase transition dipole m_{12} (in D) within the TSM frequency-dependent $\tilde{m}_{12}(\nu)$ (Eq. [132])

$$m_{12} = 9.585 \times 10^{-2} \frac{\sqrt{n_D}}{\bar{\nu}_0 f(n_D)} \left[\int \bar{\nu}\epsilon(\bar{\nu})d\bar{\nu}\right]^{1/2} \qquad [135]$$

and

$$m_{12} = 3.092 \times 10^8 [\bar{\nu}_0 \sqrt{n_D} f(n_D)]^{-1} \left[\int I_{em}(\nu) \nu^{-1} d\nu \right]^{1/2} \qquad [136]$$

where $\bar{\nu}$ is the wavenumber (cm^{-1}) and $\bar{\nu}_0 = \Delta E_{12}/hc$. When the emission spectrum is not available, the radiative rate[49]

$$k_{rad} = \int I_{em}(\nu) d\nu = \Phi_{em} \tau_{em}^{-1} \qquad [137]$$

can be used; Φ_{em} and τ_{em} are the quantum yield and emission lifetime. By defining the average frequency

$$\nu_{av} = \int I_{em}(\nu) d\nu \Big/ \int I_{em}(\nu) \nu^{-1} d\nu \qquad [138]$$

one gets

$$m_{12} = 1.786 \times 10^3 \left[\frac{k_{rad}}{\bar{\nu}_{av} \bar{\nu}_0^2 n_D f^2(n_D)} \right]^{1/2} \qquad [139]$$

Equation [139] is not very practical because an accurate definition of the average wavenumber, $\bar{\nu}_{av} = \nu_{av}/c$, demands knowledge of the emission spectrum for which Eq. [136] provides a direct route to the transition dipole. But Eq. [139] can be used in approximate calculations by assuming $\bar{\nu}_{av} = \bar{\nu}_{em}$.

    Equation [139] is exact for a two-state solute, but differs from the traditionally used connection between the transition dipole and the emission intensity by the factor $\bar{\nu}_0/\bar{\nu}_{av}$.[49] The commonly used combination $m_{12}\bar{\nu}_0/\bar{\nu}_{av}$ appears as a result of neglect of the frequency dependence of the transition dipole $\tilde{m}_{12}(\nu)$ entering Eq. [129]. It can be associated with the condensed-phase transition dipole in the two-state approximation.[43] Exact solution for a two-state solute makes the transition dipole between the adiabatic free energy surfaces inversely proportional to the energy gap between them. This dependence, however, is eliminated when the emission intensity is integrated with the factor ν^{-1}.[93]

    The transition dipole m_{12} in Eqs. [136] and [139] is the gas-phase adiabatic transition dipole. Therefore, emission intensities measured in different solvents should generate invariant transition dipoles when treated according to Eqs. [136] and [139]. A deviation from invariance can be used as an indication of the breakdown of the two-state approximation and the existence of intensity borrowing from other excited states of the chromophores (the Murrell mechanism[17,88,94]). Figure 16 illustrates the difference between Eq. [139] and
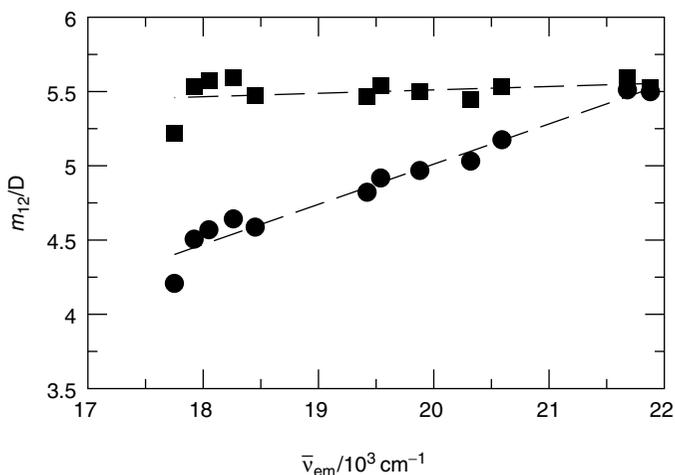
Figure 16 The transition dipole m_{12} according to Eq. [139] ($\bar{v}_{av} = \bar{v}_{em}$, circles) and $m_{12}\bar{v}_0/\bar{v}_{em}$ (squares) versus $\bar{v}_{em}$ for emission transitions in C153 in different solvents.[95] The dashed lines are regressions with the slopes 0.02 (squares) and 0.27 (circles).

the traditional formulation. It shows the dependence of m_{12} (circles) and $m_{12}\bar{v}_0/\bar{v}_{av}$ (squares, $\bar{v}_0 = 25{,}400$ cm^{-1}) on the emission frequency $\bar{v}_{em}$ for the dye C153 measured in solvents of different polarity.[95] The two sets of transition dipoles are noticeably divergent in strongly polar solvents.

Electron-Transfer Matrix Element

The transition dipole between the free energy surfaces $F_{\pm}(X)$ is not the only parameter that depends on the nuclear configuration of the solvent. The effective ET matrix element $H_{ab}^{eff}[\mathcal{P}_n]$ following from the trace of the two-state Hamiltonian over the electronic degrees of freedom also depends on the nuclear configuration of the solvent (Eq. [37]). In contrast to the case of optical transitions where the dependence on the nuclear solvent configurations is transformed into a frequency dependence of the transition dipole $\tilde{m}_{12}(v)$ (Eq. [132]), the dependence of the ET matrix element $H_{ab}^{eff}[\mathcal{P}_n]$ on the nuclear field $\mathcal{P}_n$ should be fully included into the statistical average over $\mathcal{P}_n$ when the ET rate constant is calculated in the Golden Rule perturbation scheme over $H_{ab}^{eff}[\mathcal{P}_n]$.[11] The $\mathcal{P}_n$ dependence represents a non-Condon effect of the solvent field on the rate preexponential factor. The result of the calculations[43] is the standard Golden Rule expression[9,11] for the nonadiabatic rate constant

$$k_{NA}^{(i)} = \hbar^{-1}(\pi\beta/\lambda)^{1/2}|H^{MH}|^2 FCWD_i(0) \qquad [140]$$

