Reviews in Computational Chemistry, Volume 17. Edited by Kenny B. Lipkowitz, Donald B. Boyd

ISBNs: 0-471-39845-4 (Hardcover); 0-471-22441-3 (Electronic)

Reviews in Computational Chemistry Volume 17

Reviews in Computational Chemistry Volume 17

Edited by

Kenny B. Lipkowitz and Donald B. Boyd

WILEY-VCH

NEW YORK • CHICHESTER • WEINHEIM • BRISBANE • SINGAPORE • TORONTO

ISBN 0-471-22441-3

This title is also available in print as ISBN 0-471-39845-4.

For more information about Wiley products, visit our web site at www.Wiley.com.

Preface

The aphorism ''Knowledge is power'' applies to diverse circumstances. Anyone who has climbed an organizational ladder during a career understands this concept and knows how to exploit it. The problem for scientists, however, is that there may exist too much to know, overwhelming even the brightest intellectual. Indeed, it is a struggle for most scientists to assimilate even a tiny part of what is knowable. Scientists, especially those in industry, are under enormous pressure to know more sooner. The key to using knowledge to gain power is knowing what to know, which is often a question of what some might call, variously, innate leadership ability, intuition, or luck.

Attempts to manage specialized scientific information have given birth to the new discipline of informatics. The branch of informatics that deals primarily with genomic (sequence) data is bioinformatics, whereas cheminformatics deals with chemically oriented data. Informatics examines the way people work with computer-based information. Computers can access huge warehouses of information in the form of databases. Effective mining of these databases can, in principle, lead to knowledge.

In the area of chemical literature information, the largest databases are produced by the Chemical Abstracts Service (CAS) of the American Chemical Society (ACS). As detailed on their website (www.cas.org), their principal databases are the Chemical Abstracts database (CA) with 16 million document records (mainly abstracts of journal articles and other literature) and the REGISTRY database with more than 28 million substance records. In an earlier volume of this series,* we discussed CAS's SciFinder software for mining these databases. SciFinder is a tool for helping people formulate queries and view hits. SciFinder does not have all the power and precision of the command-line query system of CAS's STN, a software system developed earlier to access these and other CAS databases. But with SciFinder being easy

*D. B. Boyd and K. B. Lipkowitz, in *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., Wiley-VCH, New York, 2000, Vol. 15, pp. *v–xxxv*. Preface.

to use and with favorable academic pricing from CAS, now many institutions have purchased it.

This volume of *Reviews in Computational Chemistry* includes an appendix with a lengthy compilation of books on the various topics in computational chemistry. We undertook this task because as editors we were occasionally asked whether such a listing existed. No satisfactory list could be found, so we developed our own using SciFinder, supplemented with other resources.

We were anticipating not being able to retrieve every book we were looking for with SciFinder, but we were surprised at how many omissions were encountered. For example, when searching specifically for our own book series, *Reviews in Computational Chemistry*, several of the existing volumes were not "hit." Moreover, these were not consecutive omissions like Volumes 2–5, but rather they were missing sporadically. Clearly, something about the database is amiss.

Whereas experienced chemistry librarians and information specialists may fully appreciate the limitations of the CAS databases, a less experienced user may wonder: How punctilious are the data being mined by SciFinder? Certainly, for example, one could anticipate differences in spelling like Mueller versus Müller, so that typing in only Muller would lead one to not finding the former name. The developers of SciFinder foresaw this problem, and the software does give the user the option to look for names that are spelled similarly. Thus, there is some degree of "fuzzy logic" implemented in the search algorithms. However, when there are misses of information that should be in the database, the searches are either not fuzzy enough or there may be wrong or incomplete data in the CAS databases. Presumably, these errors were generated by the CAS staff during the process of data entry. In any event, there are errors, and we were curious how prevalent they are.

To probe this, we analyzed the hits from our SciFinder searches. Three kinds of errors were considered: (1) *wrong*, meaning there were factual errors in an entry which prevented the citation from being found by, say, an author search (although more exhaustive mining of the database did eventually uncover the entry); (2) *incomplete*, meaning that a hit could be obtained, but there were missing pieces of data, for example, the publisher, the city of publication, the year of publication, or the name of an author or editor; (3) *spelling*, meaning that there were spelling or typographical errors apparent in the entry, but the hit could nevertheless be found with SciFinder. In our study, about 95% of the books abstracted in the CA database were satisfactory; 1% had errors that could be ascribed to the data being wrong, 3% had incomplete data, and 1% had spelling errors. These error rates are lower limits. There almost certainly exist errors in spellings of authors' names or other errors that we did not detect. Concerning the *wrong* entries, most of them were recognized with the help of books on our bookshelves, but there are probably others we did not notice. Many errors, such as missing volumes of

a series, became evident when books from the same author or on the same topic were listed together.

If we noticed a variation of the spelling of an author's name from year to year or from edition to edition, especially when Russian and Eastern European names are involved, we classified these entries as being *wrong* if the infraction is serious enough to give a wrong outcome in a search. If one is looking for books by I. B. Golovanov and A. K. Piskunov, for example, one needs to search also for Golowanow and Piskunow, respectively. The user discovers that the spelling of their co-author changes from N. M. Sergeev to N. M. Sergejew! Should the user write Markovnikoff or Markovnikov? (Both spellings can be found in current undergraduate organic chemistry textbooks.) More of the literature is being generated by people who have non-English names. But even for very British names, such as R. McWeeney and R. McWeeny, there are misspellings in the CAS database. Perhaps one of the more frequent occurrences of misspellings and errors is bestowed on N. Yngve Öhrn. Some of the CAS spellings include: N. Yngve Oehrn, Yngve Ohrn, Ynave Ohrn, and even Yngve Oehru! There also may be errors concerning the publishing houses, some not very familiar to American readers. For example, aside from variability in their spellings, the Polish publisher Panstwowe Wydawnictwo Naukowe (PWN) is entered as PAN in one of the entries of W. Kolos' books, whereas the others are PWN.

Some of this analysis might be considered "nit-picking," but an error is certainly serious if it prevents a user from finding what is actually in the database. Our exercises with SciFinder suggest that it would be helpful if CAS strengthened their quality control and standardization processes. Cross-checking and cleaning up the spellings in their databases would allow users to retrieve desired data more reliably. It would also enhance the value of the CAS databases if missing data were added retrospectively.

So, what level of data integrity is acceptable? The total percentage of errors we found in our study was 5%. Is this satisfactory? Is this the best we can hope for? Hopefully not, especially as more people become dependent on databases and the rate of production of data becomes ever faster. Clearly, there is a need for a system that will better validate data being entered in the most used CAS databases. It is desirable that the quality of the databases increases at the same time as they are mushrooming in size.

A Tribute

Many prominent colleagues who have worked in computational chemistry have passed away since about the time this book series began. These include (in alphabetical order) Jan Almlöf, Russell J. Bacquet, Jeremy K. Burdett, Jean-Louis Calais, Michael J. S. Dewar, Russell S. Drago, Kenichi Fukui, Joseph Gerratt, Hans H. Jaffe, Wlodzimierz Kolos, Bowen Liu, Per-Olov Löwdin, Amatzya Y. Meyer, William E. Palke, Bernard Pullman, Robert

Rein, Carlo Silipo, Robert W. Taft, Antonio Vittoria, Kent R. Wilson, and Michael C. Zerner.* These scientists enriched the field of computational chemistry each in his own way. Three of these individuals (Almlöf, Wilson, Zerner) were authors of past chapters in *Reviews in Computational Chemistry*.

Dr. Michael C. Zerner died from cancer on February 2, 2000. Other tributes have already been paid to Mike, but we would like to add ours. Many readers of this series knew Mike personally or were aware of his research. Mike earned a B.S. degree from Carnegie Mellon University in 1961, an A.M. from Harvard University in 1962, and, under the guidance of Martin Gouterman, a Ph.D. in Chemistry from Harvard in 1966. Mike then served his country in the United States Army, rising to the rank of Captain. After postdoctoral work in Uppsala, Sweden, where he met his wife, he held faculty positions at the University of Guelph, Canada, and then at the University of Florida. At Gainesville he served as department chairman and was eventually named distinguished professor, a position held by only 16 other faculty members on the Florida campus.

Probably, Mike's research has most touched other scientists through his development of ZINDO, the semiempirical molecular orbital method and

*After this volume was in press, the field of computational chemistry lost at least four more highly esteemed contributors: G. N. Ramachandran, Gilda H. Loew, Peter A. Kollman, and Donald E. Williams. We along with many others grieve their demise, but remember their contributions with great admiration. Professor Ramachandran lent his name to the plots for displaying conformational angles in peptides and proteins. Dr. Loew founded the Molecular Research Institute in California and applied computational chemistry to drugs, proteins, and other molecules. She along with Dr. Joyce J. Kaufman were influential figures in the branch of computational chemistry called by its practitioners "quantum pharmacology" during the 1960s and 1970s. Professor Kollman, like many in our field, began his career as a quantum chemist and then expanded his interests to include other ways of modeling molecules. Peter's work in molecular dynamics and his AMBER program are well known and helped shape the field as it exists today. Professor Williams, an author of a chapter in Volume 2 of *Reviews in Computational Chemistry*, was famed for his contributions to the computation of atomic charges and intermolecular forces. Drs. Ramachandran, Loew, and Williams were blessed with long careers, whereas Peter's was cut short much too early.

Although several of Peter's students and collaborators have written chapters for *Reviews in Computational Chemistry*, Peter's association with the book series was a review he wrote about Volume 13. As a tribute to Peter, we would like to quote a few words from this book review, which appeared in *J. Med. Chem.*, **43** (*11*), 2290 (2000). While always objective in his evaluation, Peter was also generous in praise of the individual chapters ("a beautiful piece of pedagogy," "timely and interesting," "valuable," and "an enjoyable read"). He had these additional comments which we shall treasure:

> This volume of *Reviews in Computational Chemistry* is of the same very high standard as previous volumes. The editors have played a key role in carving out the discipline of computational chemistry, having organized a seminal symposium in 1983 and having served as the chairmen of the first Gordon Conference on Computational Chemistry in 1986. Thus, they have a broad perspective on the field, and the articles in this and previous volumes reflect this.

We would like to add that Peter was an invited speaker at the Symposium on Molecular Mechanics (held in Indianapolis in 1983) and was co-chairman of the second Gordon Research Conference on Computational Chemistry in 1988. As we pointed out in the Preface of Volume 13 (p. *xiii*) of this book series, no one had been cited more frequently in *Reviews of Computational Chemistry* than Peter. Peter—and the others—will be missed.

program for calculating the electronic structure of molecules. To relieve the burden of providing user support, Mike let a software company commercialize it, and it is currently distributed by Accelrys (née Molecular Simulations, Inc.) In addition, a version of the ZINDO method has been written separately by scientists at Hypercube in their modeling software HyperChem. Likewise, ZINDO calculations can be done with the CAChe (Computer-Aided Chemistry) software distributed by Fujitsu. Several thousand academic, government, and industrial laboratories have used ZINDO in one form or another. ZINDO is even distributed by several publishing companies to accompany their textbooks, including introductory texts in chemistry.

Mike published over 225 research articles in well-respected journals and 20 book chapters, one of which was in the second volume of *Reviews in Computational Chemistry*. It still remains a highly cited chapter in our series. In addition, Mike edited 35 books or proceedings, many of which were associated with the very successful Sanibel Symposia that he helped organize with his colleagues at Florida's Quantum Theory Project (QTP). If you have never organized a conference or edited a book, it may be hard to realize how much work is involved. Not only was Mike doing basic research, teaching (including at workshops worldwide), and serving on numerous university governance and service committees, he was also consulting for Eastman Kodak, Union Carbide, and others. A little known fact is that Mike is a co-inventor of eight patents related to polymers and polymer coatings.

Mike's interests and abilities earned him invitations to many meetings. He attended four Gordon Research Conferences (GRCs) on Computational Chemistry (1988, 1990, 1994, and 1998).* Showing the value of cross-fertilization, Mike subsequently brought some of the topics and ideas of these GRCs to the Sanibel Symposia. Mike also longed to serve as chair of the GRC. The GRCs are organized so that the job of chair alternates between someone from academia and someone from industry. The participants at each biennial conference elect someone to be vice-chair at the next conference (two years later), and then that person moves up to become chair four years after the election. Mike was a candidate in 1988 and 1998, which were years when nonindustrial participants could run for election. He and Dr. Bernard Brooks (National Institutes of Health) were elected co-vice-chairs in 1998. Sadly, Mike died before he was able to fulfill his dream. At the GRC in July 2000,† tributes were paid to Mike by Dr. Terry R. Stouch (Bristol-Myers Squibb), Chairman, and by Dr. Brooks. In addition, Dr. John McKelvey, Mike's collaborator during the Eastman Kodak consulting days, beautifully recounted Mike's many fine accomplishments.

Our science of computational chemistry owes much to the contributions of our departed friends and colleagues.

---

*D. B. Boyd and K. B. Lipkowitz, in *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., Wiley-VCH, New York, 2000, Vol. 14, pp. 399–439. History of the Gordon Research Conferences on Computational Chemistry.

†See http://chem.iupui.edu/rcc/grccc.html.

This Volume

As with our earlier volumes, we ask our authors to write chapters that can serve as tutorials on topics of computational chemistry. In this volume, we have four chapters covering a range of issues from molecular docking to spin–orbit coupling to cellular automata modeling.

