Reviews in Computational Chemistry Volume 23

Edited by

Kenny B. Lipkowitz
Thomas R. Cundari

Editor Emeritus

Donald B. Boyd

BICENTENNIAL
1807
WILEY
2007
BICENTENNIAL
WILEY-VCH

Reviews in Computational Chemistry Volume 23

BICENTENNIAL
1807
WILEY
2007
BICENTENNIAL

THE WILEY BICENTENNIAL–KNOWLEDGE FOR GENERATIONS

Each generation has its unique needs and aspirations. When Charles Wiley first opened his small printing shop in lower Manhattan in 1807, it was a generation of boundless potential searching for an identity. And we were there, helping to define a new American literary tradition. Over half a century later, in the midst of the Second Industrial Revolution, it was a generation focused on building the future. Once again, we were there, supplying the critical scientific, technical, and engineering knowledge that helped frame the world. Throughout the 20th Century, and into the new millennium, nations began to reach out beyond their own borders and a new international community was born. Wiley was there, expanding its operations around the world to enable a global exchange of ideas, opinions, and know-how.

For 200 years, Wiley has been an integral part of each generation's journey, enabling the flow of information and understanding necessary to meet their needs and fulfill their aspirations. Today, bold new technologies are changing the way we live and learn. Wiley will be there, providing you the must-have knowledge you need to imagine new worlds, new possibilities, and new opportunities.

Generations come and go, but you can always count on Wiley to provide you the knowledge you need, when and where you need it!

WILLIAM J. PESCE
PRESIDENT AND CHIEF EXECUTIVE OFFICER

PETER BOOTH WILEY
CHAIRMAN OF THE BOARD

Reviews in Computational Chemistry Volume 23

Edited by

Kenny B. Lipkowitz
Thomas R. Cundari

Editor Emeritus

Donald B. Boyd

BICENTENNIAL
1807 WILEY 2007
WILEY-VCH

Kenny B. Lipkowitz
Department of Chemistry
Howard University
525 College Street, N.W.
Washington, D. C., 20059, U.S.A
ken.lipkowitz@cox.net

Thomas R. Cundari
Department of Chemistry
University of North Texas
Box 305070,
Denton, Texas 76203-5070, U.S.A.
tomc@unt.edu

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

Copyright © 2007 by John Wiley & Sons, Inc. All rights reserved

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee of the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www. copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at http://www.wiley.com/go/permission.

Limit of Liability/Disclaimer of Warranty: While the Publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Wiley Bicentennial Logo: Richard J. Pacifico

Library of Congress Cataloging-in-Publication Data:

ISBN 978-0-470-08201-0
ISSN 1069-3599

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Preface

Students wanting to become computational chemists face a steep learning curve that can be intellectually and emotionally challenging. Those students are expected to know basic physics, from quantum mechanics to statistical mechanics, along with a full comprehension of electricity and magnetism; they are required to be conversant in calculus, algebra, graph theory, and statistics; they are expected to be cognizant of algorithmic issues of computer science, and they are expected to be well versed in the experimental aspects of the topic they intend to model, whether it be in the realm of materials science, biology, or engineering. Beginning in the mid-1990s, and continuing into this century, there appeared a series of books on molecular modeling and computational chemistry that addressed the needs of such students. Those books are very well organized, are extremely well written, and have been received enthusiastically by the community at large.

The editors of *Reviews in Computational Chemistry* knew that such books would avail themselves to a hungry public, and further knew that only an introduction to the wide range of topics in this discipline could be covered in a single book. Accordingly, a decision was made to provide lengthier, more detailed descriptions of the many computational tools that a computational scientist would need for his or her career. *Reviews in Computational Chemistry* thus set out on a trajectory of providing pedagogically driven chapters for both the novice who wants to become a computational molecular scientist as well as for the seasoned professional who wants to quickly learn about a computational method outside of his or her area of expertise. In this, the 23rd volume of the series, we continue that tradition by providing seven chapters on a wide variety of topics.