with the Mulliken–Hush[6] ET matrix element

$$H^{\mathrm{MH}} = H_{ab} - \frac{(\mathbf{m}_{ab} \cdot \Delta\mathbf{m}_{ab})}{\Delta m_{ab}^2} \Delta I_{ab} \qquad [141]$$

where $\mathbf{m}_{ab}$ and $\Delta\mathbf{m}_{ab}$ refer to, respectively, the gas-phase transition and differential dipole moments calculated in the diabatic basis set; ΔI_{ab} is the diabatic gas-phase energy gap. The term Mulliken–Hush[6] here refers to the fact that the matrix element in Eq. [141] is related to the projection of the adiabatic transition dipole on the direction of the difference diabatic dipole

$$H^{\mathrm{MH}} = \frac{(\mathbf{m}_{12} \cdot \Delta\mathbf{m}_{ab})}{\Delta m_{ab}^2} \Delta E_{12} \qquad [142]$$

Under the special condition that $\mathbf{m}_{12}$ and $\Delta\mathbf{m}_{ab}$ are parallel, one obtains the MH relation[6,7]

$$H^{\mathrm{MH}} = \frac{m_{12}}{\Delta m_{ab}} \Delta E_{12} \qquad [143]$$

Equations [140]–[143] provide a connection between the preexponential factor entering the nonadiabatic ET rate and the spectroscopically measured adiabatic transition dipole m_{12}. It turns out that the Mulliken–Hush matrix element, commonly considered as an approximation valid for $m_{ab} = 0$,[7] enters exactly the rate constant preexponent as long as the non-Condon solvent effects are accurately taken into account.[43] Equation [142] stresses the importance of the orientation of the adiabatic transition dipole relative to the direction of ET set up by the difference diabatic dipole $\Delta\mathbf{m}_{ab}$. The value of H^{MH} is zero when the vectors $\mathbf{m}_{12}$ and $\Delta\mathbf{m}_{ab}$ are perpendicular.

Electronically Delocalized Chromophores

Equation [48] gives the Franck–Condon factor that defines the probability of finding a system configuration with a given magnitude of the energy gap between the upper and lower CT free energy surfaces. It can be directly used to define the solvent band shape function[96] of a CT optical transition in Eq. [134]

$$\mathrm{FCWD}_{\pm}^{\mathrm{s}}(\nu \pm mh\nu_{\mathrm{v}}) = Q_{\pm}^{-1} \sum_{k=1,2} |\Delta E'(Y_{km})|^{-1} \exp[-\beta F_{\pm}(Y_{km})] \qquad [144]$$

where

$$Q_{\pm} = \int e^{-\beta F_{\pm}(Y^{\mathrm{ad}})} dY^{\mathrm{ad}} \qquad [145]$$

In Eq. [144], the coordinates Y_{km} $(k = 1, 2)$ are two roots of the quadratic equation

$$\Delta E(Y^{\text{ad}}) = h(\nu \pm m\nu_{\text{v}}) \qquad [146]$$

given by the expression

$$
\begin{aligned}
Y_{1m} &= Y_{\min} + \Delta e \sqrt{h^2(\nu \pm m\nu_{\text{v}})^2 - \Delta E_{\min}^2} \\
Y_{2m} &= Y_{\min} - \Delta e \sqrt{h^2(\nu \pm m\nu_{\text{v}})^2 - \Delta E_{\min}^2}
\end{aligned}
\qquad [147]
$$

The appearance of the square root in Eq. [147] is an indication of one important feature of delocalized CT systems: the existence of a lower limiting frequency of the incident light that can be absorbed by a donor–acceptor complex. This effect results in asymmetries of CT absorption and emission lines as discussed below.

A real root of Eq. [146] exists only if the following condition holds:

$$h\nu \geq \pm mh\nu_{\text{v}} + \Delta E_{\min} \qquad [148]$$

for a vibronic transition with m phonons of vibrational excitation. The 0–0 transition $(m = 0)$ sets up the absolute minimum frequency

$$h\nu_{\min} = \Delta E_{\min} = \sqrt{1 - \Delta e^2} \Delta E_{12} \qquad [149]$$

where Δe is the gas-phase mixing parameter (Eq. [107]), and ΔE_{12} is the gas-phase adiabatic energy gap (Eq. [29]). The energy $\Delta E_{\min}$ corresponds to the minimum splitting between the upper and lower CT free energy surfaces (Figures 12 and 13) that occurs at the coordinate

$$Y_{\min} = \Delta e^2 \Delta E_{12} + \Delta e \Delta F_{\text{s}}^{\text{I}} \qquad [150]$$

The transition intensity is always zero at $\nu < \nu_{\min}$. The existence of the lower transition boundary makes a profound effect on optical band shapes for a large extent of mixing of adiabatic states. The general effect of the existence of the minimum frequency on optical lines is to produce line asymmetry by squeezing its red wing.[20,97] We consider here this effect for the example of transitions in the inverted CT region when both the absorption and emission lines can be observed (Figure 13). For positively solvatochromic dyes with a major multipole higher in the excited state than in the ground state, emission lines are shifted more strongly to the red side of the spectrum than the absorption lines. Therefore, the emission lines are closer to the low-energy boundary $\nu_{\min}$
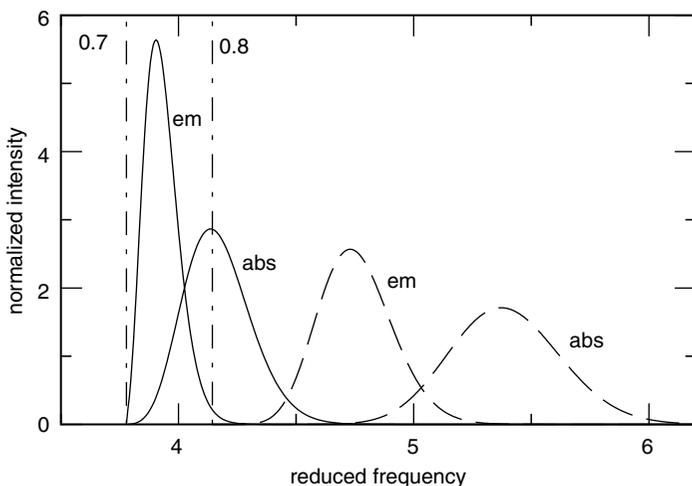
Figure 17 The normalized absorption (abs.) and emission (em.) intensities at $\Delta e = 0.7$ (solid lines) and $\Delta e = 0.8$ (long-dashed lines) versus the reduced frequency $h\nu/\lambda^I$. The dash–dotted lines indicate the lower boundary for the energy of the incident light ν_{min} (Eq. [149]).

and get narrower than the absorption lines (Figure 17). The opposite trend holds for negatively solvatochromic dyes with higher major multipoles in their ground states.