This volume begins with two chapters on docking, that is, the interaction and intimate physical association of two molecules. This topic is highly germane to computer-aided ligand design. Chapter 1, written by Drs. Ingo Muegge and Matthias Rarey, describes small molecule docking (to proteins primarily). The authors put the docking problem into perspective and provide a brief survey of docking methods, organized by the type of algorithms used. The authors describe the advantages and disadvantages of the methods. Rigid docking including geometric hashing and pose clustering is described. To model nature more closely, one really needs to account for flexibility of both host and guest during docking. The authors delineate the various categories of treating flexible ligands and explain how each works. Then an evaluation of how to handle protein flexibility is given. Docking of molecules from combinatorial libraries is described next, and the value of consensus scoring in identifying potentially interesting bioactive compounds from large sets of molecules is pointed out. Of particular note in Chapter 1 are explanations of the multitude of scoring functions used in this realm of computational chemistry: shape and chemical complementary scoring, force field scoring, empirical and knowledge-based scoring, and so on. The need for reliable scoring functions underlies the role that docking can play in the discovery of ligands for pharmaceutical development.

The first chapter sets the stage for Chapter 2 which covers protein–protein docking. Drs. Lutz P. Ehrlich and Rebecca C. Wade present a tutorial on how to predict the structure of a protein–protein complex. This topic is important because as we enter the era of proteomics (the study of the function and structure of gene products) there is increasing need to understand and predict "communication" between proteins and other biopolymers. It is made clear at the outset of Chapter 2 that the multitude of approaches used for small molecule docking are usually inapplicable for large molecule docking; the generation of putative binding conformations is more complex and will most likely require new algorithms to be applied to these problems. In this review, the authors describe rigid-body and flexible docking (with an emphasis on methods for the latter). Geometric hashing techniques, conformational search methodologies, and gradient approaches are explained and put into context. The influence of side chain flexibility, backbone conformational changes, and other issues related to protein binding are described. Contrasts and comparisons between the various computational methods are made, and limitations of their applicability to problems in protein science are given.

Chapter 3, by Dr. Christel Marian, addresses the important issue of spin–orbit coupling. This is a quantum mechanical relativistic effect, whose impact on molecular properties increases with increasing nuclear charge in a way such that the electronic structure of molecules containing heavy elements cannot be described correctly if spin–orbit coupling is not taken into account. Dr. Marian provides a history and the quantum mechanical implications of the Stern–Gerlach experiment and Zeeman spectroscopy. This review is followed by a rigorous tutorial on angular momenta, spin–orbit Hamiltonians, and transformations based on symmetry. Tips and tricks that can be used by computational chemists are given along with words of caution for the nonexpert. Computational aspects of various approaches being used to compute spin–orbit effects are presented, followed by a section on comparisons of predicted and experimental fine-structure splittings. Dr. Marian ends her chapter with descriptions of spin-forbidden transitions, the most striking phenomenon in which spin–orbit coupling manifests itself.

Chapter 4 moves beyond studying single molecules by describing how one can predict and explain experimental observations such as physical and chemical properties, phase transitions, and the like where the properties are averaged outcomes resulting from the behaviors of a large number of interacting particles. Professors Lemont B. Kier, Chao-Kun Cheng, and Paul G. Seybold provide a tutorial on cellular automata with a focus on aqueous solution systems. This computational technique allows one to explore the less-detailed and broader aspects of molecular systems, such as variations in species populations with time and the statistical and kinetic details of the phenomenon being observed. The methodology can treat chemical phenomena at a level somewhere between the intense scrutiny of a single molecule and the averaged treatment of a bulk sample containing an infinite population. The authors provide a background on the development and use of cellular automata, their general structure, the governing rules, and the types of data usually collected from such simulations. Aqueous solution systems are introduced, and studies of water and solution phenomena are described. Included here are the hydrophobic effect, solute dissolution, aqueous diffusion, immiscible liquids and partitioning, micelle formation, membrane permeability, acid dissociation, and percolation effects. The authors explain how cellular automata are used for systems of first- and second-order kinetics, kinetic and thermodynamic reaction control, excited state kinetics, enzyme reactions, and chromatographic separation. Limitations of the cellular automata models are made clear throughout. This kind of coarse-grained modeling complements the ideas considered in the other chapters in this volume and presents the basic concepts needed to carry out such simulations.

Lastly, we provide an appendix of books published in the field of computational chemistry. The number is large, more than 1600. Rather than simply presenting all these books in one long list sorted by author or by date, we have partitioned them into categories. These categories range from broad

topics like quantum mechanics to narrow ones like graph theory. The categories should aid finding books in specific areas. But it is worth remembering that all the books tabulated in the appendix, whether on molecular modeling, chemometrics, simulations, and so on, represent facets of computational chemistry. As defined in the first volume of our series,* computational chemistry consists of those aspects of chemical research that are expedited or rendered practical by computers. Analysis of the number of computational chemistry books published each year revealed an interesting phenomenon. The numbers have been increasing and occurring in waves four to five years apart.

As always, we try to be heedful of the needs of our readers and authors. Every effort is made to produce volumes that will have sustained usefulness in learning, teaching, and research. We appreciate the fact that the community of computational chemists has found that these volumes fulfill a need. In the most recent data on impact factors from the Institute of Scientific Information (Philadelphia, Pennsylvania), *Reviews in Computational Chemistry* is ranked fourth among serials (journals and books) in the field of computational chemistry. (In first place is the *Journal of Molecular Graphics and Modelling*, followed by the *Journal of Computational Chemistry* and *Theoretical Chemistry Accounts*. In fifth and sixth places are the *Journal of Computer-Aided Molecular Design* and the *Journal of Chemical Information and Computer Science*, respectively.)

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 thank the authors in this volume for their excellent chapters. Mrs. Joanne Hequembourg Boyd provided valued editorial assistance.

Kenny B. Lipkowitz and Donald B. Boyd
Indianapolis
February 2001

*K. B. Lipkowitz and D. B. Boyd, Eds., *Reviews in Computational Chemistry*, VCH Publishers, New York, 1990, Vol. 1, pp. *vii–xii*. Preface.

Contents

Contributors

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

Chao-Kun Cheng, Department of Mathematics, Virginia Commonwealth University, Richmond, Virginia 23298, U.S.A. (Electronic mail: ccheng@atlas.vcu.edu)

Lutz P. Ehrlich, LION Bioscience AG, Waldhofer Strasse 98, D-69123 Heidelberg, Germany (Electronic mail: lutz.ehrlich@lionbioscience.com)

Lemont B. Kier, Department of Medicinal Chemistry, Virginia Commonwealth University, Richmond, 23298, U.S.A. (Electronic mail: kier@hsc.vcu.edu)

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

Christel M. Marian, German National Research Center for Information Technology (GMD), Scientific Computing and Algorithms Institute (SCAI), Schloss Birlinghoven, D-53754 Sankt Augustin, Germany (Electronic mail: christel.marian@gmd.de and cm@uni-bonn.de)

Ingo Mügge, Bayer Research Center, 400 Morgan Lane, West Haven, Connecticut 06516, U.S.A. (Electronic mail: ingo.mugge.b@bayer.com)

Matthias Rarey, German National Research Center for Information Technology (GMD), Institute for Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven, D-53754 Sankt Augustin, Germany (Electronic mail: rarey@gmd.de)

Paul Seybold, Chemistry Department, Wright State University, Dayton, Ohio 45435, U.S.A. (Electronic mail: paul.seybold@wright.edu)

Rebecca C. Wade, European Media Laboratory, Villa Bosch, Schloss-Wolfsbrunnenweg 33, D-69118 Heidelberg, Germany (Electronic mail: rebecca.wade@eml.villa-bosch.de)

Contributors to Previous Volumes*

Volume 1

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.

*When no author of a chapter can be reached at the addresses shown in the original volume, the current affiliation of the senior or corresponding author is given here as a convenience to our readers.

†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 Indianapolis, Indianapolis, Indiana 46202 (Electronic mail: dykstra@chem.iupui.edu).

¶Current address: University of Washington, Seattle, Washington 98195 (Electronic mail: lybrand@proteus.bioeng.washington.edu).

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

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, United Kingdom (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

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

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

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

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: DuPont Pharmaceuticals Research Laboratories, 150 California Street, Suite 1100, San Francisco, California 94111-4500 (Electronic mail: jblaney@combichem.com).

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

Volume 7

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

Zdeněk 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: Department of 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: fromzdenek@hotmail.com).

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

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

Volume 9

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

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.

---

*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).

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

Volume 11

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

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., c/o Aventis Pharma Ltd., Rainham Road South, Dagenham, Essex, RM10 7XS, United Kingdom (Electronic mail: david.clark@argentadiscovery.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

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

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

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

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

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

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).

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.

Reviews in Computational Chemistry Volume 17

Author Index

Subject Index

Computer programs are denoted in boldface; databases and journals are in italics.

Reviews in Computational Chemistry, Volume 17. Edited by Kenny B. Lipkowitz, Donald B. Boyd

ISBNs: 0-471-39845-4 (Hardcover); 0-471-22441-3 (Electronic)

CHAPTER 1

Small Molecule Docking and Scoring

Ingo Muegge* and Matthias Rarey†

**Bayer Research Center, 400 Morgan Lane, West Haven, Connecticut 06516, and †German National Research Center for Information Technology (GMD), Institute for Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven, D-53754 Sankt Augustin, Germany*

INTRODUCTION

Molecular recognition is a central phenomenon in biochemistry. The highly specific recognition of, for example, enzymes and their substrates, protein receptors and their signal inducing ligands, or antigens and their antibodies in biological systems, is crucial to make complex life forms work. A detailed understanding of molecular recognition mechanisms is of particular interest in drug discovery, because most drugs interact with protein targets such as enzymes or receptors. To understand in detail the energetics of a protein interacting with a ligand that may be a potential drug candidate, one has to know the structure of the protein–ligand complex at atomic resolution. Protein–ligand structures are usually solved by X-ray crystallography[1–4] or by nuclear magnetic resonance (NMR) spectroscopy.[5–8] The Protein Data Bank (PDB) presently contains about 15,000 protein structures[9,10] and grows at a pace of about seven additional structures per day.[11,12] For a particular protein structure that is crystallized for a drug discovery program, fast screening methods have been developed that can process up to several thousand compounds soaked into the crystal.[13] NMR screening methods have also been employed

that connect structural measurements (chemical shifts of amino acid residues in the binding site) to binding affinity evaluation.[14] However, these methods have obvious limitations; they depend on protein samples and crystals. Also, only a limited number of small molecules can be analyzed by physical chemistry experiments due to cost and time constraints.

The rapid advancement in X-ray crystallography and NMR spectroscopy provides a large number of solved protein structures. However, computational approaches are promising alternatives to crystallographic and NMR screening techniques. Computational methods that predict the three-dimensional (3D) structure of a protein–ligand complex are sometimes referred to as molecular docking approaches.[15] Protein structures can be employed to dock ligands into the binding site of the protein and to study the intermolecular interactions. The prediction of ligand-binding modes can help in guiding, for instance, medicinal chemists exploring structure-activity relationships (SAR) in the lead optimization phase of a drug discovery effort. (A lead is a compound that shows biological activity and has the potential of being structurally modified for improved bioactivity.) Docking is applied here as a tool for ligand design complementing other structure-based methods such as de novo design.[16] In the absence of experimental structures, homology models[17,18] or pseudo-receptor models[19] are sometimes used. Note, however, that the quality of a protein structure—whether experimental or computational—is crucial, because small changes in protein structure can influence the outcome of docking experiments dramatically.[20] Ideally, the atomic resolution of crystal structures should be below 2.5 Å.[21]

A challenge for molecular docking as a ligand design tool lies in the identification of the correct binding geometry of the ligand in the binding site (binding mode). In some cases, finding the correct binding mode is complicated by the observation that similar ligands unexpectedly bind in quite different orientations in the receptor site. Examples include the inhibitor MJ33 in phospholipase A2[22] and BANA113 in influenza virus neuraminidase[23] discussed in the section on Applications. To find the correct binding mode of a ligand in the receptor site, an adequate sampling of conformational space available to a flexible ligand molecule in the protein binding pocket is required. The high flexibility of a typical ligand requires effective sampling methods. It also somewhat separates protein–ligand docking from the related protein–protein docking problem to be reviewed by Ehrlich and Wade[24] in this book. In contrast to a globular protein, a small molecule is often not constrained in its overall shape. Also, shape alone is not a sufficient descriptor to identify low-energy conformations of a protein–ligand complex.[25]

Several docking algorithms (e.g., DOCK,[26–28] FlexX,[29–31] Hammerhead,[32] GOLD,[21] AutoDOCK,[33] FLOG[34]) have been developed that place rigid or flexible ligands in mostly rigid—but recently also flexible—protein binding sites. All these methods/programs consist of two more or less intertwined parts: the sampling of the configuration space and the scoring of protein–ligand

complexes. For sampling purposes, different algorithms have been applied ranging from numerical simulation to combinatorial optimization. These algorithms will be discussed in the second section (Algorithms for Molecular Docking). The aim of scoring is to identify the correct binding mode by its lowest energy, if the configuration with the lowest energy is assumed to be the "correct" (observable) one. Various functions have been devised to measure the protein–ligand binding affinity in the docking algorithms. Many of the functions are not strictly related to binding free energies. Therefore, functions, designed to rank different protein–ligand complexes according to their binding affinities, are usually referred to as scoring functions.[35–38] Scoring functions will be discussed in the third section (Scoring).