Most bench chemists who use software for computing quantum mechanical properties, structures, and energies of molecular systems are well aware of the n^4 bottleneck associated with the calculation of the required electron repulsion integrals and quickly find this scaling problem to be a major impediment to their studies. In Chapter 1, Christian Ochsenfeld, Jörg Kussmann, and Daniel Lambrecht provide a tutorial on the topic of linear scaling methods in quantum chemistry. The authors begin by putting into perspective the existing scaling problems associated with approximating the solution to the Schrödinger equation. They review the basics of self-consistent field (SCF) theory within the Born–Oppenheimer (BO) approximation and focus the readers' attention on the interplay between the cubic scaling for diagonalization of the Fock matrix and the quadratic scaling for the formation of that matrix. They then describe how one can reduce this problem by selecting numerically significant integrals using Schwarz or multipole-based integral estimates, illustrating those concepts with easy-to-follow diagrams and demonstrating the results with simple plots. The calculation of integrals by multipole expansion is then presented, beginning with a very simple example that shows the novice how individual pair-wise interactions between point charges can be collected into charge distributions that, when combined with a clever tree algorithm, can avoid the quadratic scaling step. The authors provide the reader with a basic understanding of multipole expansions and then describe the fast multipole method (FMM) and its generalization for continuous (Gaussian) distributions, the continuous FMM method, before providing an overview of other multipole expansions and tree codes used to speed up the calculation of two-electron integrals. Exactly how this linear scaling is accomplished is illustrated nicely by the authors through the use of an example of a molecule of substantial size—an octameric fragment of DNA. Having described how to reduce the scaling behavior for the construction of the Coulomb part of the Fock matrix, the authors bring to the fore the remaining component within Hartree–Frock (HF) theory, the exchange part [also required for hybrid density functional theory (DFT)]. As before, with a didactic style, the authors show how one can exploit localization properties of the density matrix to achieve linear scaling of the exchange part of the Hamiltonian. Then the authors show how one can avoid the conventional diagonalization of the assembled Fock matrix and reduce what would be a cubic scaling process to one that is linear. The tensor formalism is introduced, properties of the one-particle density matrix are described, and the density-matrix-based energy functional is introduced to solve the SCF problem. The authors go beyond just the computation of energies by then explaining energy gradients and molecular response properties. The chapter concludes with an overview of what it takes to reduce the scaling behavior of post-HF methods for large systems. Linear scaling techniques in quantum chemistry are becoming more widely implemented in software packages. The bench chemist who is inclined to use this software ought not treat it as a black box, but instead should be cognizant of the assumptions, approximations and pitfalls associated with linear scaling methodology. This chapter makes all of this visible to the user in a clear and coherent manner.

The BO approximation is sufficient in quantum chemistry for describing most chemical processes. However, many nonadiabatic processes exist in nature that cannot be described adequately within this context, examples of which include the ubiquitous photophysical and photochemical processes associated with photosynthesis, vision and charge transfer reactions, among others. Nonadiabatic phenomena occur when two or more potential energy surfaces approach each other, and the coupling between those surfaces becomes important. Conical intersections are the actual crossings of those surfaces. In Chapter 2, Spiridoula Matsika highlights where the BO approximation breaks down, the differences between adiabatic and diabatic representations for studying nuclear dynamics, and the significance of the noncrossing rule. She follows this with an introduction to, and explanation of, conical intersections by addressing the Jahn–Teller effect, symmetry allowed conical intersections, accidental intersections, the branching plane, and how topography is used to characterize conical intersections. Post-HF methods, including MCSCF, MRCI, and CASPT2, and single reference methods are surveyed along with the many considerations a user must take into account when choosing a particular electronic structure method for computing conical intersections. These explanations are followed by a description of how to actually locate conical intersections using Lagrange multipliers and projected gradient techniques. Then, with this background in hand, the author provides us with several applications to show what can be done to analyze such intersections. Matsika's review of the field covers inorganic and organic molecules but focuses primarily on biologically relevant systems, especially on nucleic acids. Most of the tutorial focuses on two-state conical intersections, but a description of three-state intersections is also given. That is followed by a discussion of spin-orbit coupling that, when included in the Hamiltonian, provides new and qualitatively different effects in the radiationless behavior of chemical systems. In this part of the review, the author points out how such effects can couple states of different spin multiplicity whose intersections would otherwise not be conical, along with explaining the influence of such coupling on systems with an odd number of electrons for which there are qualitative changes in the characteristics of the intersection. Novice molecular modelers intending to carry out quantum mechanical calculations are encouraged to peruse this chapter to determine whether their systems are susceptible to nonadiabatic processes that would require evaluation of conical intersections, and then to read this tutorial to ensure that a proper treatment of the system is being pursued. We also urge the reader to see the chapter by Michael A. Robb, Marco Garavelli, Massimo Olivucci, and Fernando Bernardi in Volume 15 that also examined some of these issues.