The lower free energy surface has two minima in the normal CT region (Figure 12). Two absorption transitions exist in this case, even for self-exchange reactions. The reason is the symmetry breaking induced by a non-zero adiabatic transition dipole leading to $\Delta e < 1$ (the standard MH picture, Figure 15, is recovered when $m_{12} = 0$). The energy splitting between the two minima of the lower free energy surface gives rise to two transition frequencies

$$h\nu_{abs}^{(1)} = \lambda_v + \lambda^I + \Delta F_s^I + \Delta e \Delta E_{12} \qquad [151]$$

and

$$h\nu_{abs}^{(2)} = \lambda_v + \lambda^I - \Delta F_s^I - \Delta e \Delta E_{12} \qquad [152]$$

The combination of Eq. [134] with Eq. [144] provides an effective formalism for the band shape analysis of CT spectra when a substantial degree of electronic delocalization is involved. Equation [134] is exact for a TS donor–acceptor complex and, therefore, can be used for an arbitrary degree of electronic delocalization as long as the assumption of decoupling of the vibrational and solvent modes holds. Figure 18 illustrates the application of the band shape
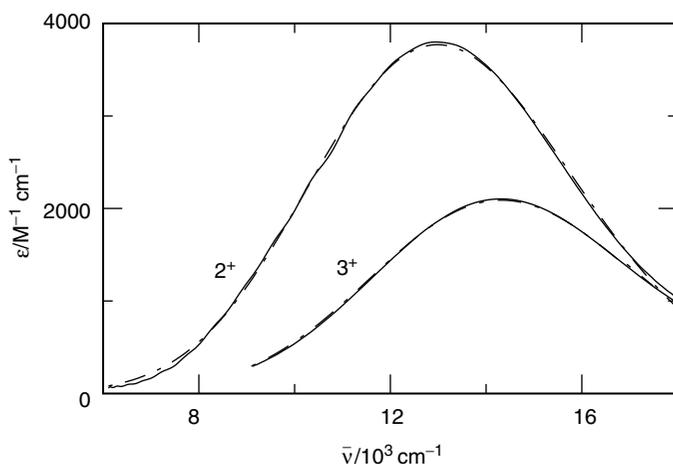
Figure 18 Fits of experimental spectra in acetonitrile (solid lines)[87] to Eqs. [134] and [144] (dash–dotted lines, almost indistinguishable from the experimental spectra on the graph scale). The labeling of the donor–acceptor complexes is according to Ref. 87.

analysis via Eqs. [134] and [144] to two CT complexes studied in Ref. 87 when the traditional band shape analysis[16,17] fails to fit the experimental spectra. The fitting procedure employs the simulated annealing technique in the space of four parameters: λ^I, λ_v, ν_v, and ΔE_{12}.

Polarizable Chromophores

The model of polarizable dipolar chromophores suggests that the 3D nuclear reaction field of the solvent serves as a driving force for electronic transitions. Even in the case of an isotropic solute polarizability, two projections of the reaction field should be included: the longitudinal (parallel to the difference solute dipole) component and the transverse (perpendicular to the difference dipole) component. The δ function in Eq. [18] eliminates integration over only one of these two field component. The integral still can be taken analytically resulting in a closed-form solution for the Franck–Condon factor

$$\text{FCWD}_i^s(\nu) = \beta A_i \sqrt{\frac{\lambda_i |\alpha_i|^3}{|h\nu - X_0|}} e^{-\beta(|\alpha_i||h\nu - X_0| + \lambda_i \alpha_i^2)} I_1\left(2\beta\sqrt{|\alpha_i|^3 \lambda_i |h\nu - X_0|}\right)$$

$$[153]$$

where $I_1(x)$ is the first-order modified Bessel function. The normalization factor

$$A_i = \left(1 - e^{-\beta \lambda_i \alpha_i^2}\right)^{-1} \qquad [154]$$

is included to ensure the identity

$$h \int_{-\infty}^{\infty} \text{FCWD}_i^s(h\nu)d\nu = 1 \qquad [155]$$

In Eq. [153], the parameters α_i are given by Eqs. [74] and [99]. The reorganization energies are defined through the second spectral cumulants and are connected to each other according to Eq. [73]. The boundary of the permissible energy gaps between the two-electron states sets up the range of light frequencies for which the transition intensity is nonzero. The magnitude of the spectral boundary is defined for dipolar chromophores through the difference dipole moment and the polarizability difference

$$X_0 = \Delta I + \Delta E^{\text{disp}} + \Delta F^{\text{ind}} - \frac{\Delta \tilde{m}_0^2}{2|\Delta \tilde{\alpha}_0|} \qquad [156]$$

where ΔE^{disp} and ΔF^{ind} are the differences in dispersion and induction stabilization energies between the two states. When the polarizability does not change with the transition ($\Delta \alpha_0 = 0$), the spectral boundary moves to infinity, $X_0 \to \infty$, and no limiting frequency exists.

The Franck–Condon factors of polarizable chromophores in Eq. [153] can be used to generate the complete vibrational/solvent optical envelopes according to Eqs. [132] and [134]. The solvent-induced line shapes as given by Eq. [153] are close to Gaussian functions in the vicinity of the band maximum and switch to a Lorentzian form on their wings. A finite parameter α_1 leads to asymmetric bands with differing absorption and emission widths. The functions in Eq. [153] can thus be used either for a band shape analysis of polarizable optical chromophores or as probe functions for a general band shape analysis of asymmetric optical lines.