Large databases (ca. 10^6 compounds) of small, drug-like molecules can be assembled today by purchase from a supplier, by high-speed analogue synthesis, or by combinatorial synthesis.[39,40] High throughput methods have been developed that are able to screen these compounds in a few days for bioactivity against a protein target relevant to drug discovery.[41] Potentially, computational docking methods can also be used as a "virtual" (i.e., totally computational or in silico) screening method.[42,43] The major advantage of docking as a screening tool is that it does not deplete assay, compound, and laboratory resources. It is cost effective and has the advantage of screening libraries of virtual (i.e., hypothetical or not yet synthesized) compounds. To compete with biological screening methods, however, computational tools must be automated, fast, and reliable in ranking putative protein–ligand complexes according to their binding affinities. To be competitive in speed, ca. 10,000 protein–ligand complexes must be evaluated per day per processor.[44] That is, algorithms have to be developed that can find the correct binding mode of a compound by effectively sampling its available conformational and configurational (i.e., orientational) space in the binding pocket within 10 seconds of central processing unit (CPU) time. This time includes the evaluation of functions that estimate the energy of every docking arrangement. Currently, the best performing docking algorithms take about 1–3 min of CPU time for a ligand–protein docking experiment.

The use of docking as a virtual screening tool is more challenging than using it as a ligand design tool. If many structurally diverse compounds are docked, they need to be ranked according to their predicted binding affinity to the protein. In practice, it is rather unlikely to find a strongly binding ligand in a screening database of compounds. Hence, the docking–scoring approach has to be able to identify weak binders in a pool of nonbinders. Docking–scoring approaches used today tend to have a large number of false positives (usually between 96 and 100% of the computational hits)—compounds with high scores but with no experimentally observable binding to the protein. Although scoring functions have been worked on for two decades now, there is only incremental progress. No single scoring function facilitates a reliable ranking of protein–ligand complexes today. Therefore, the currently preferred

scheme in scoring applications involves using many scoring functions and then eliminating false positives by consensus scoring (i.e., making decisions based on what a combination of scoring functions predicted). Encouraging work on enzyme targets has been presented recently showing that consensus scoring is able to reduce the rate of false positives in a virtual screen significantly.[37] For p38 mitogen-activated protein (MAP) kinase, inosine monophosphate dehydrogenase, and human immunodeficiency virus (HIV) protease, it has been found that 11–29% of the compounds picked by a consensus scoring scheme are biologically active compared to only 3–9% of the compounds picked by a single scoring function. The success of a virtual screening experiment is determined by finding at least one novel compound with at least low micromolar (μM) biological activity. Since typically only ca. 100 compounds from an in silico screen are tested experimentally, there is still a high chance not to find a single hit. Therefore, the reported hit increase from consensus scoring represents a significant improvement in the likely success rate of a virtual screen. To be competitive with biological high throughput screening methods, however, it is equally important to keep the number of false negatives (biologically active compounds with low scores) below 10%.

Docking methods can also be used similar to de novo design methods in docking small molecular fragments to the binding site.[45] The placements of these fragments can be used for the design of combinatorial libraries. For instance, combining the screening capabilities with a fragment approach, Rose and co-workers[46] recently proposed docking as a virtual "SAR by NMR" tool. Following the ideas of Fesik and co-workers[14] who proposed an NMR screening technique of several small fragments to be optimized separately and later linked together, this "virtual NMR screening" technique docks small fragments in different binding pockets and finds the optimal linker in a combinatorial fashion.

Since scoring of protein–ligand complexes is such a central issue in docking, this review discusses new developments in scoring functions in some detail. Docking techniques and their applications in ligand design, virtual screening, and library design are reviewed.

ALGORITHMS FOR MOLECULAR DOCKING

In 1972, Platzer et al.[47,48] performed energy calculations on enzyme–ligand complexes using molecular dynamics simulations. About 10 years later, Kuntz et al.[26] published the DOCK algorithm, which was the first approach to tackle the molecular docking problem with a combinatorial approach instead of a simulation. Since then, interest in fast molecular docking algorithms has steadily grown, and a variety of algorithms has been developed. The growth arose from two principal causes. First, more structures of pharmaceutical

targets became available, and second, computers became fast enough that the docking approach for large data sets seemed feasible.

In this section, we will give a brief survey of docking methods organized by the type of algorithms used. The main focus is on explaining the different algorithms and discussing their advantages and drawbacks rather than on evaluating them against each other. Other reviews of algorithms used in structure-based ligand design have been published.[15,27,49–54]

The Docking Problem

The protein–ligand docking problem is a geometric search problem. The degrees of freedom to consider are the relative orientation of the two molecules as well as their conformations. For the protein, we often already know the overall 3D structure, but this is mostly not true for the ligand. In addition, our focus is on the ligand-binding site, which is in nearly all cases a concave region of the protein surface like a cleft or a cavity. It is common to assume the protein is a rigid object although this is not true in general and the degree of structural changes in the receptor site depend on the protein itself. Examples of how to handle protein flexibility during docking calculations are given below.

The output from a typical docking algorithm includes a list of protein–ligand complexes rank-ordered by a given scoring function. In an ideal situation, the highest ranking complex would resemble the binding mode that would be observed experimentally, assuming the experiments are performed. Currently, this cannot be reliably achieved for two reasons. First, we do not have a scoring function that always has its global optimum in agreement with the experiment, and second, we do not have fast optimization algorithms for finding the global optimum for a given scoring function. Nevertheless, methods available today are still quite useful in practical applications like virtual screening. The major goal in a virtual screening run is to select a small number of molecules from a large pool of compounds that will then be further analyzed by experimental methods. In spite of having a high error rate, virtual screening is able to select biologically active molecules with a significantly higher rate than a random selection does or a selection does that optimizes the chemical diversity of the compounds picked. Such improvement in the rate of finding hits in a database is something referred to as enrichment.

Since we are interested in searching large sets of compounds for putative new lead structures, the speed of a docking algorithm becomes a critical issue. The only way to gain speed in an optimization process without losing quality of the results, is to guide the search by problem-specific information. In this sense, a time-efficient docking method does not consist of a search engine and a scoring function as separate parts. The information about scoring is rather an integral part of the search engine. Several examples of this design principle are shown in the following sections.

Placing Fragments and Rigid Molecules

The docking problem can be simplified by neglecting the conformational degrees of freedom of the ligand molecule. Although this simplification is not appropriate for the general protein–ligand docking scenario, algorithms based on this approximation are of great importance and can be applied to docking of small or rigid molecules, molecule fragments, or conformational ensembles of molecules.

Clique-Search Based Approaches

The docking of two rigid molecules can be understood as a problem of matching characteristic features of the molecules in space.[55] A match is an assignment of a ligand feature to a protein feature. Such a feature can be either a volume segment of the protein or the ligand or a complementary interaction such as a hydrogen-bond donor and acceptor. The search procedure maximizes the number of matches under the constraint that they are compatible in 3D space, (i.e., that they can be realized simultaneously). Compatibility means that we can find a transformation that simultaneously superimposes all ligand features onto the matched protein features.

To search for compatible matches, a distance compatibility graph is used. The nodes of the graph comprise all possible matches between the protein and the ligand; the edges connect pairs of nodes that are compatible. Compatibility means mostly distance compatibility within a fixed tolerance ε (i.e., the difference in distance Δd between the ligand and the protein features differs by $\Delta d \leq \varepsilon$). A necessary condition for a set of matches to be simultaneously formed is that all pairs of matches are distance compatible. Looking at the distance compatibility graph, a set of matched features is represented by a set of nodes. The distance compatibility between two matched features is represented by an edge. Therefore, a set of matched features is distance compatible exactly, if all pairs of corresponding nodes are connected by an edge. Such a fully connected subgraph is called a clique in graph theory. Searching for maximal cliques in graphs is a well-known problem, and although this is a hard optimization problem in theory, fast algorithms exist for practical applications.[56] We obtain an initial orientation of the ligand molecule in the receptor site of the protein by superimposing the matched features of a clique (Figure 1).

The algorithm for rigid-body docking in the DOCK program suite[26] is based on the idea of searching for distance-compatible matches. Starting with the molecular surface of the protein,[57–59] a set of spheres is created inside the receptor site. The spheres represent the volume that could be occupied by a ligand molecule (Figure 2). Figure 3 shows an example of the HIV inhibitor VX-478 bound in the active site of the enzyme.[60] The ligand is represented either by spheres inside the ligand or directly by its atoms (Figure 2). In the early versions of DOCK, an enumeration algorithm searched for sets with

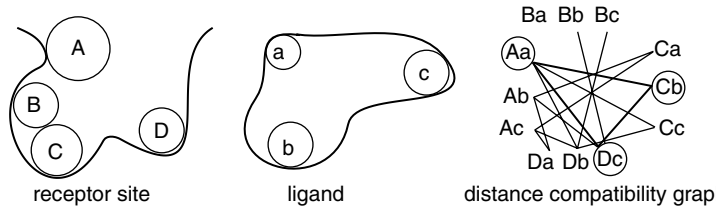

Figure 1 On the left-hand side, a protein receptor site and a ligand are schematically drawn. Some features are highlighted marked by letters. The corresponding distance compatibility graph is shown on the right. Each distance compatible pair of matched features is connected by an edge. The three encircled nodes are an example for a clique.

up to four distance-compatible matches. Each set was used for an initial fit of the ligand into the receptor site. The set was augmented using further compatible matches. The position of the ligand was then optimized and scored.

Since its first introduction in 1982, the DOCK software has been extended in several directions. The matching spheres can be labeled with chemical properties,[61] and distance bins are used to speed up the search process.[62,63] Recently, the search algorithm for distance-compatible matches was changed[28] to the clique-detection algorithm introduced by Crippen and co-workers.[55] Furthermore, several scoring functions are now applied in combination with the DOCK algorithm.[64–68]

Several other approaches such as LUDI,[69,70] CLIX,[71] and ADAM[72] are based on this clique-search paradigm although somehow differing in the features used for matching and in the way they are represented.

Geometric Hashing

Geometric hashing[73] originated from the area of computer vision and was first applied to molecular docking problems by Fischer et al.[74,75] In computer vision, the geometric hashing scheme was developed for the problem of recognizing (partially occluded) objects in camera scenes. In principle,

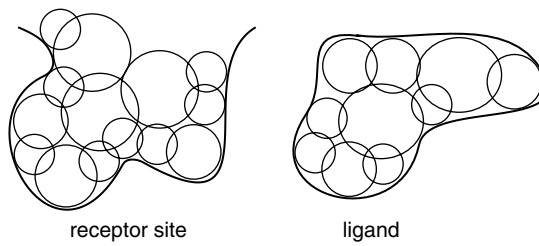

Figure 2 In DOCK, the receptor is geometrically described using overlapping spheres of different sizes that are complementary to the molecular surface of the ligand-binding site. The ligand is described with spheres inside its surface. During the docking process protein and ligand spheres are matched.

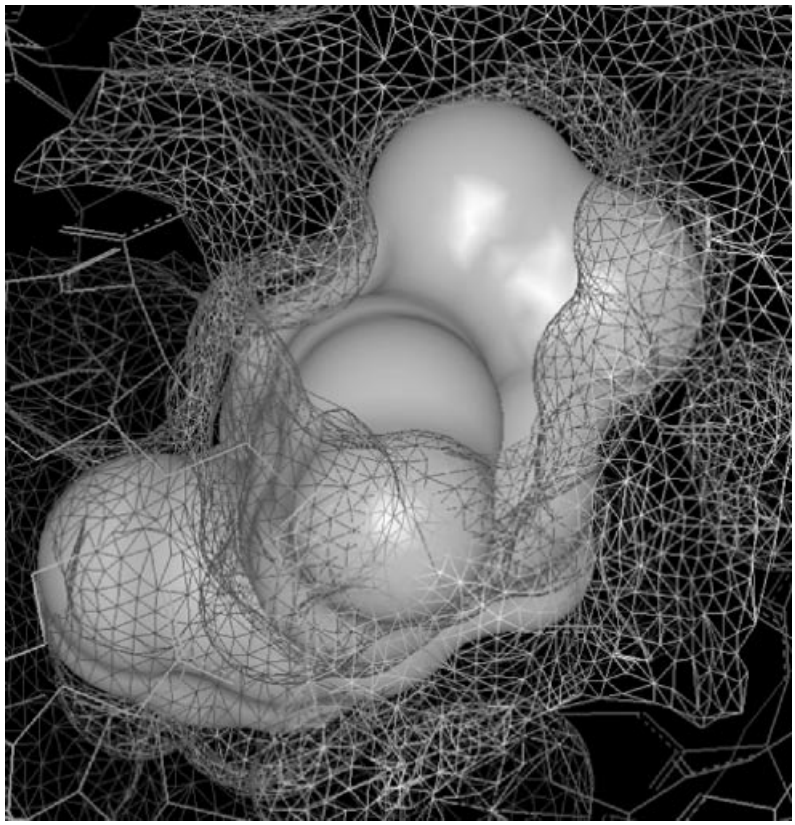

Figure 3 Crystal structure (PDB entry 1hpv) of the HIV-1 protease inhibitor VX-478 (solid surface) bound in the active site of the enzyme (line representation). (Ref. 60.)

geometric hashing can be regarded as an alternative to clique searching for the matching of features.

Hashing is a frequently used computer science technique allowing fast access to data. The basic idea is to create a key for a data entry that can be used as a memory address for the data entry. Since there are typically more addresses available than computer memory, a so-called hashing function is used to map the addresses of the data entry to a smaller address space. In geometric hashing, geometric features of objects like distances are used to create the hashing key. Therefore, objects having certain geometric features can be easily accessed via the geometric hash table.