Most of us are content with computing structures, energies, and some properties of molecules in the ground or excited states. For other researchers, however, kinetic information is required when rate constants for chemical reactions must be evaluated. How does one go about computing rate constants for, say, a large catalytic system like an enzyme in which a critical step is the transfer of a light particle such as a hydride or a proton that is subject to quantum tunneling effects? In Chapter 3, Antonio Fernandez-Ramos, Benjamin Ellingson, Bruce Garrett, and Donald Truhlar provide a tutorial on the topic of variational transition state theory (VTST) with multidimensional tunneling that gives us a good starting point to answer that question. The tutorial begins with a description of conventional transition state theory (TST), highlighting the tenets upon which it is constructed, the merits of its use, and pointing out that it only provides an approximation to the true rate constant because it assesses the one-way flux through a dividing surface that is appropriate for small, classic vibrations around a saddle point. Canonical and other types of variational TSTs are introduced by the authors at this point along with highlighting the influences that quantum effects have on the reaction coordinate; a section on practical methods for quantized VTST calculations is subsequently presented to address these concerns. Here the authors cover some algorithms used to calculate the reaction path by describing the minimum energy path and an algorithm for a variational reaction path, showing us how to evaluate partition functions using both rectilinear and curvilinear coordinates, describing the influence of anharmonic vibrational levels on those partition functions, and demonstrating how to calculate the number of states needed for microcanonical variational theory calculations. The authors then focus on quantum effects on reaction coordinate motion; such effects are usually dominated by tunneling but also include nonclassical reflection, both of which are incorporated in a multiplicative transmission coefficient. In this part of the tutorial, multidimensional tunneling corrections are highlighted for the novice. Because the reaction path is a curvilinear coordinate, the curvature of that path couples motion along the reaction coordinate to local vibrational modes that are perpendicular to it. The coupling causes the system to take a shorter path than the reaction coordinate by tunneling. Both small- and large-curvature tunneling motions, with and without vibrational excitations, are compared in this part of the tutorial. In the second part of Chapter 3, the authors deal with VTST in complex systems. Because an analytical potential energy surface (PES) is usually not available, the authors begin by describing how one can build the PES from electronic structure calculations using "on the fly" quantum methods for direct dynamics calculations, i.e., without the fitting of those energies in the form of a potential function, and then they explain how one can derive those surfaces by interpolation through the use of their multiconfiguration molecular mechanics algorithm (MCMM) and by a mapping procedure. This, in turn, is followed by a description of how to incorporate both low-level and high-level calculations to generate the PES so as to make the calculation of rate constants very fast. In this chapter, the authors also cover other topics of relevance to the prediction of accurate rate calculations, including reactions in liquids and, because there is more than one reaction coordinate, how to use ensemble-averaged VTST. Finally, the authors provide two insightful examples, one in the gas phase and the other in solution, to demonstrate the speed and accuracy of modern methods for predicting rate constants.