Hybrid Model

Both electronic delocalization and polarizability of the donor–acceptor complex lead to a significant asymmetry between the absorption and emission optical lines as is often observed in experiment.[66,98,99] The importance of this effect can be assessed by comparing the dependence of the observed spectral width on solvent polarity with the prediction of MH theory. Equations [6] and [12] can be combined to give

$$\beta h^2 \langle (\delta \nu)^2 \rangle_{\text{abs/em}} = h\Delta \nu_{\text{st}} + \lambda_\nu (\beta h \nu_\nu - 2) \qquad [157]$$

The MH theory thus predicts that the absorption and emission widths are equal to each other (Eq. [11]) and are linear functions of the Stokes shift
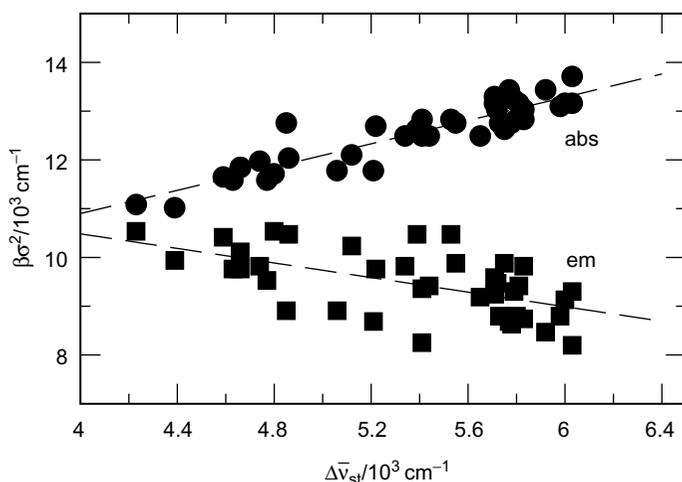
Figure 19 Absorption (circles) and emission (squares) widths (Eq. [80]) versus the Stokes shift for the coumarin dye C153 in 40 molecular solvents according to Ref. 99. The dashed lines are regressions drawn to guide the eye.

$h\Delta\nu_{st}$ with unit slope. This prediction can be dramatically violated for some optical dyes. An illuminating example of such a breakdown is the steady-state spectroscopy of C153 (Figure 19; data according to Ref. 99). The spectral widths in Figure 19 are obtained from half-intensity widths $\Gamma_{abs/em}$ according to Eq. [81] (in Eq. [81], $i = 1$ and $i = 2$ stand for absorption and emission, respectively). As is seen from Figure 19, not only do the spectral widths differ, but also the slopes of $\sigma^2_{abs/em}$ versus $h\Delta\nu_{st}$ have different signs for absorption and emission transitions. This phenomenon is actually well explained by considering a combined effect of the dye polarizability and the electronic coupling between the ground and excited electronic states on the optical band shape.

Within the TSM, the emission width is lower than the absorption width for electronic transitions with a higher magnitude of the dipole moment in the excited state compared to the ground state, as is seen in Figure 17. This is indeed the feature observed in Figure 19. Despite this qualitative agreement, the TSM is very unrealistic due to the neglect of excited electronic states of the chromophore, leading, for example, to a negative excited state polarizability. The polarizability of the excited state of essentially all known chromophores is, on the contrary, positive, and, in the majority of cases, is higher than that of the ground state.[69] To incorporate correctly the chromophore polarizability on the one hand and generate explicit electronic delocalization on the other, a hybrid model was developed.[100] The two states participating in the transition are explicitly considered. Transitions to all other excited states of the chromophore are assumed to result in polarization of the electron density defined by the dipolar polarizability $\bar{\alpha}_{0i}$ ($i = 1, 2$). The total vacuum

polarizability of the solute, α_{0i}, treated as input available from experiment or independent calculations, is thus split into the polarizability from the $1 \leftrightarrow 2$ transition and the component $\bar{\alpha}_{0i}$ from all other transitions. The solvent effect on the transition between the states 1 and 2 then includes three components: (1) solvation of the fixed charges (dipole moments) of the chromophore, (2) self-polarization of the solute's electronic cloud due to polarizability, and (3) change in the electronic occupation numbers induced by the off-diagonal coupling of the transition dipole to the solvent field.

Figure 20 compares the solvent-induced FCWD calculated in the TSM (dash–dotted lines, Eq. [144]), the polarizable model (dashed line, Eq. [153]), and the hybrid model (solid lines). The latter incorporates the effects of both the electronic delocalization between the ground and excited states and polarizability due to the coupling of these two states to all other excited states of the chromophore. The latter model was called the adiabatic polarizable model (APM).[100] The APM thus includes the linear and all nonlinear polarizabilities arising from transitions between the ground and excited states and only linear polarizability for all other states. The emission line is broader than the absorption line due to a higher excited state polarizability when electron delocalization is neglected (Figure 20, dashed lines). The inclusion of electronic delocalization through the transition dipole narrows the emission line and reduces the maxima separation (APM, solid lines). Finally, the neglect of polarizability from higher lying electronic states in the TSM (dash–dotted lines) generates an even narrower emission band. The line shape is therefore a result of a compensation between the polarizability effect tending to increase both the emission width and the Stokes shift for $\Delta\alpha_0 > 0$ and the opposite effect of electronic delocalization.
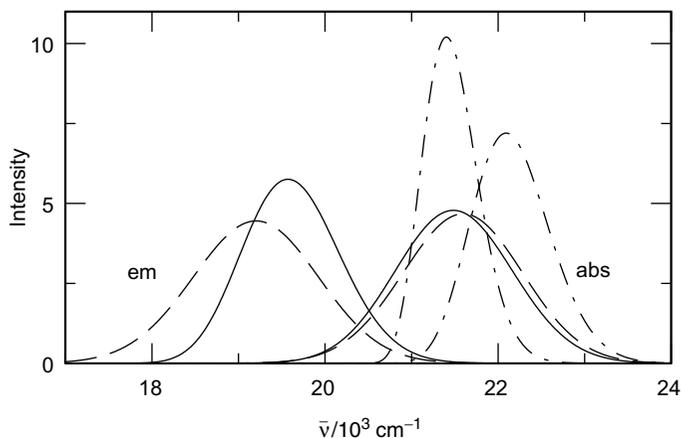