The geometric hashing algorithm consists of two phases: a preprocessing phase and a recognition phase. In the preprocessing phase, the geometric hash table is constructed from a single ligand or a set of ligands to be docked. A hash entry contains the ligand name and a so-called reference frame allowing

for orienting the ligand in space. The entry is stored in the table several times. Each entry is addressed by the distance of a feature to the corners of the reference frame. In the recognition phase, the protein features are used to vote for hash entries. A vote means that there is a protein feature that will match a ligand feature if the ligand is oriented as defined in the hash entry. A hash entry with a large number of votes represents a ligand and an orientation with several matched features. Such hash entries with large vote counts represent initial orientations and are further analyzed.

Two aspects make geometric hashing attractive for molecular docking problems: it is time-efficient and deals with partial matching or partially occluded objects in terms of pattern recognition. The latter is extremely important because in most docking applications not all the ligand features are matched with those of the protein because parts of the ligand surface are in contact with bulk water.

To apply geometric hashing to molecular docking, Fischer et al.[74,75] used the sphere representation of DOCK as the underlying model. Because docking is performed in 3D space, in principle three points (here spheres or atoms) are necessary to define a reference frame. Consequently, the number of hash table entries unacceptably increases with the fourth power of the number of ligand atoms. Therefore, a reference frame is described by only two points after omitting one unfixed degree of freedom (rotation around the axis defined by the two points). With this model in mind, the geometric hashing approach can be directly applied to the molecular docking problem.

Pose Clustering

Akin to geometric hashing, pose clustering[76] is an approach for solving the molecular docking problem that comes from pattern recognition.[77] The algorithm was originally developed for detecting objects in two-dimensional (2D) scenes with an unknown camera location. As in the case of geometric hashing, pose clustering can be regarded as an alternate algorithm for matching of features. The algorithm matches each triplet of features of the first object with each triplet of the second object. From a match, a location for the first object with respect to the second can be computed by superimposing the triangles. The calculated locations are stored and clustered. If a large cluster is found, a location with a high number of matching features is detected.

For applying pose clustering to molecular docking, the LUDI model of molecular interactions[69,70] can be used as the underlying representation. For each interacting group, an interaction center and an interaction surface are defined as shown in Figure 4. An interaction between two groups A and B occurs if the interaction center of A lies approximately on the interaction surface of B and vice versa. Discrete points forming the features of the protein approximate the interaction surfaces of the protein. The ligand features are the interaction centers of the ligand molecule.

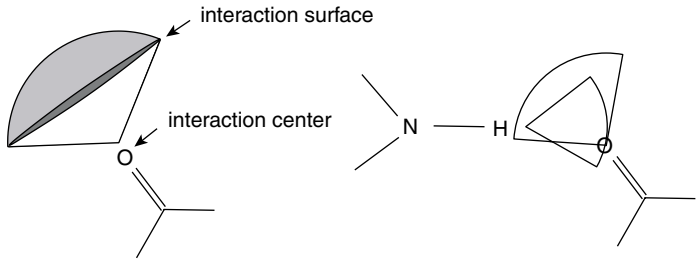

Figure 4 The interaction model taken from LUDI consists of an interaction center and a spherical interaction surface describing distance and angle constraints of the interaction. An interaction occurs if the interaction center of the first group is located close to the interaction surface of the second group and vice versa.

In the docking application, the matches are limited in two ways. First, the interaction types must be compatible, for example, a hydrogen-bond donor interaction center can only be matched with a hydrogen-bond acceptor interaction surface. Second, the triangle edges must be approximately of the same length. A hashing scheme is necessary to efficiently access and match surface triangles onto a triangle query of a ligand interaction center. The hashing scheme stores edges between two points that are addressed by the two interaction types and the edge length. A list-merging algorithm creates all triangles based on lists of fitting triangle edges for two of the three query triangle edges. For each match created by this procedure, the interaction centers of the ligand can be placed on the interaction surfaces of the protein. However, in order to form an interaction, the interaction center of the protein must also lie close to the interaction surface of the ligand. This additional directionality constraint has to be checked for the three interactions. Finally, a transformation is calculated that superimposes the two triangles.

For clustering the transformations, a complete-linkage hierarchical clustering is used.[78] A hierarchical clustering algorithm iteratively merges the two clusters that are closest to each other. In complete-linkage clustering, the distance between two clusters is defined to be the maximum distance between the cluster elements. In this application, the distance between two transformations is the root-mean-square distance (rmsd) between ligand atoms after applying the transformations. Using complete-linkage clustering ensures that no two transformations are put into the same cluster if they have an rmsd greater than a given threshold. For each of the found clusters, post–processing steps such as searching for additional interactions, checking for protein–ligand overlap, and scoring are performed.

Flexible Ligand Docking

Most small molecules in pharmaceutical research have at least a few rotatable bonds or even flexible ring systems. Seventy percent of drug-like

molecules possess between two and eight rotatable bonds.[79] Since the energetic differences between alternative conformations are often low compared to the binding affinity, handling of conformational degrees of freedom is of great importance in molecular docking calculations. The basic categories of approaches for handling flexible ligands, which are summarized here, are ensembles, fragmentation, genetic algorithms, and simulation.

Conformation Ensembles

In principle, every conformer of a set of flexible ligands could be stored in a database, and then each conformation could be evaluated with rigid-body docking algorithms. The size of the ensemble is critical since the computing time increases linearly with the number of conformations and the quality of the result drops with larger differences between the most similar conformation of the ensemble and the actual complex conformation. Thus a balance must be struck between computing time requirements and the desire to cover all of conformational space.

Kearsley and co-workers[80] developed the Flexibase/FLOG docking algorithm based on conformation ensembles. Flexibases store a small set of diverse conformations for each molecule from a given database. The conformations are created with distance geometry methods[81–83] and a rough energy minimization. A set of up to 25 conformations per molecule is selected by rmsd dissimilarity criteria. Each conformation is then docked using the rigid-body docking program FLOG,[34] which is similar to the DOCK algorithm discussed above.

Lorber and Shoichet[84] presented a different approach. They created a database with on average a set of ca. 300 conformations per molecule. For each molecule, a rigid part was defined (e.g., an aromatic ring system). The conformation ensemble was created such that the atoms of this rigid part were superimposed. Then, the DOCK algorithm was applied to the rigid part and all conformations were subsequently tested for overlap and scored. With this method, a significant speedup could be achieved compared to an independent docking of the conformations.

Fragmentation

The most popular approach for handling ligand flexibility is fragmentation. Here, the ligand is somehow divided into smaller pieces, called fragments, which can be treated as conformationally rigid or by a small conformational ensemble. In principle, there are two strategies for handling the fragments. We can either start by placing one fragment in the receptor site and then add the remaining fragments to the orientations we find for the first one or we can place all (or a subset of) fragments independently and try to reconnect the fragments in favorable orientations until they constitute a complete ligand. We call the first strategy ''incremental construction,'' the second strategy ''place & join.''

Although place & join is frequently used in de novo design approaches,[16,85] incremental construction is the preferred fragmentation-based

strategy for molecular docking. There are several reasons for favoring incremental construction. First, if a ligand is divided into fragments, not every fragment must lie at a low energy position by itself. Second, considering the individual energy contributions, distorting the ligand conformation, especially bond lengths and angles, is very expensive. Therefore, it makes sense to directly place fragments such that these distortions do not occur, which can be guaranteed by incremental construction. With place & join, however, a ligand's bonds are formed after placing the fragments, which may result in distorted conformations. A final energy minimization is often necessary, which could move the fragments away from their previously calculated favorable positions, thus negating much of the previous placement efforts.

The first place & join algorithm was developed by DesJarlais et al.[86] The ligand is manually divided into two fragments having one atom in common. Then, placement lists are created for each fragment using the docking algorithm DOCK. The algorithm searches through these lists for placement pairs in which the common atom is located approximately at the same point. Finally, the fragments are reconnected, energy minimized, and scored.

Sandak et al.[87–89] applied the geometric hashing paradigm to develop a place & join algorithm. As before, the ligand is divided into fragments with one overlapping atom, called the hinge. For each ligand atom triplet of a fragment, a hash table entry is created addressed with the pairwise distances between the atoms. Each entry contains fragment identification as well as the location of the hinge. In the matching phase, protein sphere triplets (DOCK spheres) are used to extract ligand atom triplets with similar distances. The method yields a vote for a hinge location for each match. A ''yes'' vote is cast for the identity of the ligand together with the location and orientation of the candidate ligand frame. Hinge locations with high votes are then selected; the fragments are reconnected accordingly and finally scored.

Place & join algorithms are advantageous in cases where the molecule consists of a small set of medium-sized rigid fragments. If the fragments are too small, it is difficult to place them independently. Another difficulty is the already mentioned problem of getting correct bond lengths and angles at the connecting atom without destroying the previously found favorable interactions of the fragments to the protein.

Docking algorithms based on incremental construction typically consist of three phases: The selection of a set of so-called base or anchor fragments, the placement of these base fragments, and the incremental construction phase itself. An incremental construction algorithm can start with several base fragments. In contrast to the place & join algorithms, the placements in incremental construction are not combined but taken as an anchoring orientation to which the remaining parts of the ligand can be added.

Both place & join and incremental construction originated from the area of de novo ligand design.[16,85] Moon and Howe[90] created the peptide design

tool GROW based on the incremental construction strategy. The first docking algorithm based on incremental construction was developed by Leach and Kuntz[91] based on the DOCK program. In the Leach–Kuntz approach, a single anchor fragment is first selected manually and docked into the receptor site using a variant of the DOCK algorithm, which handles hydrogen-bonding features in the matching phase. A subset of placements is selected favoring a high number of matched hydrogen-bond pairs, a high score of the anchor placement (see the Scoring section), and low similarity to other placements. Subsequently, incremental construction is started for the anchor placements chosen. In this phase, a backtracking algorithm is used that enumerates the space of nonoverlapping placements of the whole ligand in the receptor site. After adding a fragment to the current placement, a refinement routine is used to eliminate steric strain and to improve the hydrogen-bonding geometry. The final placements are then filtered, refined, and scored with a force field method. Although there are several manual steps in this procedure, the work demonstrated that the incremental construction idea could be applied to the docking problem.

The docking algorithm in the FlexX package is also based on incremental construction.[29,92,93] FlexX is a fully automated approach to molecular docking developed for virtual screening purposes. FlexX was the first tool that demonstrated computing times short enough for medium-sized screening efforts with flexible ligand molecular docking.

In the first phase of FlexX's operation, a small set of base fragments is selected. A simple scoring function is used to select fragments that are suitable for placement, that is, they contain a reasonably high number of interacting groups and exhibit a low number of low-energy conformations. A necessary condition for a successful calculation is that the selected base fragment binds to the protein and is not mostly exposed to water in the final protein–ligand complex. Therefore, it is extremely important to select a set of base fragments distributed over the ligand.

The base fragments are placed with the pose clustering algorithm mentioned above. Base fragment conformations are enumerated within the placement algorithm. The advantage of the pose clustering algorithm is that it is based on the molecular interactions instead of the shape of the fragment, thus allowing for the handling of much smaller fragments down to the size of a single functional group. All calculated placements form the input to the incremental construction phase. In contrast to Leach's algorithm, a "greedy strategy" is applied that selects a large number k of placements with the highest estimated scores (typically, k is ca. 800). Each iteration of the incremental construction algorithm contains the following steps: (1) adding the next fragment in all possible conformations to all placements from the previous iteration (or the base placement phase), (2) searching for new protein–ligand interactions, (3) optimizing the ligand position to improve the interaction geometry and reduce steric strain, (4) selecting a subset of placements with high

score, and (5) clustering the placements to achieve a reasonable degree of diversity in the solution set.

Ligand conformations within FlexX are based on the MIMUMBA model:[94] to each rotatable bond, a set of low-energy torsional angles is assigned, previously derived from a statistical analysis of the Cambridge Structural Database (CSD).[95] Ring system conformations are precomputed using the 3D structure generator CORINA.[96,97] For scoring protein–ligand complexes, a variant of Böhm's empirical scoring function[98] is used.

Some extensions of the FlexX approach have been recently developed. In one publication, the interaction model was extended such that hydrophobic fragments can be handled with the pose clustering algorithm.[99] To consider discrete water molecules and metal ions located in the protein–ligand interface, a technique called the particle concept was invented and integrated into the FlexX software. The particle concept allows for the automatic placement and energetic consideration of approximately spherical objects during the docking calculation.[100]

Two other approaches based on incremental construction have been published. The program Hammerhead[32] differs from FlexX in the construction strategy. Instead of adding small fragments (cut between each rotatable bond), the ligand is divided into a small set of large fragments. During the construction phase, the next fragment is added such that the connecting atom (or bond) overlaps and interactions to the protein can be formed. There is no discrete sampling at the torsional angle of the added fragment. However, high-energy conformations will also be created, and the situation in which a fragment does not interact directly with the protein is more difficult to handle than in FlexX.

Makino and Kuntz[101] developed a fully automated incremental construction docking algorithm based on backtracking. A single anchor fragment is selected maximizing hydrogen-bonding features. During the incremental construction phase, the number of conformations for each fragment is limited to reduce the size of the search space. This method is called a limited backtrack search. For scoring, the AMBER force field[102–104] is used with a modification that allows for the handling of multiple protonation states.

Genetic Algorithms and Evolutionary Programming

Since 1995, genetic algorithms (GA) are applied to the molecular docking problem in several approaches.[105–109] A genetic algorithm[110,111] is a general purpose optimization scheme that mimics the process of evolution. The individuals are configurations in the search space. A so-called fitness function is used to decide which individuals survive and produce offspring, which are then used in the next iteration of optimization.