Chemists have traditionally worked on a scale of size ranging from Angstrom units to nanometers; we take a molecular view of the scientific problems at hand in which atomic-level detail is *de rigueur*. What happens, though, if the career path you take or the research project you are engaged with involves, say, long-chain polymeric systems consisting of a few thousand monomers in a melt where the relevant length scales run from bond lengths to the contour length of the chain, which is on the order of micrometers, and where the relevant relaxation times increase as $N^{3.4}$ for chains of length N? One approach to addressing such a problem is to invoke coarse-grained techniques, and in Chapter 4, Roland Faller shows us how this is accomplished. The author sets the stage for such a computational scene by first pointing out that one needs to define the system to be evaluated and then one needs to select a suitable model to combine simulations on a variety of length and time scales. An explanation is then provided about how one assigns interaction sites on the coarse-grain scale. Because it may be necessary to use two or more models to cover the range of relevant interactions of interest to the scientist or engineer, the author emphasizes that a meaningful mapping between scales is needed for meaningful results. For example, atomistic models can treat lengths of scale from a few hundred picometers to tens of nanometers, whereas meso-scale models are useful from the multi-nanometer scale up to a few micrometers in size, but if we want to enter the realm of micrometers and beyond, a second or third mapping is needed. A brief tutorial on the various types of existing mapping strategies is given for the novice modeler. First, static mapping methods are discussed, including single-chain distribution models, iterative structural coarse-graining, and mapping onto simple models. Then the author teaches us about dynamic mapping, including mapping by chain diffusion, mapping through local correlation times, and direct mapping of the Lennard–Jones time. Following that part of the tutorial, the author describes coarse-grained Monte Carlo simulations and reverse mapping, in which atomistic detail is reintroduced at the end of the simulation. Faller then goes beyond polymers to describe examples of coarse-grain modeling of lipid bilayer systems. Nowadays the scientific community is expecting more than it has in the past from a computational chemist in terms of both the quality and the scope of the modeling endeavor. Because advances in computing machinery will not likely allow us to take a fully atomistic approach to such modeling in the next decade, this chapter, written from the perspective of an engineer, gives us the insights needed to carry out simulations on both small and large scales.

Many of the readers of this book series work in the pharmaceutical industry where informatics is especially relevant. Different databases are available free of charge in some cases but more usually for a fee, even if that fee comes from within the company where large investments are made in developing a proprietary database. One might want to know in advance if a given database is more diverse than another, or one might want to answer the question: "How much additional chemical diversity is added when we double the size of the current database?" Given the costs of generating compound libraries (real or virtual), answering such a question requires that a management team should have good insights about information in general; otherwise, poor decisions could have costly ripple effects that negatively influence both big and small companies alike. In Chapter 5, Jeffrey Godden and Jürgen Bajorath provide a tutorial on the analysis of information content that focuses on Shannon entropy (SE) applied to molecules. Here, any structural representation of a molecule, including the limitless number of molecular descriptors currently in use today or to be used in the future, is to be understood as a communication carrying a specific amount of information. The authors begin their tutorial by providing a historical account of how this area of informatics developed, and they explain the relationship between Shannon entropy used in the telecommunications industry and information being conveyed in a typical molecular database. Simple equations and simple examples are used to illustrate the concepts. The authors then use these concepts to show the reader how one would compare descriptors in, say, the Available Chemical Directory (ACD) with those in the Molecular Drug Data Report (MDDR). Here it is emphasized that because SE is a nonparametric distribution measure, this entropy-based approach is well suited for information content analysis of descriptors with different units, different numerical ranges, and different variability. Now one can begin answering questions such as "Which descriptors carry high levels of information for a specific compound set?" [which in turn could be used for deriving a statistically meaningful quantitative structure-activity relationship/quantitative structure-property relationship (QSAR/QSPR) model] and "Which descriptors carry low levels of information?" The authors continue their tutorial by describing the influence of boundary effects on such analyses, and they give hints about what to do and what not to do for the novice modeler who would otherwise become trapped in one or more computational pitfalls that are not visible to a beginner. An extension of the method called differential Shannon entropy (DSE) analysis is then introduced, and the reader is shown how DSE can reveal descriptors that are sensitive to systematic differences in the properties of different databases or classes of molecules. A brief glimpse into the information content of organic molecules is given, and then uses of SE in quantum mechanical calculations, molecular dynamics simulations, and other types of modeling are presented. The authors end this chapter with examples of SE and DSE analysis for the modeling of physicochemical properties and for accurate classification of molecules, a topic that is described in the following chapter.