Figure 20 Absorption (abs.) and emission (em.) solvent-induced FCWD of C153 in acetonitrile calculated according to the APM model (solid lines), the TSM (dash–dotted lines), and the polarizable model (Eq. [153], dashed lines).
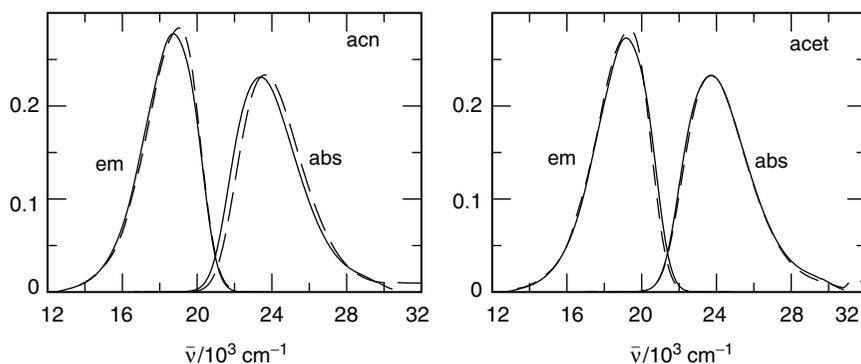
Figure 21 Normalized experimental (dashed lines) and calculated (solid lines) absorption (abs.) and emission (em.) spectra of C153 in acetonitrile (acn) and acetone (acet).

The application of the APM model to the absorption and emission spectra of C153 gives good agreement with experimentally observed spectra in a broad range of solvent polarities.[100] The quality of the calculations is illustrated in Figure 21 where the experimental (dashed lines) and calculated (solid lines) absorption and emission spectra are compared for acetonitrile and acetone as the solvent. The distinction between the optical band shapes calculated on various levels of the theory shown in Figure 20 and the excellent agreement with the experimental results shown in Figure 21 indicate that transitions to higher excited states (polarizability) and solvent-induced mixing between the adiabatic states (transition dipole) are both crucial for reproducing the optical band shape of C153. For this chromophore, the electronic mixing effect is significant due to its high transition dipole moment, $m_{12} = 5.78$ D, close in magnitude to the difference in the excited- and ground-state dipole moments, $\Delta m_0 \approx 7.5$ D. Depending on the relative magnitudes of the polarizability change, $\Delta \alpha_0$, and the transition dipole, m_{12}, polarizability and electronic mixing effects may become more or less important for other optical dyes. For all such cases, the APM provides a general framework for analyzing the FCWD of activated and optical transitions by lifting the two restrictions of the MH theory: the TSM and the neglect of electronic overlap in the FCWD (assumptions 1 and 2 in the Introduction). In fact, the APM also provides a general framework for analyzing the effects of coupling between the vibrational solute modes and the solvent fluctuations (assumption 3),[100] but this problem still requires further studies, both experimental and theoretical.

SUMMARY

The concept of free energy surfaces has proven its vitality over many years of fruitful applications to electron transfer kinetics. The direct connection

of the ET free energy surfaces to the solvent-induced component of the optical Franck–Condon provides a unique possibility to apply the statistical mechanical analysis of ET and CT energetics and to test it on experiment. The band shape analysis of optical profiles is thus the key factor in a successful interplay between theory and experiment.

This chapter outlines some recent advances in the statistical mechanical analysis of the CT energetics. The basic strategy used in this approach is to introduce new physical features of CT activation into the system Hamiltonian used to build the free energy surfaces. These are then applied to calculate the Franck–Condon factors and determine how the changes in the physics of the problem affect the optical observables. This development highlights two fundamental results. First, the MH model of fixed charges solvated in a dense, condensed-phase environment leads to a very accurate representation of the ET energetics in terms of two intersecting parabolas. The static nonlinear solvation effects are generally weak and do not substantially distort the parabolas. There is, however, ample room to modify the free energy surfaces when changes in the electronic density of the donor–acceptor complex are allowed either through polarizability or electronic delocalization. The CT free energies then inherit nonlinear features, and a number of interesting consequences for optical observables can be anticipated. These fascinating phenomena will be the subject of future research.

ACKNOWLEDGMENTS

REFERENCES

1. R. A. Marcus, *Adv. Chem. Phys.*, **106,** 1 (1999). Electron Transfer Past and Future.

2. P. F. Barbara and W. Jarzeba, *Adv. Photochem.*, **15,** 1 (1990). Ultrafast Photochemical Intramolecular Charge and Excited State Solvation.

3. B. Bagchi and N. Gayathri, *Adv. Chem. Phys.*, **107,** 1 (1999). Interplay Between Ultrafast Polar Solvation and Vibrational Dynamics in Electron Transfer Reactions: Role of High-Frequency Vibrational Modes.

4. B. Bagchi and R. Biswas, *Adv. Chem. Phys.*, **109,** 207 (1999). Polar and Nonpolar Solvation Dynamics, Ion Diffusion, and Vibrational Relaxation: Role of Biphasic Solvent Response in Chemical Dynamics.

5. F. O. Raineri and H. L. Friedman, *Adv. Chem. Phys.*, **107,** 81 (1999). Solvent Control of Electron Transfer Reactions.

6. N. S. Hush, *Progr. Inorg. Chem.*, **8,** 391 (1967). Intervalence-Transfer Absorption. Part 2. Theoretical Considerations and Spectroscopic Data. See also, R. S. Mulliken and W. B. Person, *Molecular Complexes,* Wiley, New York, 1969.

7. M. D. Newton, *Adv. Chem. Phys.,* **106,** 303 (1999). Control of Electron Transfer Kinetics: Models for Medium Reorganization and Donor–Acceptor Coupling.

8. J. J. Regan and J. N. Onuchic, *Adv. Chem. Phys.,* **107,** 303 (1999). Electron-Transfer Tubes.

9. L. D. Landau and E. M. Lifshits, *Quantum Mechanics,* Pergamon, Oxford, UK, 1965.

10. A. M. Kuznetsov, *Charge Transfer in Chemical Reaction Kinetics,* Presses Polytechniques et Universitaires Romandes, Lausanne, Switzerland, 1997.