Two aspects of the system must be modeled in order to use the idea of genetic algorithms for an application like molecular docking. First, a linear representation of a configuration such as a bit string is needed describing all

degrees of freedom of the problem. The linear representation is called a chromosome. Finding the chromosome description is the most difficult part of modeling. The description should be free of redundancy. If possible, any constraints of configuration space should be directly modeled such that configurations violating constraints are never generated during the optimization. Second, a fitness function has to be developed. The fitness function is closely related to scoring functions for molecular docking with one extension. Scoring functions normally work on 3D coordinates. Therefore the chromosome of an individual has to be interpreted in order to apply the scoring function (genotype-to-phenotype conversion). Since most of the computing time is spent evaluating the fitness function, the conversion and the evaluation have to be efficient.

The optimization scheme itself is more or less independent of the application. Typically, several GA parameters such as the population size, the number of generations, crossover, and mutation rates must be chosen.[111] It is important to find a balance between optimizing the fitness function and maintaining diversity in the population, so that there is a variety of individuals to select from at each iteration.

Jones et al.[21,105] developed one of the first genetic algorithms for molecular docking. Their ideas are implemented in the software tool GOLD. Among the genetic algorithm-based docking approaches, GOLD is probably the most widely used software. A configuration is represented in GOLD by two bit strings. The first string stores the conformation of the ligand (and the protein) by defining the torsional angle of each rotatable bond. The second string stores a mapping between hydrogen-bond partners in the protein and the ligand. For fitness evaluation, a 3D structure is created from the chromosome representation by first generating the ligand conformation. According to the mapping stored in the second string, hydrogen-bond atoms are superimposed to hydrogen-bond site points in the receptor site. Finally, a scoring function is used as the fitness function for evaluating hydrogen bonds, the ligand internal energy, and the protein–ligand van der Waals (vdw) energy.

Oshiro et al.[106] developed two variants of a docking method based on a genetic algorithm and the DOCK approach. The variants differ in the way the relative orientation of the ligand to the protein is described. The first variant encodes the matching of ligand atoms to protein spheres in the chromosome. A superposition is used to generate the 3D orientation of the ligand. The second variant stores the relative orientation directly by a translation vector and three Euler angles. For scoring, a simplified version of the AMBER force field[102–104] was used as the scoring function.

Gehlhaar et al.[108,112] developed a docking algorithm called EPDOCK based on evolutionary programming. In contrast to a genetic algorithm, offspring are created from one parent by mutation only. Each member of a population is competing for survival in a kind of a tournament. EPDOCK contains

its own scoring function based on atomic pair-wise potentials for steric interactions and hydrogen bonding.

Molecular Simulation

Whereas all previously discussed methods for solving the docking problem are based on some kind of combinatorial optimization algorithm, there are several approaches for tackling the problem by simulation techniques. Instead of trying to enumerate the population of a discrete low-energy subspace of the problem, the simulation approaches begin their calculation with a starting configuration and move locally to configurations with lower energy.

Simulated annealing[113] is a well-known simulation technique, which is also frequently used for solving complex optimization problems without any physical interpretation of the simulation itself. The overall simulation routine iterates the following steps. (1) Starting with a configuration A with an energy or score value $E(A)$, we calculate a random local move to a new configuration B with energy $E(B)$. (2) The acceptance of the new configuration is based on the Metropolis criterion; namely, the configuration is accepted if its energy is lower or with probability $P = \exp(-[E(B) - E(A)]/[k_B T])$ if its energy is higher. Here k_B is the Boltzmann constant, and T is the current absolute temperature. During the simulation, T is reduced based on a so-called cooling schedule such that accepting configurations with increased energy becomes less likely over time. The AutoDOCK program for protein–ligand docking developed by Goodsell and co-workers[33,114,115] is based on this strategy. For energy calculation, molecular affinity potentials[116] are precalculated on a grid. Yue[117] developed a program for optimizing distance constraints for rigid-body docking based on simulated annealing.

Molecular Dynamics Molecular docking problems can in principle be solved with molecular dynamics (MD) simulations. This methodology is quite general because various energy models can be used to simulate the process of molecular recognition. However, since the method is following a path from a starting orientation to low-energy configurations, the simulation process can be quite time consuming. In addition, several simulations with different starting orientations must be performed to get a statistically significant result. Thus far, no MD based technique exists for handling really large data sets in a reasonable time frame. However, the methodology is of value for analyzing small sets of ligands.

In MD, a force field is used to calculate the forces on each atom of the simulated system. Then, following Newtonian mechanics, velocities and accelerations are calculated, and the atoms are slightly moved with respect to a given time step. Introducing molecular dynamics and force fields is clearly beyond the scope of this chapter, so we refer the reader to earlier reviews in this book series.[118–123] However, some aspects of docking based on MD simulations should be briefly mentioned.

To avoid very long computing times, scientists look for methods to make greater moves of the ligand in a single step. Larger steps decrease the dependency of the outcome of the docking simulation from the starting structure and allow for a better sampling of the search space. A good example for such a method is the "helicopter view" developed by DiNola and coworkers.[124,125] For a limited time at the beginning of the simulation, the temperature of the system is increased for some degrees of freedom (protein–ligand relative orientation) and the repulsive terms of the energy function are decreased. This increases the amount of structural changes between two simulation steps and makes it easier for the system to escape from local minima.

Given and Gilson[126] developed a four-phase docking protocol based on MD simulation. First, a set of low-energy ligand conformations is created using MD with alternating heating (to perturb the structure) and cooling (to minimize the structure). Then the ligand is several times randomly placed into the receptor site and minimized. In the final phase, the most stable configurations are further investigated using MD with more alternating heating and cooling. The goal of the last phase is to explore the search space around the stable conformations in more detail.

Monte Carlo Algorithms In an MD simulation, the local movement of the atoms is performed due to the forces present at each step. In contrast, in a Monte Carlo (MC) simulation, the local moves of the atoms are performed randomly.

Two components are of major importance in an MC algorithm: the description of the degrees of freedom and the energy evaluation. The degrees of freedom should ideally be described such that high-energy states are avoided. Describing the ligand flexibility by internal coordinates (bond angles, torsional angles) can be an advantage compared to a description by 3D coordinates in Cartesian space because the force field and MC moves are also coded in internal coordinates. The most time-consuming part of an MC simulation is the energy (score) calculation. This step must be as time efficient as possible. Often, energy potentials are precalculated on a grid to speed up this step. There are several examples for MC-based docking algorithms. Hart and Read[127] developed a combined scheme of MC with simulated annealing. In the MC run, random orientations are created and moved such that the protein–ligand overlap is reduced. The results are then optimized using a simulated annealing scheme. Protein–ligand overlap and scores are calculated based on precalculated grids. McMartin and Bohacek's[128] QXP program is a similar approach and can be applied to molecular docking and structural superposition. Wallqvist and Covell[129] also use MC for optimizing the final ligand orientation. Instead of a random set of starting structures, a surface matching algorithm is used to align the ligand to a complementary region of the protein.

The ICM (Internal Coordinate Mechanics) software package developed by Abagyan et al.[130] combines MC as the optimization scheme with an internal

coordinate description. In contrast to other approaches that make purely random moves, ICM is able to model probability functions for variable sets. This ability allows for biasing the MC calculation toward low-energy states.

Apostolakis et al.[131] developed a multistep docking protocol based on MC combined with energy minimization. After a random creation of starting structures, a minimization is performed using a modified van der Waals potential. The energy function for evaluating the conformations is a sum of three terms: a force field energy, a hydrophobic solvation term that is proportional to the solvent accessible surface (SAS) of the complex, and an electrostatic solvation term obtained from the solution of the linearized Poisson–Boltzmann equation.[132] The highest scoring configurations are further analyzed with a Monte Carlo simulation method called MCM that performs Monte Carlo moves interleaved with energy minimization steps.

PRODOCK is an MC-based docking algorithm developed by Trosset and Scheraga.[133–135] The software allows use of either the AMBER 4.0 or the ECEPP/3 force field with a grid-based energy evaluation. For calculating energy values within the grid, Bezier splines were used[136] allowing for a more accurate estimation of the energy value as well as providing information about the derivatives at this point. Like ICM, an internal description of the degrees of freedom is used. The MC simulation is interleaved with energy minimization steps like in the Apostolakis' approach.

Alternative Methods

There is a large variety of methods for molecular docking that do not fit into one of the most popular algorithmic classes mentioned above. Although only some of them are more widely used today, they form a rich set of computational methodologies, which can be used for all kinds of subproblems occurring in structure-based ligand design.

Distance Geometry Distance geometry is a well-known technique from the area of structure determination via NMR technology.[137,138] Instead of describing a molecule by coordinates in Euclidean space, it is described by a so-called distance matrix containing all interatomic distances in the molecule. Based on distance matrices, a set of allowable conformations can be described in a comprehensive form from a distance interval for each atom pair.

The distance geometry methodology can be directly used in docking algorithms based on clique search and distance compatibility as discussed above. Matches between protein site points and ligand atoms are compatible, if the site point distance lies within the distance interval of the ligand atom pair. The drawback of this approach is that the distance matrix is overdetermined, that is, for most distance matrices there is more than one set of Cartesian coordinates of the atoms that satisfy the distance criteria. If matches are formed, the distances between the matched ligand atoms are fixed to the distance of the matched features in the protein. Fixing the atom–atom distance between a given atom pair to a single value usually causes other distance

intervals to shrink. New upper and lower bounds for distances can then be approximated exploiting the triangle and tetrangle inequalities.[139,140] Nevertheless, a very limited number of distance matrices can be converted back to 3D space. Examples for methods based on distance geometry have been reported.[141–143]

Random Search Most random search approaches involve Monte Carlo algorithms already discussed above. However, there are exceptions in docking algorithms that we would like to address here. If a scoring function for evaluating protein–ligand complexes is available, random search algorithms can be applied to the docking problem. Random placements can be created either independent of each other (Markov chain) or from a (random) starting structure by subsequent random moves (somewhat mimicking MD simulations). In most cases, the structure generation is combined with a numerical optimization method that drives the placements of the ligand to the closest local minima as done in LIGIN.[144]

Baxter et al.[145] use tabu search as the underlying method of their docking software called PRO_LEADS. Tabu search starts with an initial random structure; new structures are created by random moves. During the optimization iterations, a list (the tabu list) is maintained containing the best and most recently visited configurations. Moves resulting in configurations close to one in the tabu list are rejected except if they are better than the best scoring one. The tabu list technique improves the sampling properties of the random search algorithm by avoiding revisitation of previously sampled configurations.

Hybrid Methods Due to the complexity of the docking problem, all methods have their pros and cons. No one method will work well for every problem. Combining different methods is therefore a reasonable approach, which can result in methods containing the best of each.

Two recently published methods combine rapid fragment-based searching techniques with sophisticated MD or MC simulations. Wang et al.[146] developed a multistep approach based on rigid-body docking and MD. First, a set of low-energy conformers is created. Each conformer is docked rigidly into the receptor site using DOCK. The high scoring placements are then optimized using an MD-based simulated annealing optimization in combination with the AMBER force field. Hoffmann et al.[147] combined the incremental construction algorithm in FlexX with an MD-based procedure for postoptimization. FlexX is first used to create a sample of a few hundred ligand placements. The goal of the second phase is to improve the overall ranking of the solutions and to identify the correct placement. The placements are energy minimized using the CHARMm22 force field.[148,149] The first phase yields a large number of plausible ligand conformations. Reranking using the software package CAMLab[150] prioritizes those conformations. The final score consists of three contributions, a force field energy, an electrostatic part of the solvation energy, and a nonpolar part of the solvation energy. The electrostatic part of the solvation energy is calculated by solving the Poisson equation using fast

multigrid methods; the nonpolar part is approximated by the total solvent accessible surface of the protein–ligand complex.

Handling Protein Flexibility

Most approaches for protein–ligand docking treat the protein as a rigid structure. Although this is an oversimplification in the general case, for many proteins this is a reasonable assumption, but for other proteins the assumption fails. Protein flexibility plays an important role in protein–ligand binding, and algorithms considering these additional degrees of freedom should work better.

In principle, protein flexibility can directly be handled by MD- or MC-based algorithms as described above. A method to improve time efficiency during the MD simulation with a flexible protein was developed by Wasserman and co-workers.[151,152] In their work, the protein is divided into rigid and flexible parts. Every atom, sufficiently far away from the receptor site, is considered as fixed in space; all energy contributions from these atoms can be precalculated on a grid. During the MD simulation, only the protein's receptor site atoms and the ligand atoms are allowed to move and are considered explicitly in the energy calculation. The method can be combined with an implicit model of solvation.

Although such approaches lead to a significant speedup over methods that treat the protein as fully flexible, they are still too time consuming for screening purposes. Therefore, one is interested in introducing protein flexibility into the much faster, fragmentation-based approaches.

To accomplish this, Leach[153] developed a docking algorithm similar to incremental construction such that the degrees of freedom are fixed sequentially. Degrees of freedom considered in Leach's method are a discrete set of ligand orientation and conformation as well as a set of rotamers for selected protein side chains. To search efficiently through the space of possible configurations, Leach used a variant of a branch & bound scheme, called A*-algorithm with dead-end elimination (DEE). The A*-algorithm[154] is a search tree algorithm that involves the expansion of nodes. It is capable of finding the optimal (least cost) path to the global energy minimum of an energy function for the assignment of side chain conformations. The DEE[154] algorithm on the other hand identifies rotamers of side chains that are incompatible with the global minimum energy structure by satisfying the inequality

$$E_{i_r,\text{rigid}} + \sum_j \min(s)\varepsilon_{i_r,j_s} > E_{i_t,\text{rigid}} + \sum_j \max(s)\varepsilon_{i_t,j_s} \qquad [1]$$

where $E_{i_r,\text{rigid}}$ is the interaction energy between rotamer r of residue i and the backbone, ε_{i_r,j_s} is the interaction energy of rotamer r of residue i and rotamer s of residue j, and $\min(s)\varepsilon_{i_r,j_s}$ and $\max(s)\varepsilon_{i_t,j_s}$ are determined by screening all

other rotamers that are currently permitted. To address the problem that the true global energy minimum of the protein–ligand complex does not necessarily correspond to that of the force field, Leach applied a modified DEE/A* algorithm in which an energy threshold is introduced that allows all structures within a certain energy difference to the global minimum to satisfy Eq. [1]. The large number of possible structures that survive Eq. [1] then defines the space searched by the A* algorithm.