Many of us make binary, black/white, either/or type decisions every day: "Should I buy this house now or wait?", "Should I say something to my boss or not?", and so on. These types of queries are commonly posed in a scientific setting as well, where, for example, the question might be on a health-related issue like "Is this cell cancerous or not?", and in the business world, where we might ask, "Is this lead molecule toxic or not?" Robust methods for simple classification do exist. One of the more popular and successful techniques involves a group of supervised learning methods called support vector machines (SVM) that can be applied to classification as well as to regression. In Chapter 6, Ovidiu Ivanciuc covers the topic of SVMs in chemistry. Following a historical introduction that covers the development of SVM and other kernel-based techniques, the author provides a non-mathematical introduction to SVM, beginning with the classification of linearly separable classes and then continues by teaching us about partitioning classes that cannot be separated with a linear classifier, which is a situation where mapping into a high-dimensional feature space is accomplished with nonlinear functions called kernels. The author uses a simple "synthetic" dataset to demonstrate the concepts for the beginner, and he provides simple MATLAB-generated plots to illustrate what should and should not be done for both classification and regression. The next topic unveiled in this tutorial is pattern classification, which is used, for example, in clinical diagnostics and in speech analysis as well as for chemical problems where one might need to recognize, say, the provenance of agricultural products like wine, olive oil, or honey based on chemical composition or spectral analysis. Again, very simple examples and clear plots are presented to show the utility of this method for pattern classification along with restrictions on its use. Because SVMs are based on structural risk minimization, which in turn is derived from statistical learning theory, the machine algorithm is considered deterministic. Accordingly, concepts related to the expected risk or to the expected error are next introduced by describing the Vapnik–Chervonenkis dimension, a construct used to indicate how high in complexity a classifier must be to minimize the empirical risk. With this background, pattern classification with linear support machines is described for the reader, showing how to establish the optimum separation hyperplane for a given finite set of learning patterns. The equations needed to accomplish this are developed in a clear and concise manner, and again, simple examples are given for SVM classification of linearly separable data, and then for nonlinearly separable data. Because there are cases where complex relationships exist between input parameters and the class of a pattern, the author devotes a full section to nonlinear support machines, showing first how patterns are mapped to a feature space, and then describing feature functions and kernels, including linear kernels, polynomial kernels, and Gaussian and exponential radial basis function kernels, along with others like neural, Fourier series, spline, additive, and tensor product kernels. Also covered in this section of the chapter are weighted SVMs for imbalanced classification, and multiclass SVM classification. A significant portion of the review describes SVM regression. Here simple examples and clear diagrams are used to illustrate the concepts being described. This precedes a section on optimizing the SVM model, i.e., finding good prediction statistics. Given this background, the author then spends the remainder of the chapter first on practical aspects of SVM classification, providing guidelines for their use in cheminformatics and QSAR, and then on applications of SVM regression. Several examples from this section of the review include predicting the mechanism of action for polar and nonpolar narcotic compounds, classifying the carcinogenic activity of polycyclic aromatic hydrocarbons, and using SVM regression for developing a QSAR for benzodiazepine receptor ligands. The chapter ends with a literature review of SVM applications in chemistry. SVM resources on the Web are identified, and then SVM software for chemists interested in cheminformatics and computational chemistry are tabulated in a convenient, easy-to-read list that describes what those programs can do.

In the final chapter, Donald B. Boyd presents a historical account of the growth of computational chemistry that covers hardware, software, events, trends, hurdles, successes, and people in the pharmaceutical industry. In the 1960s, there were no computational chemists in that industry. That term had not yet been invented. A smattering of theoretical chemists, statisticians, quantum chemists, and crystallographers were among the computer-savvy scientists at that time period who set the stage for modern computational science and informatics to be played out in the pharmaceutical industry. The chapter conveys to the novice molecular modeler what it was like to rely on huge, offsite mainframes like the IBM 7094 or onsite machines used mostly for payroll and bookkeeping with little time available for scientific computing. The smell and sounds of a computer center, replete with loud chunking noises from card punch machines and high-pitch ripping sounds from line printers, are well depicted for the young reader who grew up with quiet personal computers, graphical interfaces, and the Internet, all of which were only futuristic thoughts in the 1960s. Also brought to light is the fact that the armamentarium of the computational scientist in those days consisted of programs like Extended Hückel Theory and early versions of semiempirical HF-based quantum methods like CNDO. Molecular mechanics for fast geometry optimization of pharmaceutically relevant molecules was just being developed in academic laboratories. Preparation of input data involving atomic coordinates was a tedious process because it involved using handheld mechanical models, protractors, and tables of standard bond lengths and bond angles. But despite hardware and software limitations, there were useful insights from such computational endeavors deemed valuable by management and that, in turn, led eventually to the accelerated growth of computational chemistry in the 1980s and the fruition of such research in the 1990s. Woven into this historical tapestry are the expensive threads of hardware purchases beginning with IBM and CDC mainframes and followed by interactive machines like the DEC-10, department-sized super-minicomputers like the VAX 11/780, PCs, Macintoshes, UNIX workstations, supercomputers, array processors, servers, and now clusters of PCs. Interlaced throughout this story are the historical strands of software availability and use of venerable programs like CNDO/2, MINDO/3, MOPAC, MM2, and molecular modeling packages like CHEMGRAF, SYBYL, and MacroModel along with other programs of utility to the pharmaceutical industry, including MACCS and REACCS. Along with the inanimate objects of computers and software, this chapter reveals some social dynamics involving computational chemists, medicinal chemists, and management. Stitched throughout this chapter are the nascent filaments of what we now call informatics, showing how the fabric of that industry evolved from dealing with a small number of molecules to now treating the enormous numbers of potential drug candidates coming from experimental combi-chem studies or from virtual screening by computer. This chapter conveys to the reader, in a compelling way, both the hardships and the successes of computational chemistry in the pharmaceutical industry.