11. J. Ulstrup, *Charge Transfer Processes in Condensed Media,* Springer, Berlin, 1979.

12. R. A. Marcus, *Annu. Rev. Phys. Chem.,* **15,** 155 (1964). Chemical and Electrochemical Electron-Transfer Theory.

13. C. Creutz, *Progr. Inorg. Chem.,* **30,** 1 (1980). Mixed Valence Complexes of d^5–d^6 Metal Centers.

14. P. F. Barbara, T. J. Meyer, and M. A. Ratner, *J. Phys. Chem.,* **100,** 13148 (1996). Contemporary Issues in Electron Transfer Research.

15. M. Lax, *J. Chem. Phys.,* **11,** 1752 (1952). The Franck–Condon Principle and Its Application to Crystals.

16. P. Chen and T. J. Meyer, *Chem. Rev.,* **98,** 1439 (1998). Medium Effects on Charge Transfer in Metal Complexes.

17. M. Bixon and J. Jortner, *Adv. Chem. Phys.,* **106,** 35 (1999). Electron Transfer–From Isolated Molecules to Biomolecules.

18. R. A. Marcus, *J. Phys. Chem.,* **93,** 3078 (1989). Relation Between Charge Transfer Absorption and Fluorescence Spectra and the Inverted Region.

19. M. B. Robin and P. Day, *Adv. Inorg. Chem. Radiochem.,* **10,** 247 (1967). Mixed Valence Chemistry—A Survey and Classification.

20. S. F. Nelsen, *Chem. Eur. J.,* **6,** 581 (2000). "Almost Delocalized" Intervalene Compounds.

21. M. R. Wasielewski, *Chem. Rev.,* **92,** 435 (1992). Photoinduced Electron Transfer in Supermolecular Systems for Artificial Photosynthesis.

22. J. W. Verhoeven, *Adv. Chem. Phys.,* **106,** 603 (1999). From Close Contact to Long-Range Intramolecular Electron Transfer.

23. K. Y. Wong and P. N. Schatz, *Progr. Inorg. Chem.,* **28,** 369 (1981). A Dynamic Model for Mixed-Valence Compounds.

24. M. Born and K. Huang, *Dynamic Theory of Crystal Lattices,* Clarendon Press, Oxford, UK, 1985.

25. D. Chandler and P. Wolynes, *J. Chem. Phys.,* **74,** 4078 (1981). Exploring the Isomorphism Between Quantum Theory and Classical Statistical Mechanics of Polyatomic Fluids.

26. G. A. Voth, *Adv. Chem. Phys.,* **93,** 135 (1996). Path-Integral Centroid Methods in Quantum Statistical Mechanics and Dynamics.

27. S. I. Pekar, *Research in Electron Theory of Crystals,* United States Atomic Energy Commission, Washington, DC, 1963.

28. A. S. Alexandrov and N. Mott, *Polarons and Bipolarons,* World Scientific, Singapore, 1995.

29. A. B. Myers, *Annu. Rev. Phys. Chem.,* **49,** 267 (1998). Molecular Electronic Spectral Broadening in Liquids and Glasses.

30. K. Ando and J. T. Hynes, *Adv. Chem. Phys.,* **110,** 381 (1999). Acid–Base Proton Transfer and Ion Pair Formation in Solution.

31. H. J. Kim and T. J. Hynes, *Am. Chem. Soc.,* **114,** 10508 (1992). A Theoretical Model for S_N1 Ionic Dissociation in Solution. 1. Activation Free Energetics.

32. A. Warshel and W. W. Parson, *Annu. Rev. Phys. Chem.,* **42,** 279 (1991). Computer Simulations of Electron-Transfer Reactions in Solution and in Photosynthetic Reaction Centers.

33. L. W. Ungar, M. D. Newton, and G. A. Voth, *J. Phys. Chem. B,* **103,** 7367 (1999). Classical and Quantum Simulation of Electron Transfer Through a Polypeptide.

34. For a review, see: M. Tachiya, *J. Phys. Chem.,* **93,** 7050 (1989). Relation between the Electron-Transfer Rate and the Free Energy Change of Reaction.

35. D. Chandler, *Phys. Rev. E,* **48,** 2898 (1993). Gaussian Field Model of Fluids with an Application to Polymeric Fluids.

36. C. H. Wang, *Spectroscopy of Condensed Media,* Academic Press, Orlando, FL, 1985.

37. G. van der Zwan and J. T. Hynes, *J. Phys. Chem.,* **89,** 4181 (1985). Time-Dependent Fluorescence Solvent Shift, Dielectric Friction, and Nonequilibrium Solvation in Polar Solvents.

38. D. V. Matyushov and B. M. Ladanyi, *J. Chem. Phys.,* **108,** 6362 (1998). Dispersion Solute–Solvent Coupling in Electron Transfer Reactions. I. Effective Potential.

39. G. Fischer, *Vibronic Coupling,* Academic Press, London, UK, 1984.

40. H. J. Kim and J. T. Hynes, *J. Chem. Phys.,* **96,** 5088 (1992). Equilibrium and Nonequilibrium Solvation and Solvent Electronic Structure. III. Quantum Theory.

41. L. S. Shulman, *Techniques and Applications of Path Integration,* Wiley, New York, 1996.

42. D. V. Matyushov and G. A. Voth, *J. Phys. Chem. A,* **104,** 6470 (2000). Reorganization Parameters of Electronic Transitions in Electronically Delocalized Systems. 1. Charge Transfer Reactions.

43. D. V. Matyushov and B. M. Ladanyi, *J. Phys. Chem. A,* **102,** 5027 (1998). Spontaneous Emission and Nonadiabatic Electron Transfer Rates in Condensed Phases.

44. G. D. Mahan, *Many-Particle Physics,* Plenum Press, New York, 1990.

45. R. J. D. Miller, G. L. McLendon, A. J. Nozik, W. Schmickler, and F. Willig, *Surface Electron Transfer Processes,* VCH Publishers, New York, 1995.

46. A. V. Gorodyskii, A. I. Karasevskii, and D. V. Matyushov, *J. Electroanal. Chem.,* **315,** 9 (1991). Adiabatic Outer-Sphere Electron Transfer Through the Metal–Electrolyte Interface.