Knegtel et al.[155] extended the DOCK algorithm so that it could handle partial protein flexibility. Instead of using a single protein 3D structure as input, an ensemble of protein structures is used for the docking calculation. During the precalculation of scoring grids, all protein structures are used at the same time such that the grid represents an average structure. In this averaging process, the repulsive part of the van der Waals term is neglected for flexible parts of the protein. The advantage of this method is its speed and simplicity resulting from the fact that the combinatorics of the protein structure is avoided. This is, of course, a very rough way of approximating protein flexibility, especially for the protein–ligand overlap estimation.

The screening software SPECITOPE, developed by Schnecke et al.[156], uses distance matrix comparisons as a first filter step. However, flexibility of molecules is not modeled by distance intervals. Instead, a weighting scheme is defined to scale down the contributions of more flexible atom pairs in the overall score. In addition, a special optimizer is added that allows the removal of protein–ligand steric clashes after the placement calculation.

Recently, Claussen et al.[157] presented a new algorithm based on the idea of protein structure ensembles implemented in the program FlexE. The algorithm considers the alternative protein conformations in a truly combinatorial way by calculating the highest scoring protein conformation for each partially placed ligand during an incremental construction algorithm. The new method is applied to a series of protein ensembles with known ligands. It can be shown that FlexE successfully docks sets of ligands into proteins where protein flexibility is required.

Docking of Combinatorial Libraries

The development of combinatorial chemistry and its application to drug design[39,40] has had a dramatic impact not only on experimental screening methods but also on virtual screening. Due to the increasing number of molecules that can be theoretically synthesized, the demand for the throughput of experimental and virtual screening systems increased. For real-world pharmaceutical research, this development led to high-throughput screening based mostly on a miniaturization of the equipment and parallel processing.

For virtual screening, the situation is somewhat different. Screening can be understood as a search process for finding the best lead structures available in a database. Search problems typically become easier to solve if the search

space is structured. If we search the shortest path to go between two places, we would never try all possible paths and measure their lengths since we already know that most paths will not be the shortest due to the structure of the search space, which can be expressed, for example, by the triangle inequality. In virtual screening, hypothetical combinatorial chemistry can be used to introduce structure into a search space. Combinatorial libraries arise systematically trying all combinations of a limited set of building blocks. The resulting database space of the enumerated combinatorial library becomes structured by the combination pattern of the enumeration scheme. The structuring of the database can be exploited to reduce the runtime of virtual screening calculations.

In the context of combinatorial libraries, we distinguish between three kinds of docking problems:

1. Combinatorial Docking Problem: Given a library of ligands, calculate the docking score (and the geometry of the complex) for each molecule of the library.
2. R-Group Selection Problem: Given a library, select molecules for the individual R-groups in order to form a smaller sublibrary with an enriched number of hits for the target protein.
3. De novo Library Design Problem: Given a catalog of available reagents, design a library (including the rules of synthesis) that will optimize the number of hits for the target protein.

Methods for treating these problems have emerged from the area of molecular docking and de novo ligand design (see, e.g., Kubinyi[158] for an overview on combinatorial docking methods).

Early algorithms for the combinatorial docking problem analyzed the similarity among the members of a given ligand data set to speed up the search process. One approach was to generate a minimal tree structure representing the whole ligand data set.[159] Another approach is to speed up conformational searching based on clustering similar molecules.[160] In both cases, the derived hierarchy of molecules can be used in an incremental construction docking method.

Recent combinatorial docking tools such as PRO_SELECT[161] and CombiDOCK[162] are based on the incremental construction method. In both, a library is formed by a template (or core) molecule with a set of attachment points at which a predefined set of substituents can be connected. The template is then positioned in the receptor site without considering the substituents. Starting from a few orientations of the template, the substituents are attached to the template independently and the fit to the receptor is assessed. In case of PRO_SELECT, substituents are selected based on a score (see next section for an introduction to the scoring concept) and additional criteria such as 2D similarity to other substituents and feasibility of synthesis. CombiDOCK calculates a score for enumerated library molecules by combining fragment scores.

A new algorithm for the combinatorial docking problem based on the FlexX incremental construction algorithm was reported recently.[163] The algorithm is part of a new combinatorial docking extension of FlexX called FlexX^c. FlexX^c considers the combinatorial library as a tree structure. Each node represents either the core or an R-group of the library. Molecules from the library, required during the docking calculations, are created on the fly. Therefore, the library is handled in its so-called closed form, enumerating the library molecules in advance is not necessary. With this method, libraries with a few hundred thousand molecules can be handled in main memory during a docking calculation. The tree representation allows the handling of complex libraries consisting of several R-groups, that can be arbitrarily linked. One exception is that no ring closures over different R-groups are allowed. A recursive combinatorial library extends the incremental construction method to multiple molecules; in each incremental construction step, all possible R-group molecules are added sequentially.

Some studies based on ligand de novo design software have been published for R-group selection problems. Kick et al.[164] applied a variant of the BUILDER program[165] to the preselection of substituents for a library targeted to cathepsin D. Recently, Böhm applied the LUDI[69,166] program to the docking of two groups of fragments that can be connected pairwise in a single-step reaction to search for new thrombin inhibitors.[167] In principle, all programs for fragment-based de novo ligand design can be applied in a similar way to the R-group selection problem.

Finally, we mention two methods for de novo library design. Caflisch[168] applied the MCSS[169] technique generating fragment placements, which are subsequently connected. The DREAM++ software combines tools for fragment placement and selection.[170] The selection process is done such that only a small set of well-characterized organic reactions is needed to create the library.

SCORING

The binding affinity between a protein and a ligand that form a noncovalent complex can be experimentally determined, for example, by microcalorimetry[171] or surface plasmon resonance.[172] The affinity depends on several environmental parameters such as pH and ionic strength of the solvent.[173] Experimental errors typically lie in the range of 0.1–0.25 kcal/mol, although repeated assays of the same compound can give even wider variability in some biological systems. Even assuming there exist good experimental data for calibrating a computer determination of binding free energies, making reliable predictions is a very challenging task that has been pursued for decades.[174] The binding free energy of a protein–ligand complex can be intuitively understood as the difference in how much the ligand likes to be in the protein

binding pocket and how much it likes to be in the solvent (usually aqueous solution) at infinite separation from the protein. Although there have been arguments against it,[175] the binding free energy is usually dissected into a manifold of ensemble-averaged terms that are sometimes hard to calculate, for example, solvation energies of the ligand, the protein, and the complex ($\Delta G_{\text{solv}}^{\text{lig}}, \Delta G_{\text{solv}}^{\text{prot}}, \Delta G_{\text{solv}}^{\text{complex}}$), interaction energy between the protein and the ligand in the complex (ΔG_{int}), entropy changes for the ligand and the protein ($T\Delta S$), conformational changes (reorganization energy) in ligand and protein ($\Delta\lambda$) upon complex formation. The binding free energy can then be written as

$$\Delta G_{\text{bind}} = \Delta G_{\text{solv}}^{\text{complex}} - \Delta G_{\text{solv}}^{\text{prot}} - \Delta G_{\text{solv}}^{\text{prot}} + \Delta G_{\text{int}} - T\Delta S + \Delta\lambda \qquad [2]$$

Disregarding the fact that there is no unique set of terms in which the binding energy must be expressed, it is key to calculate *all* the terms and not an arbitrary subset.[176] One problem is that large energy contributions have to be balanced in order to estimate small binding free energies,[177,178] a calculation prone to large statistical and systematic errors.[179,180] A second problem arises from the fact that some contributions such as entropy can only be crudely estimated.[181,182] A third problem is that binding free energies can only be rigorously calculated by fully sampling the conformational space available to the ligand in the protein-binding pocket. Algorithms such as free energy perturbation[179] or the linear response approximation,[183,184] which are employed to facilitate the sampling/free energy calculation task, are too time consuming to be used in docking a large set of ligands. The fast estimation of binding free energies requires many simplifications in the model. Simple functions that are designed to rank protein–ligand complexes according to their binding affinities are therefore referred to as scoring functions rather than energy functions. Scoring functions used in docking studies have been reviewed recently.[35,38,174,185] Here, we briefly summarize older techniques and discuss new developments such as knowledge-based scoring functions and consensus scoring.

Scoring functions are the Achilles' heel of docking algorithms. They can be divided into several categories. Scoring functions derived from a force field use nonbonded interaction terms as the score,[21,102,103,186] sometimes in combination with solvation terms.[67,187] Empirical scoring functions employ multivariate regression methods to fit coefficients of physically motivated structural functions by using a training set of protein–ligand complexes with measured binding constants.[69,98,188–193] Knowledge-based scoring functions use statistical atom pair potentials derived from structural databases as the score.[194–200] In addition, more heuristic scoring schemes[112] such as chemical scores, contact scores, or shape complementary scores are also used.[37,62,201–205] The fact that the well-established docking programs do not all use the same scoring function underlines that no optimal scoring function has yet been found. In fact, no scoring function has been developed so far that reliably and consistently

identifies the correct binding mode of a protein–ligand complex.[206] Scoring functions are usually designed and tested to identify binding modes of known ligands that bind to a well-resolved protein with high affinity. In thoroughly validated docking studies, about 70% of known protein–ligand complexes can usually be reproduced correctly by flexible docking methods.[21,29]

Shape and Chemical Complementary Scores

Hydrophilic and hydrophobic surface complementarity are often used for scoring. Bohacek and McMartin divided the Lee and Richards[207] accessible protein surface into three separate zones representing hydrophobic, hydrogen-bond donating, or H-bond accepting properties. For nine thermolysin–inhibitor complexes, a correlation between complementary contacts and binding affinity with an r^2 of 0.99 could be achieved. Gehlhaar et al.[112] defined a piecewise linear potential (PLP) function used for protein–ligand interactions. The potential combines hydrogen-bond donor, hydrogen-bond acceptor, both donor and acceptor, and nonpolar atom types to steric or hydrogen-bond interactions. The parameters for this simple and very fast scoring scheme were optimized employing docking experiments on dihydrofolate reductase (DHFR), FK506-binding protein, and HIV-1 protease. Consequently, the crystal structures represent the global minima of the PLP scoring function. Intraligand interactions were treated using nonbonded force field terms and a torsional potential term.

Shape and chemical complementarity are also accounted for in DOCK where atomic contact potentials for the receptor can be used to construct a cubic lattice that fills the volume of the binding site.[62] A point on the lattice receives a score of one for every atom within a certain distance and highly negative scores for any contact that is too close. Polar and apolar contacts can be distinguished. A ligand is scored mapping its atoms onto the nearest lattice points and summing up over all mapped points. Chemical matching (chemical scoring) can be applied in DOCK by ''coloring'' the DOCK spheres with respect to the properties of the nearby receptor atoms.[55,209] During the sampling process, a sphere can only be matched with a ligand atom of complementary color. Chemical color labels can decipher properties such as ''H-bond donor,'' ''H-bond acceptor,'' ''hydrophobe,'' ''electronegative,'' ''neutral,'' and so on. These colors are not predefined in DOCK but have to be assigned by hand.

Walls and Sternberg[210] considered surface complementary for docking protein–protein complexes using a vdW-like soft core potential. The use of accessible surface area on binding and complexation has been discussed, for example, by Flower.[205] Surface descriptors have been used to calculate solvation energies as part of the energetics of molecular association.[144,188,211] Volume overlap was also used to quantify interactions of chemical compounds.[201] The HINT exponential function and empirical atomic

hydrophobicities were used to evaluate docked conformations.[65] Viswanadhan et al.[212] employed atomic hydrophobicity constants derived for Ghose and Crippen atom types[213] to calculate hydrophobic contributions to the binding energy. Ligand intramolecular strain energy is sometimes used to complement intermolecular scoring functions.[202,203] SPECITOPE uses a pharmacophore filter to filter out unlikely ligands.[156] A scoring function is then applied that consists of the number of hydrogen bonds between protein and ligand and a hydrophobic complementarity term. The hydrophobic measure was derived from a knowledge base of 56 protein structures.[214] Shape complementary scores are especially useful in rigid docking experiments. They have the ability to identify correct binding modes. However, shape and chemical complementary scores were not very useful for ranking different putative protein–ligand complexes. SANDOCK uses several scoring schemes that combine tuned geometric, hydrophobic, and hydrogen bonding terms that are calculated on solvent accessible surface points of ligand and protein.[215] Recently, Goldman and Wipke[216] reviewed the field of shape descriptors and also introduced new quadratic shape descriptors (QSD) for molecular docking.

Force Field Scoring

Force field scoring functions, especially when used on a precomputed grid, are fast and transferable. A conceptual advantage is that force field terms are well studied and have a physical basis. A disadvantage is that force field scores measure only parts of the relevant energies, namely, potential energies (rather than free energies) sometimes enhanced by solvation or entropy terms. As a consequence, electrostatics often dominates in docking. Electrostatic complementarity helps form hydrogen bonds and salt bridges, but the electrostatic contribution to the calculated binding score is usually largely overestimated leading to systematic problems in ranking putative protein–ligand complexes. Unfortunately, this cannot be prevented by scaling electrostatic interactions down by a uniform dielectric constant ε, because in vacuo ion-pair interactions need to be scaled down much farther than dipole interactions (e.g., see discussion in Refs. 217,218).