Reviews in Computational Chemistry is highly rated and well received by the scientific community at large; the reason for these accomplishments rests firmly on the shoulders of the authors whom we have contacted to provide the pedagogically driven reviews that have made this ongoing book series so popular. To those authors we are especially grateful.

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 and previous volumes for their excellent chapters.

Kenny B. Lipkowitz
Washington

Thomas R. Cundari
Denton

April 2006

Contents

1. **Linear-Scaling Methods in Quantum Chemistry** 1
Christian Ochsenfeld, Jörg Kussmann, and Daniel S. Lambrecht

Introduction 1
Some Basics of SCF Theory 4
Direct SCF Methods and Two-Electron Integral Screening 8
    Schwarz Integral Estimates 9
    Multipole-Based Integral Estimates (MBIE) 11
Calculation of Integrals via Multipole Expansion 15
    A First Example 16
    Derivation of the Multipole Expansion 20
    The Fast Multipole Method: Breaking the Quadratic Wall 27
    Fast Multipole Methods for Continuous Charge Distributions 32
    Other Approaches 34
Exchange-Type Contractions 35
The Exchange-Correlation Matrix of KS-DFT 40
Avoiding the Diagonalization Step—Density Matrix-Based SCF 42
    General Remarks 42
    Tensor Formalism 43
    Properties of the One-Particle Density Matrix 47
    Density Matrix-Based Energy Functional 49
    "Curvy Steps" in Energy Minimization 53
    Density Matrix-Based Quadratically Convergent SCF (D-QCSCF) 55
    Implications for Linear-Scaling Calculation of SCF Energies 56
SCF Energy Gradients 57
Molecular Response Properties at the SCF Level 59
    Vibrational Frequencies 60
    NMR Chemical Shieldings 61
    Density Matrix-Based Coupled Perturbed SCF (D-CPSCF) 62

Outlook on Electron Correlation Methods for Large Systems 64
  Long-Range Behavior of Correlation Effects 67
  Rigorous Selection of Transformed Products via Multipole-Based Integral Estimates (MBIE) 72
  Implications 72
Conclusions 73
References 74

2. Conical Intersections in Molecular Systems 83
Spiridoula Matsika

Introduction 83
General Theory 85
  The Born–Oppenheimer Approximation and its Breakdown: Nonadiabatic Processes 85
  Adiabatic-Diabatic Representation 87
  The Noncrossing Rule 88
  The Geometric Phase Effect 89
  Conical Intersections and Symmetry 90
  The Branching Plane 91
  Characterizing Conical Intersections: Topography 93
  Derivative Coupling 96
Electronic Structure Methods for Excited States 97
  Multiconfiguration Self-Consistent Field (MCSCF) 98
  Multireference Configuration Interaction (MRCI) 99
  Complete Active Space Second-Order Perturbation Theory (CASPT2) 101
  Single Reference Methods 101
  Choosing Electronic Structure Methods for Conical Intersections 102
Locating Conical Intersections 102
Dynamics 104
Applications 105
  Conical Intersections in Biologically Relevant Systems 106
Beyond the Double Cone 110
  Three-State Conical Intersections 110
  Spin-Orbit Coupling and Conical Intersections 112
Conclusions and Future Directions 115
Acknowledgments 116
References 116