47. H. Heitele, *Angew. Chem. Int. Ed. Engl.,* **32,** 359 (1993). Dynamic Solvent Effect on Electron-Transfer Reactions.

48. T. Kakitani, N. Matsuda, A. Yoshimori, and N. Mataga, *Progr. Reaction Kinetics,* **20,** 347 (1995). Present and Future Perspectives of Theoretical Aspects of Photoinduced Charge Separation and Charge Recombination Reactions in Solution.

49. J. B. Birks, *Photophysics of Aromatic Molecules,* Wiley, London, UK, 1970.

50. T. Kakitani and N. Mataga, *J. Phys. Chem.,* **89,** 8 (1985). New Energy Gap Laws for the Charge Separation Process in the Fluorescence Quenching Reaction and the Charge Recombination Process of Ion Pairs Produced in Polar Solvents.

51. J.-K. Hwang and A. Warshel, *J. Am. Chem. Soc.,* **109,** 715 (1987). Microscopic Examination of Free-Energy Relationships for Electron Transfer in Polar Solvents.

52. R. A. Kuharski, J. S. Bader, D. Chandler, M. Sprik, M. L. Klein, and R. W. Impey, *J. Chem. Phys.,* **89,** 3248 (1988). Molecular Model for Aqueous Ferrous–Ferric Electron Transfer.

53. E. A. Carter and J. T. Hynes, *J. Chem. Phys.,* **94,** 5961 (1991). Solvation Dynamics for an Ion Pair in a Polar Solvent: Time-Dependent Fluorescence and Photochemical Charge Transfer.

54. H.-X. Zhou and A. Szabo, *J. Chem. Phys.,* **103,** 3481 (1995). Microscopic Formulation of Marcus' Theory of Electron Transfer.

55. T. Ichiye, *J. Chem. Phys.,* **104,** 7561 (1996). Solvent Free Energy Curves for Electron Transfer Reactions: A Nonlinear Solvent Response Model.

56. R. B. Yelle and T. Ichiye, *J. Phys. Chem. B,* **101,** 4127 (1997). Solvation Free Energy Reaction Curves for Electron Transfer in Aqueous Solutions: Theory and Simulations.

57. P. L. Geissler and D. Chandler *J. Chem. Phys.,* **113,** 9759 (2000). Importance Sampling and Theory of Nonequilibrium Solvation Dynamics in Water.

58. Y.-P. Liu and M. D. Newton, *J. Phys. Chem.,* **99,** 12382 (1995). Solvent Reorganization and Donor/Acceptor Coupling of Electron Transfer Processes: Self-Consistent Reaction Field Theory and Ab Initio Applications.

59. R. Kubo and Y. Toyozawa, *Progr. Theor. Phys.*, **13,** 160 (1955). Application of the Method of Generating Function to Radiative and Non-Radiative Transitions of a Trapped Electron in a Crystal.

60. J. L. Skinner and D. Hsu, *J. Phys. Chem.*, **90,** 4931 (1986). Pure Dephasing of a Two-Level System.

61. D. V. Matyushov and G. A. Voth, *J. Chem. Phys.*, **113,** 5413 (2000). Modeling the Free Energy Surfaces of Electron Transfer in Condensed Media.

62. S. M. Hubig, T. M. Bockman, and J. K. Kochi, *J. Am. Chem. Soc.*, **118,** 3842 (1996). Optimized Electron Transfer in Charge-Transfer Ion Pairs. Pronounced Inner-Sphere Behavior of Olefin Donors.

63. N. Tétreault, R. S. Muthyala, R. S. H. Liu, and R. P. Steer, *J. Phys. Chem. A,* **103,** 2524 (1999). Control of Photophysical Properties of Polyatomic Molecules by Substitution and Solvation: The Second Excited Singlet State of Azulene.

64. T. Asahi, Y. Matsuo, H. Masuhara, H. Koshima, *J. Phys. Chem. A,* **101,** 612 (1997). Electronic Structure and Dynamics of Excited State in CT Microcrystals as Revealed by Femtosecond Diffuse Reflectance Spectroscopy.

65. A. Painelli and F. Terenziani, *J. Phys. Chem. A,* **104,** 11041 (2000). Optical Spectra of Push–Pull Chromophores in Solution: A Simple Model.

66. P. van der Meulen, A. M. Jonkman, and M. Glasbeek, *J. Phys. Chem. A,* **102,** 1906 (1998). Simulation of Solvation Dynamics Using a Nonlinear Response Approach.

67. D. V. Matyushov and G. A. Voth, *J. Phys. Chem. A,* **103,** 10981 (1999). A Theory of Electron Transfer and Steady-State Optical Spectra of Chromophores with Varying Electronic Polarizability.

68. C. Reichardt, *Chem. Rev.*, **94,** 2319 (1994). Solvatochromic Dyes as Solvent Polarity Indicators.

69. W. Liptay, in *Excited States,* E. C. Lim, Ed., Academic Press, New York, 1974, Vol. 1, pp. 129–229. Dipole Moments and Polarizabilities of Molecules in Excited Electronic States.

70. L. R. Pratt, *Mol. Phys.*, **40,** 347 (1980). Effective Field of a Dipole in Non-Polar Polarizable Fluids.

71. P. Vath, M. B. Zimmt, D. V. Matyushov, and G. A. Voth, *J. Phys. Chem. B,* **103,** 9130 (1999). A Failure of Continuum Theory: Temperature Dependence of the Solvent Reorganization Energy of Electron Transfer in Highly Polar Solvents.

72. D. V. Matyushov and G. A. Voth, *J. Chem. Phys.*, **111,** 3630 (1999). A Perturbation Theory for Solvation Thermodynamics: Dipolar-Quadrupolar Liquids.

73. W. Liptay, in *Modern Quantum Chemistry, Part. II,* O. Sinanoğlu, Ed., Academic Press, New York, 1965, pp. 173–198. The Solvent Dependence of the Wavenumber of Optical Absorption and Emission.