Force field scoring is based on the idea of using only the enthalpic contributions to estimate the binding free energy. Most often, nonbonded interaction energy terms of standard force fields are used, such as an in vacuo electrostatic term (sometimes modified by scaling constants that assume the protein to be an electrostatic continuum) and a vdW term.[185,186,219,220] The force field score used in programs such as DOCK[26] and GREEN[221] consists of the intermolecular terms of the AMBER energy function,[102,103] except for an explicit hydrogen-bonding term.[64]

$$E_{\text{nonbond}} = \sum_{i}^{\text{lig}} \sum_{j}^{\text{prot}} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + 332 \frac{q_i q_j}{D r_{ij}} \right] \qquad [3]$$

where each term is summed up over ligand atoms i and protein atoms j. The parameters A_{ij} and B_{ij} are the vdW repulsion and attraction parameters of the 6-12 potential, r_{ij} is the distance between i and j, q is the point charge on an atom, and D is a dielectric constant (treating the protein as a homogeneous continuum). In addition, intraligand interactions are added to the score. A significant reduction in computer time is achieved by precomputing terms of the potential that are sums over protein atoms. Assuming a rigid protein structure upon docking, protein terms can be calculated on a 3D grid.[33,116] In contrast to continuous scoring, where for each ligand atom all interactions with engulfing protein atoms have to be calculated repeatedly, the gridded score interpolates from eight surrounding grid points only. This gridding approach usually leads to a 100-fold speed up in docking calculations. In DOCK, the electrostatic potential can be calculated on a grid using a discretized continuum method.[222]

Docking programs that do not use incremental build up procedures to flexibly dock a ligand in a binding site often use soft vdW potentials. For example, FLOG uses a 6-9 Lennard-Jones function for vdW interactions, local dielectric constants in a Coulomb representation of the electrostatic interactions, and additional terms for hydrogen-bond potentials and hydrophobic potentials to score protein–ligand conformations.[34] GOLD uses a scoring function that combines a soft intermolecular vdW potential (Lennard-Jones 4-8 potential) with additional terms for hydrogen bonding and intraligand energy.[21] Hydrogen-bond energies between donors and acceptors were precalculated using model fragments and accounting for desolvation.[223] Containment potentials that force a ligand to dock in a certain region are sometimes applied.[224] Sometimes complete force field energy functions including bonded terms are used as the scoring function.[134]

Nonbonded interaction terms in scoring functions have also been combined with solvation and entropy estimates. The importance of solvation for binding is illustrated, for example, by the finding that solvation energy *alone* can sometimes highly correlate with binding free energy.[225] Entropy terms on the other hand can often be neglected when molecules of similar size and internal degrees of freedom are compared. Caflisch and co-workers[131,226] combine the total CHARMM[227] energy with solvation energy terms decomposed into nonpolar and electrostatic contributions. Solvent effects are often approximated by rescaling the in vacuo Coulomb interactions by $1/\varepsilon$, where the dielectric constant ε ranges between 1 and 80.[126] A generalized Born/surface area approach is sometimes used.[228]

DOCK introduces ligand solvation into the scoring function by

$$E_{\text{bind}} = E_{\text{nonbond}} - E_{\text{solv,elec}} - E_{\text{solv,np}} \qquad [4]$$

where the electrostatic hydration energy $E_{\text{solv,elec}}$ is calculated using continuum electrostatics (Born equation) and the nonpolar component $E_{\text{solv,np}}$ is

calculated based on the surface area of the ligand in the solvent.[67] The change in solvation free energy upon protein–ligand binding may also be accounted for by atomic solvation parameters.[211,229,230] Force field terms used in scoring functions depend on several parameters such as dielectric constants, nonbonded cutoff schemes, scaling of surface charges, and modification of vdW repulsion. An analysis of these parameters using the CHARMM energy function was given recently by Vieth et al.[231] It was concluded that a soft-core vdW potential is needed for the kinetic accessibility of the binding site.

Empirical Scoring Functions

The need for fast ranking of protein–ligand complexes led to the design of empirical scoring functions. They use multivariate regression methods that fit coefficients of physically motivated contributions to the binding free energy to reproduce measured binding affinities of a training set of known 3D protein–ligand complexes.[98,188–192] Horton and Lewis[188] developed an empirical scoring function for protein–protein interaction using polar and apolar surface area as well as an entropy term as descriptors. The empirical scoring function for protein–ligand complexes designed by Böhm[98,189] is probably the most prominent one. It was originally developed for the de novo design program LUDI[69] and later used (slightly modified) in the docking program FlexX.[29] The free energy of binding for a protein–ligand complex is estimated in FlexX as the sum of free energy contributions from the number of rotatable bonds in the ligand, hydrogen bonds, ion-pair interactions, hydrophobic and π-stacking interactions of aromatic groups, and lipophilic interactions:

$$\begin{aligned}\Delta G = \Delta G_0 + \Delta G_{\text{rot}} N_{\text{rot}} + \Delta G_{\text{hb}} \sum_{\text{neutral\_Hbonds}} f(\Delta R, \Delta\alpha) \\ + \Delta G_{\text{io}} \sum_{\text{ionic\_int}} f(\Delta R, \Delta\alpha) + \Delta G_{\text{aro}} \sum_{\text{aro\_int}} f(\Delta R, \Delta\alpha) \\ + \Delta G_{\text{lipo}} \sum_{\text{lipo.cont}} f^*(\Delta R)\end{aligned} \quad [5]$$

where $\Delta G_0, \Delta G_{\text{rot}}, \Delta G_{\text{hb}}, \Delta G_{\text{io}}, \Delta G_{\text{aro}}$, and ΔG_{lipo} are adjustable parameters that were fitted. The parameter $f(\Delta R, \Delta\alpha)$ is a scaling function penalizing deviations from the ideal geometry, and N_{rot} is the number of freely rotatable bonds. The interaction of aromatic groups (aro_int term) is an addition to Böhm's original force field design.[98,189] The lipophilic contributions (lipo.cont) are calculated as a sum of atom pair contacts in contrast to evaluating a surface grid as in Böhm's scoring function. Böhm's scoring function and its FlexX implementation were being improved and additional terms were being tested (see, e.g., Refs. 31,189). Böhm's scoring function is somewhat sensitive to the orientation and conformation of the ligand in the binding pocket.

The placement of hydrogen atoms influences estimated binding energies significantly. Jain introduced an empirical scoring function to overcome this issue.[192] His function includes hydrophobic, polar complementary, entropy, and solvation terms that are constructed somewhat arbitrarily using Gaussian-like and sigmoidal basis functions with 17 tunable parameters. The problem of treating surface-related properties, which are particularly important for solvation contributions and protein–ligand interaction terms, was elegantly solved by using nonoccluded atom pair interactions rather than surface area terms. This treatment speeds up the computation of the binding score. Calibrated on a data set of 34 ligand–protein complexes, the mean error of predicted binding affinities (under leave-one-out cross-validation conditions) was 1.0 log K_i units making this scoring function the best one reported so far. Jain's scoring function is implemented in the docking program Hammerhead.[32]

VALIDATE is a program that defines protein–ligand binding by a master equation combining molecular mechanics terms with physicochemical properties.[35,191] Eight coefficients scale terms for vdW and Coulomb interactions, surface complementarity, lipophilicity, conformational entropy and enthalpy, and lipophilic and hydrophilic complementarity between protein and ligand surfaces. These coefficients are fitted using 51 diverse receptor–ligand complexes as the training set. Another simple correlation method used physical descriptors such as molecular weight combined with numbers of ligand–protein contacts, interaction energy, and conformational entropy.[193]

The docking program PRO_LEADS[232,233] uses an empirical scoring function that estimates lipophilic and metal-binding contributions and that includes a hydrogen-bond term and a flexibility penalty term.[190] Eighty-two protein–ligand complexes with measured binding affinities served as training set for determining five coefficients in the scoring function. The derived function gave a cross-validated error of about 2 kcal/mol for the training sets. An application of Bayesian regression improved the predictive power of the model.[234] An empirical scoring function similar to that of Murray and coworkers[190] was devised by Wang et al.[204] Eleven parameters were fitted using a training set of 170 protein–ligand complexes leading to a cross-validated deviation of 1.5 kcal/mol for this training set.

Empirical scoring functions are now used in many docking programs. The functions are fast and show good predictive power for known protein–ligand complexes. Conceptual disadvantages come from the fact that the functions are trained solely on crystal structures of protein–ligand complexes with medium-to-strong binding affinities. This method of derivation (i.e., solely from such crystal structures) results in empirical scoring functions that have no effective penalty term for bad conformations. The most serious concern, however, is that it is unclear to what extent empirical scoring functions can be used on test sets structurally different from those used in the training sets. Unfortunately, most scoring studies present only very limited test sets

making it hard to compare one study to another. Nevertheless, success stories have been reported, for example, finding biotin with the highest score amongst 80,000 compounds from the Available Chemicals Directory[235] (ACD) docked to streptavidin.[32] Note, however, that finding a high affinity ligand in a pool of nonbinders is relatively easy.[43] More challenging is finding weakly potent compounds in pools of nonbinders.[45]

Knowledge-Based Scoring Functions

In contrast to fitting approaches pursued in empirical scoring functions described above, knowledge-based scoring functions constitute a deductive approach. Since the forces that govern the interaction between proteins and ligands are so complex, known protein–ligand structures are taken as the only reliable source of information. The goal is to extract forces and potentials from a database of known protein–ligand complexes that establish a score for their binding affinities.

The idea of deriving knowledge-based potentials for scoring protein–ligand complexes has been inspired by the success of similar potentials in protein folding and protein structure evaluation.[236–242] Protein–ligand atom pair potentials can be calculated from structural data under the assumption that an observed crystallographic complex represents the optimum placement of the ligand atoms relative to the protein atoms. There are upwards of ca. 12,000 protein structures in the PDB, and many of them containing cocrystallized ligands. Hence there are millions of observed distances between ligand and protein atoms. Statistical preferences can be derived between protein and ligand atom types that are similar but not strictly equal to potentials of mean force (PMF).[243,244] The protein cannot be considered as a liquid of different atom types in its equilibrium state. The protein atoms are rather constrained by covalent bonds. During the folding process, entropy is exported and dipoles become preorganized in a specific way. Therefore, statistical mechanics of liquids cannot be rigorously applied to proteins. Statistical potentials that are derived from statistical mechanics are really only statistical preferences.[200,245] Nevertheless, due to similarity to the concept of statistically averaged forces, we loosely call them potentials of mean force here.

Verkhivker et al.[199] were first to use knowledge-based potentials in a model that described the total binding free energy between protein and ligand as a sum over protein–ligand, protein–water, and ligand–water interactions, nonpolar and polar desolvation, ligand and protein conformational changes, and terms accounting for the change in rotational and translational degrees of freedom of protein–ligand and water upon complex formation. To express protein–ligand interaction by a mean-field term, 12 protein–ligand atom pairs were defined, and PMF were determined based on 30 HIV-1 and simian immunodeficiency virus (SIV) protein–ligand complexes. The complete scoring function containing the mean-field term for protein–ligand interactions was then

successfully applied to reproduce the binding affinities of seven other HIV-1 protease complexes. However, the mean-field contribution alone did not show any correlation with the experimental data.

A preference-based free energy score of enzyme–inhibitor binding was developed by Wallqvist et al.[200] It was applied in docking experiments of HIV-1 inhibitors.[129] Atomic interaction scores between adjacent buried surfaces of protein and ligand atoms in the binding site were derived by analyzing 38 high-resolution protein crystal complexes taken from the PDB. Atom pair preference scores P_{ij} between protein atoms i and ligand atoms j for 21 atom types (10 carbon, 5 oxygen, 5 nitrogen, 1 sulfur atom type) were calculated and combined with three constants, α_{ij}, γ, and δ to predict protein–ligand binding affinities,

$$\Delta G_{\text{pred}} = -\sum_i \sum_j \alpha_{ij}(\gamma + \delta \ln P_{ij}) + \beta \qquad [6]$$

This scoring function showed a reliability of 1.5 kcal/mol when tested on eight protein–ligand complexes.[200] The preference-based scoring scheme was used in docking experiments following a surface complementary screen. It was found that 20 rigid enzyme–inhibitor complexes could be reassembled with all-atom root-mean-square deviations of 1.0 Å from the native complexes.[129]

A coarse grained knowledge-based potential was derived by DeWitte and Shakhnovich[194] as an energy function in the de novo design program SMoG. In the SMoG score, atom pair preferences g_{ij} are combined with a simple distance term Δ_{ij} that is zero if atoms i and j are more than 5 Å apart (and one otherwise).

$$G = \sum_{ij} g_{ij}\Delta_{ij} \qquad [7]$$

A more comprehensive mean-field scoring function was reported recently.[198] Potentials of mean force are derived from a statistical analysis of protein and ligand atom occurrences at certain distances by using 697 protein–ligand complexes taken from the PDB as the knowledge base. Sixteen protein atom types and 34 ligand atom types are defined and lead to 544 PMF interactions of which 282 are statistically significant. The PMF score is calculated as the sum over all protein–ligand atom pair interaction free energies $A_{ij}(r)$ as a function of the atom pair distance r by

$$\text{PMF\_score} = \sum_{\substack{kl \\ r<r^{ij}_{\text{cutoff}}}} A_{ij}(r) \qquad [8]$$

where kl is a ligand–protein atom pair of type ij. r^{ij}_{cutoff} designates the distance at which atom pair interactions are truncated. This cutoff distance has been

found to be optimal at 6 Å for carbon–carbon interactions and 9 Å otherwise. These values appear to be reasonable given the short-range character of hydrophobic interactions compared to, e.g., polar interactions. Note, however, that all $A_{ij}(r)$ are derived with a reference sphere radius of 12 Å. The PMF A_{ij} can be calculated by

$$A_{ij}(r) = -k_B T \ln \left[f^{j}_{\text{Vol\_corr}}(r) \frac{\rho^{ij}_{\text{seg}}(r)}{\rho^{ij}_{\text{bulk}}} \right] \qquad [9]$$

where k_B is the Boltzmann factor, T is the absolute temperature, and $f^{j}_{\text{Vol\_corr}}(r)$ is a ligand volume correction factor that is introduced because intraligand interactions are not accounted for.[198,246] The parameter $\rho^{ij}_{\text{seg}}(r)$ designates the number density of atom pairs of type *ij* at a certain atom pair distance *r*. The parameter $\rho^{ij}_{\text{bulk}}(r)$ is the number density of a ligand–protein atom pair of type *ij* in a reference sphere with radius *R*.[247] This general and fast scoring function was able to reproduce the binding affinities of 77 diverse ligand–protein complexes with a standard deviation of 1.8 log K_i units from measured binding affinities. For use in docking studies, the PMF score is combined with a vdW term to account for short-range interactions.[45] The PMF scoring function was implemented into the DOCK4.0 program. For faster scoring, it was also implemented on a grid similar to the force field score in DOCK. Flexible docking experiments on FK506 binding protein,[45] neuraminidase,[20] and stromelysin[248] showed high predictive power and robustness of the PMF score.