3. **Variational Transition State Theory with Multidimensional Tunneling** 125
Antonio Fernandez-Ramos, Benjamin A. Ellingson, Bruce C. Garrett, and Donald G. Truhlar

Introduction 125
Variational Transition State Theory for Gas-Phase Reactions 127
  Conventional Transition State Theory 127
  Canonical Variational Transition State Theory 131
  Other Variational Transition State Theories 136
  Quantum Effects on the Reaction Coordinate 138
Practical Methods for Quantized VTST Calculations 139
  The Reaction Path 140
  Evaluation of Partition Functions 148
  Harmonic and Anharmonic Vibrational Energy Levels 158
  Calculations of Generalized Transition State Number of States 163
Quantum Effects on Reaction Coordinate Motion 163
  Multidimensional Tunneling Corrections Based on the Adiabatic Approximation 164
  Large Curvature Transmission Coefficient 172
  The Microcanonically Optimized Transmission Coefficient 188
Building the PES from Electronic Structure Calculation 190
  Direct Dynamics with Specific Reaction Parameters 191
  Interpolated VTST 192
  Dual-Level Dynamics 199
Reactions in Liquids 203
Ensemble-Averaged Variational Transition State Theory 206
Gas-Phase Example: $H + CH_4$ 212
Liquid-Phase Example: Menshutkin Reaction 217
Concluding Remarks 221
Acknowledgments 222
References 222

4. **Coarse-Grain Modeling of Polymers** 233
Roland Faller

Introduction 233
Defining the System 235
  Choice of Model 235
  Interaction Sites on the Coarse-Grained Scale 237

Static Mapping 238
  Single-Chain Distribution Potentials 238
  Simplex 239
  Iterative Structural Coarse-Graining 240
  Mapping Onto Simple Models 245
Dynamic Mapping 246
  Mapping by Chain Diffusion 247
  Mapping through Local Correlation Times 247
  Direct Mapping of the Lennard-Jones Time 250
Coarse-Grained Monte Carlo Simulations 250
Reverse Mapping 252
A Look Beyond Polymers 254
Conclusions 257
Acknowledgments 258
References 258

5. **Analysis of Chemical Information Content Using Shannon Entropy** **263**
Jeffrey W. Godden and Jürgen Bajorath

Introduction 263
Shannon Entropy Concept 265
Descriptor Comparison 269
Influence of Boundary Effects 273
Extension of SE Analysis for Profiling of Chemical Libraries 275
Information Content of Organic Molecules 278
Shannon Entropy in Quantum Mechanics, Molecular Dynamics, and Modeling 279
Examples of SE and DSE Analysis 280
Conclusions 286
References 287

6. **Applications of Support Vector Machines in Chemistry** **291**
Ovidiu Ivanciuc

Introduction 291
A Nonmathematical Introduction to SVM 292
Pattern Classification 301
The Vapnik–Chervonenkis Dimension 306
Pattern Classification with Linear Support Vector Machines 308
  SVM Classification for Linearly Separable Data 308
  Linear SVM for the Classification of Linearly Non-Separable Data 317

Nonlinear Support Vector Machines 323
  Mapping Patterns to a Feature Space 323
  Feature Functions and Kernels 326
  Kernel Functions for SVM 329
  Hard Margin Nonlinear SVM Classification 334
  Soft Margin Nonlinear SVM Classification 335
  ν-SVM Classification 337
  Weighted SVM for Imbalanced Classification 338
  Multi-class SVM Classification 339
SVM Regression 340
Optimizing the SVM Model 347
  Descriptor Selection 347
  Support Vectors Selection 348
  Jury SVM 348
  Kernels for Biosequences 349
  Kernels for Molecular Structures 350
Practical Aspects of SVM Classification 350
  Predicting the Mechanism of Action for Polar and Nonpolar Narcotic Compounds 352
  Predicting the Mechanism of Action for Narcotic and Reactive Compounds 355
  Predicting the Mechanism of Action from Hydrophobicity and Experimental Toxicity 359
  Classifying the Carcinogenic Activity of Polycyclic Aromatic Hydrocarbons 360
  Structure-Odor Relationships for Pyrazines 361
Practical Aspects of SVM Regression 362
  SVM Regression QSAR for the Phenol Toxicity to *Tetrahymena pyriformis* 363
  SVM Regression QSAR for Benzodiazepine Receptor Ligands 366
  SVM Regression QSAR for the Toxicity of Aromatic Compounds to *Chlorella vulgaris* 367
  SVM Regression QSAR for Bioconcentration Factors 369
Review of SVM Applications in Chemistry 371
  Recognition of Chemical Classes and Drug Design 371
  QSAR 376
  Genotoxicity of Chemical Compounds 378
  Chemometrics 379
  Sensors 381
  Chemical Engineering 383
  Text Mining for Scientific Information 384