74. G. U. Bublitz and S. G. Boxer, *Annu. Rev. Phys. Chem.*, **48,** 213 (1997). Stark Spectroscopy: Applications in Chemistry, Biology, and Materials Science.

75. F. W. Vance, R. D. Williams, and J. T. Hupp, *Int. Rev. Phys. Chem.*, **17,** 307 (1998). Electroabsorption Spectroscopy of Molecular Inorganic Compounds.

76. B. S. Brunshwig, C. Creutz, and N. Sutin, *Coord. Chem. Rev.*, **177,** 61 (1998). Electroabsorption Spectroscopy of Charge Transfer States of Transition Metal Complexes.

77. D. V. Matyushov and B. M. Ladanyi, *J. Chem. Phys.*, **107,** 1375 (1997). Nonlinear Effects in Dipole Solvation. II. Optical Spectra and Electron Transfer Activation.

78. T. Asahi, M. Ohkohchi, R. Matsusaka, N. Mataga, R. P. Zhang, A. Osuka, and K. Maruyama, *J. Am. Chem. Soc.*, **115,** 5665 (1993). Intramolecular Photoinduced Charge Separation and Charge Recombination of the Product Ion Pair States of a Series of Fixed-Distance Dyads of Porphyrins and Quinones: Energy Gap and Temperature Dependences of the Rate Constants.

79. H. Heitele, F. Pöllinger, T. Häberle, M. E. Michel-Beyerle, and H. A. Staab, *J. Phys. Chem.*, **98,** 7402 (1994). Energy Gap and Temperature Dependence of Photoinduced Electron Transfer in Porphyrin–Quinone Cyclophanes.

80. T. Fonseca, B. M. Ladanyi, and J. T. Hynes, *J. Phys. Chem.,* **96,** 4085 (1992). Solvation Free Energies and Solvent Force Constant.

81. D. V. Matyushov and B. M. Ladanyi, *J. Chem. Phys.,* **110,** 994 (1999). A Perturbation Theory and Simulations of the Dipole Solvation Thermodynamics: Dipolar Hard Spheres.

82. H. Meirovitch, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., Wiley-VCH, New York, 1998, Vol. 12, pp. 1–74. Calculation of the Free Energy and the Entropy of Macromolecular Systems by Computer Simulations.

83. J. Åqvist and T. Hansson, *J. Phys. Chem.,* **100,** 9512 (1996). On the Validity of Electrostatic Linear Response in Polar Solvents.

84. M. V. Basilevsky, G. E. Chudinov, and M. D. Newton, *Chem. Phys.,* **179,** 263 (1994). The Multi-Configurational Adiabatic Electron Transfer Theory and Its Invariance under Transformations of Charge Density Basis Functions.

85. B. S. Brunshwig and N. Sutin, *Coord. Chem. Rev.,* **187,** 233 (1999). Energy Surfaces, Reorganization Energies, and Coupling Elements in Electron Transfer.

86. A. Myers Kelley, *J. Phys. Chem. A,* **103,** 6891 (1999). Resonance Raman Intensity Analysis of Vibrational and Solvent Reorganization in Photoinduced Charge Transfer.

87. S. F. Nelsen, R. F. Ismagilov, K. E. Gentile, and D. R. Powell, *J. Am. Chem. Soc.,* **121,** 7108 (1999). Temperature Effects on Electron Transfer within Intervalence Bis(Hydrazine) Radical Cations.

88. I. R. Gould, R. H. Young, L. J. Mueller, A. C. Albrecht, and S. Farid, *J. Am. Chem. Soc.,* **116,** 8188 (1994). Electronic Structures of Exciplexes and Excited Charge-Transfer Complexes.

89. D. V. Matyushov, R. Schmid, and B. M. Ladanyi, *J. Phys. Chem. B,* **101,** 1035 (1997). A Thermodynamic Analysis of the π^* and $E_T(30)$ Polarity Scales.

90. C. Wang, B. K. Mohney, B. B. Akhremitchev, and G. C. Walker, *J. Phys. Chem. A,* **104,** 4314 (2000). Ultrafast Infrared Spectroscopy of Vibrational States Prepared by Photoinduced Electron Transfer in $(CN)_5FeCNRu(NH_3)_5^-$.

91. C. J. F. Böttcher, *Theory of Electric Polarization,* Elsevier, Amsterdam, The Netherlands, 1973.

92. D. P. Craig and T. Thirunamachandran, *Molecular Quantum Electrodynamics. An Introduction to Radiation Molecule Interaction,* Dover, Mineola, NY, 1984.

93. I. R. Gould, D. Noukakis, L. Gomez-Jahn, R. H. Young, G. L. Goodman, and S. Farid, *Chem. Phys.,* **176,** 439 (1993). Radiative and Nonradiative Electron Transfer in Contact Radical-Ion Pairs.

94. R. S. Mulliken and W. B. Person, *Molecular Complexes,* Wiley, New York, 1969.

95. J. L. Lewis and M. Maroncelli, *Chem. Phys. Lett.,* **282,** 197 (1998). On the (Uninteresting) Dependence of the Absorption and Emission Transition Moments of Coumarin 153 on Solvent.

96. D. V. Matyushov and G. A. Voth, *J. Phys. Chem. A,* **104,** 6470 (2000). Reorganization Parameters of Electronic Transitions in Electronically Delocalized Systems. 2. Optical Spectra.

97. C. Lambert and G. Nöll, *J. Am. Chem. Soc.,* **121,** 8434 (1999). The Class II/III Transition in Triarylamine Redox Systems.

98. D. Bingemann and N. P. Ernsting, *J. Chem. Phys.,* **102,** 2691 (1995). Femtosecond Solvation Dynamics Determining the Band Shape of Stimulated Emission from a Polar Styryl Dye.

99. L. Reynolds, J. A. Gardecki, S. J. V. Frankland, M. L. Horng, and M. Maroncelli, *J. Phys. Chem.,* **100,** 10337 (1996). Dipole Solvation in Nondipolar Solvents: Experimental Studies of Reorganization Energies and Solvation Dynamics.

100. D. V. Matyushov and M. D. Newton, *J. Phys. Chem. A,* **105,** 8516 (2001). Understanding the Optical Band Shape: Coumarin-153 Steady-State Spectroscopy.