SVM Resources on the Web 385
SVM Software 387
Conclusions 391
References 392

7. How Computational Chemistry Became Important in the Pharmaceutical Industry 401
Donald B. Boyd

Introduction 401
Germination: The 1960s 404
Gaining a Foothold: The 1970s 408
Growth: The 1980s 414
Gems Discovered: The 1990s 424
Final Observations 437
Acknowledgments 443
References 443

Author Index 453

Subject Index 471

Contributors

Jürgen Bajorath, Department of Life Science Informatics, B-IT International Center for Information Technology, Rheinische Friedrich-Wilhelms-Universität, Görresstrasse 13, D-53113 Bonn, Germany (Electronic mail: bajorath@bit.uni-bonn.de)

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

Benjamin A. Ellingson, Department of Chemistry and Supercomputing Institute, University of Minnesota, 207 Pleasant Street S.E., Minneapolis, MN 55455, U. S. A. (Electronic mail: elling@t1.chem.umn.edu)

Roland Faller, Department of Chemical Engineering and Materials Science, University of California-Davis, 1 Shields Avenue, Davis, CA 95616, U. S. A. (Electronic mail: rfaller@ucdavis.edu)

Antonio Fernandez-Ramos, Departamento de Quimica Fisica, Universidade de Santiago de Compostela, Facultade de Quimica, 15782 Santiago de Compostela, Spain (Electronic mail: qframos@gmail.com)

Bruce C. Garrett, Chemical and Materials Sciences Division, Pacific Northwest National Laboratory, MS K9-90, P.O. Box 999, Richland, WA 99352, U. S. A. (Electronic Mail: bruce.garrett@pnl.gov)

Jeffrey W. Godden, Department of Life Science Informatics, B-IT International Center for Information Technology, Rheinische Friedrich-Wilhelms-Universität, Görresstrasse 13, D-53113 Bonn, Germany (Electronic mail: godden@bit.uni-bonn.de)

Ovidiu Ivanciuc, Sealy Center for Structural Biology, Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, 301 University Boulevard, Galveston, TX 77555, U. S. A. (Electronic mail: oiivanci @utmb.edu)

Jörg Kussmann, Institut für Physikalische und Theoretische Chemie, Universität Tübingen, Auf der Morgenstelle 8, D-72076 Tübingen, Germany (Electronic mail: joerg.kussmann@uni-tuebingen.de)

Daniel S. Lambrecht, Institut für Physikalische und Theoretische Chemie, Universität Tübingen, Auf der Morgenstelle 8, D-72076 Tübingen, Germany (Electronic mail: lambrecht@uni-tuebingen.de)

Spiridoula Matsika, Department of Chemistry, Temple University, 1901 N. 13th Street, Philadelphia, PA 19122, U. S. A. (Electronic mail: smatsika @temple.edu)

Christian Ochsenfeld, Institut für Physikalische und Theoretische Chemie, Universität Tübingen, Auf der Morgenstelle 8, D-72076 Tübingen, Germany (Electronic mail: christian.ochsenfeld@uni-tuebingen.de)

Donald G. Truhlar, Department of Chemistry and Supercomputing Institute, University of Minnesota, 207 Pleasant Street S.E., Minneapolis, MN 55455, U. S. A. (Electronic mail: truhlar@umn.edu)

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.

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.

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.

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.

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.

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.

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.

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.