Based on sets of 351 and 188 PDB entries, respectively, two mean-field functions similar to the PMF-score of Muegge and Martin[198] (Eqs. [8, 9]) were derived called BLEEP1 and BLEEP2.[197] In contrast to BLEEP1, BLEEP2 also contains interactions with water. Scoring tests with BLEEP1 and BLEEP2 on 90 protein–ligand complexes from the PDB demonstrate a statistically significant correlation with measured binding constants for BLEEP2 ($r^2 = 0.55$) and somewhat inferior results for BLEEP1 ($r^2 = 0.45$).[249] Standard deviations were not reported but seem to be on the order of less than 2 log K_i units. Differences between the PMF score (Eq. [8]) and BLEEP are small. They consist mainly of different definitions of atoms types, a different reference state (and reference radius) and a different treatment of intraligand interactions. Recent PMF scoring experiments with the BLEEP test set yielded results comparable to those of BLEEP1 and BLEEP2.[246]

Gohlke et al.[195] combined a short distance mean-field term (cutoff 6 Å) with a solvation term based on solvent accessible surface preferences in a knowledge-based function called DrugScore. DrugScore was used to rank ligand placements generated by FlexX for two test sets of 91 and 68 protein–ligand complexes from the PDB. It was found that 75% of the best ranking binding modes could be recognized with rmsd of less than 2 Å from the native binding modes.

Knowledge-based scoring functions have the advantage of being more general than empirical scoring functions. The PDB was used as knowledge base for developing the PMF-score, BLEEP, and DrugScore. Whereas PMF-score and BLEEP use a mean-field term only to score protein–ligand complexes, DrugScore and earlier approaches use combinations of mean-field terms and extra terms, e.g., for solvation. A recent study of reference states used in knowledge-based scoring functions was conducted to analyze the dependence of the spherical reference state on the length of its radius. It was found that a reference state radius of at least 8 Å is needed to implicitly capture solvation effects in the PMF scoring function.[247] The study suggests that short-ranged mean-field potentials used in DrugScore or by Verkhivker et al.[199] need additional terms in their models. Knowledge-based scoring functions seem to be a promising alternative in docking programs. The few docking studies conducted so far with mean-field potential scoring functions as the primary scoring function (that drives the sampling) suggest that they are robust and predictive for finding the correct binding mode of a putative protein–ligand complex[20,248] and for ranking protein–ligand complexes according to their binding affinities.[45,248]

Comparing Scoring Functions in Docking Experiments: Consensus Scoring

As illustrated above, there are plenty of scoring functions available today. Unfortunately, none of them seems to be good enough to be unanimously adopted by developers of docking programs. Indeed, there are probably no two docking programs that use exactly the same scoring function. To make matters worse, comparison studies of scoring functions are not numerous and provide conflicting evidence. For instance, Knegtel et al.[250] analyzed DOCK with chemical scoring, energy scoring, and FlexX predicting the correct binding mode of 32 known thrombin inhibitors by flexible docking. They found that the chemical score outperformed both the energy score and the FlexX score and identified 10–35% of the native-binding conformations with rmsd of less than 2 Å compared to the crystal coordinates. In contrast, Gohlke et al.[195] found that for diverse sets of 91 and 68 protein–ligand complexes from the PDB, the DOCK energy score performed similar to the DrugScore and outperformed the chemical score in identifying correct binding modes. Ha et al.[248] compared FlexX with energy score and PMF-score in DOCK for 61 biphenyl inhibitors of the matrix metalloproteinase stromelysin (MMP-3). It was found that the PMF score outperformed both the energy score and the empirical score in FlexX in identifying the correct binding mode. In fact, the only wrongly predicted binding modes with the DOCK/PMF score were due to sampling insufficiencies rather than scoring problems. Most notably, the PMF score was the only one that found a statistically significant correlation between predicted score and measured binding affinities of

the inhibitor–MMP3 complexes. In two other docking studies, one involving FK506 binding protein[45] and the other neuraminidase,[20] the superiority of PMF scoring over energy scoring in DOCK4 was also reported. Stahl[251] reported that using PMF scoring in database screening experiments is less sensitive to small coordinate changes than is the FlexX score. Finally, a comparison for a variety of sets of protein–ligand complexes from the PDB showed the superiority of PMF scoring over SMoG and Böhm's score.[198]

To better discern between active and inactive inhibitors or ligands in docking experiments, more than one scoring function can be used. Sometimes two different scoring functions are used, one to find the correct binding mode and the other to locally optimize and score the resulting ligand–protein complex. For instance, Wallqvist and Covell[129] used a surface complementarity score to find a limited set of binding modes and then invoked a preference-based free energy surface score to quantify the binding affinity. Gohlke et al.[195] used FlexX to generate a limited number of possible binding modes and employed DrugScore as an "after-score" to identify the correct binding mode. Stahl[251] also relied on FlexX to generate protein–ligand conformations and then evaluated them by using the PMF score. Scoring results can also be improved by introducing filter functions that remove structures with unfavorable properties.[252]

Recently, Charifson et al.[37] presented a comprehensive consensus scoring approach applying two docking approaches and 13 scoring functions. This Vertex group used the energy scoring function of DOCK as primary scoring function along with GAMBLER, a genetic algorithm docker, that employs the piecewise linear potential[112] (PLP) as its primary score. Scoring functions to after-score the binding modes included empirical scoring functions (Böhm's score,[69] ChemScore,[190,234] SCORE,[204] PLP[112]), force fields (Merck molecular force field nonbonded energy,[253] DOCK energy score,[28,64] DOCK chemical score, FLOG score,[34] strain energy) and some other methods (Poisson-Boltzmann,[254] buried lipophilic surface area,[205] DOCK contact score,[62] volume overlap[201]). No knowledge-based scoring functions were included. The docking/scoring study involving three different enzymes showed that by using the intersection of the top scoring compounds of each scoring hit list, significant reductions in the list of false positives in the docking screen could be obtained. The confirmed hit rates of weakly active chemical entities (IC_{50} < 50 μM) were between 2 and 7% (IC_{50} is the concentration of drug required to inhibit a biological or biochemical process by 50%). Consensus scoring is a significant improvement over typical hit rates that are on the order of only 0–3% in virtual screens. By comparing the different scoring functions and combinations thereof, it was found that ChemScore, PLP, and DOCK energy score performed best singly and in consensus.

Simplex minimization during rescoring led to improved docked orientations and scores.[63] Finally, it should be mentioned that a consensus scoring spreadsheet is also available in SYBYL,[255] which combines four scoring

functions including FlexX as the primary score,[29] the PMF score,[198] the DOCK energy score,[64] and the GOLD score[21] as after-scores.

FROM MOLECULAR DOCKING TO VIRTUAL SCREENING

In the previous sections, we introduced algorithms for protein–ligand docking, some of which are fast enough for analyzing several hundred molecules per day per processor. Here, we summarize the tasks beyond the docking step itself that are necessary in virtual screening. The virtual screening process is shown schematically in Figure 5. In general, we can distinguish between four groups of tasks: protein structure preparation, ligand database preparation, the docking calculation itself, and the postprocessing, each of which we now describe in turn.

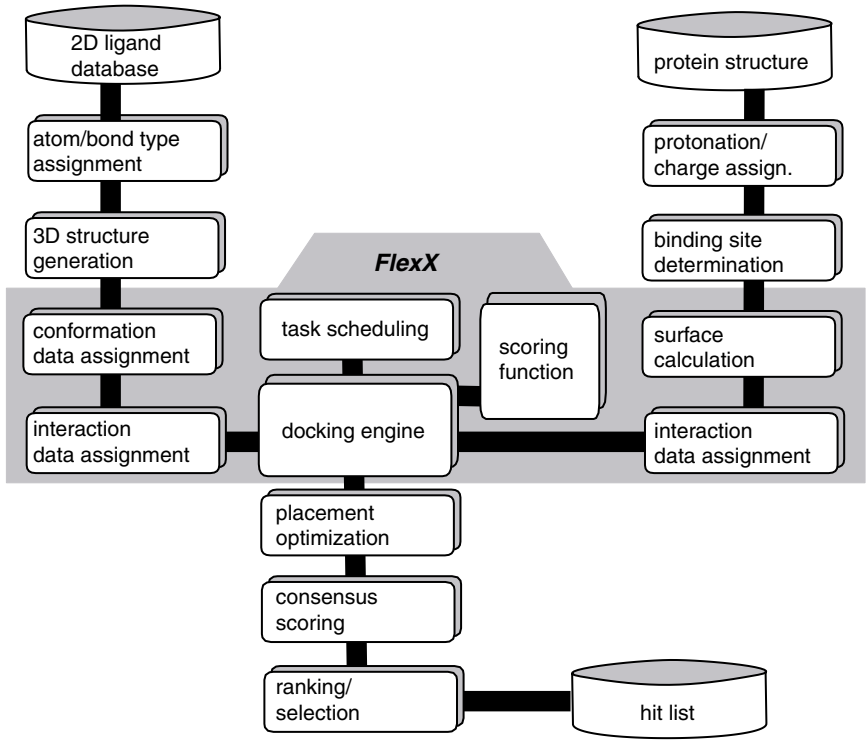

Figure 5 Flow chart of docking in the example of the FlexX program.

Protein Data Preparation

Preparing a protein 3D structure is usually the same for single docking calculations as for virtual screening. Typical steps are the calculation of protonation states and the assignment of atomic charges. The receptor site has to be determined, which is a critical step because a too small receptor site limits the available search space, whereas a larger receptor site costs more computing time. Although algorithms for receptor site determination are available (e.g., LIGSITE[256]), it is generally a good idea to invest time in the manual selection of the binding site because automated methods often fail.[206] Other steps sometimes necessary for docking are the determination of the protein surface atoms and the assignment of interaction data such as chemical coloring of atoms and site points. Whether these steps are necessary depends on the docking software; these steps are internally included in some programs (e.g., in the program FlexX) but have to be done separately in others (e.g., in DOCK).

Ligand Database Preparation

Since we often have to consider several thousand small molecules in a ligand database, manual steps in data preparation must be avoided. Typically, information on available ligands is stored in 2D form (atom types and connectivity) in MDL databases[257] that frequently serve as data repositories for the pharmaceutical industry. Starting from a database of 2D structures, atom and bond types must first be checked and corrected; protonation states and charges have to be assigned. This procedure is typically done with rule-based format translation programs such as db_translate.[255] Next, 3D structures have to be created for the database using programs such as CORINA[258] or CONCORD.[259] Finally, physicochemical data necessary for calculating ligand conformations (e.g., rotational barriers or allowed side-chain rotamers) as well as protein–ligand interactions (e.g., site points that guarantee proper hydrogen-bonding directionality) have to be assigned. Within FlexX, the assignment of physicochemical data is done automatically via a set of so-called static data files containing this information on the basis of subgraphs that could flexibly be matched to the ligand molecule. For example, FlexX contains the MIMUMBA torsional angle database,[94] which contains information about frequently occurring torsional angles for small molecules determined for about a thousand different molecular fragments.

Docking Calculation

The docking calculations are usually iterated separately for each molecule of a database. Since the calculation times per molecule are still quite large even for fast algorithms (about a minute per ligand), the use of a workstation cluster or parallel hardware is of great advantage. Within FlexX, a task

scheduler is employed that distributes docking processes on heterogeneous workstation clusters based on the Parallel Virtual Machine[260] (PVM) software. The scheduler automatically detects whether a parallel-processing environment is available and starts docking processes on a user defined subset of the available computers. The scheduler is very flexible allowing, for example, user initiated interruptions of the calculation, handling of partial and total failure of the computing environment, and recovery of interrupted calculations. In addition, the scheduler contains an automatic output file merging procedure, which is able to handle the output of several hundred docking processes running in parallel. For the user, the whole process is transparent, that is, the resulting output files are identical for calculations performed on a single processor and those performed on a 1024-processor parallel computer. Of course, in the latter case the result is calculated about a thousand-fold faster.

Postprocessing

After the docking calculation, a set of placements and a score are calculated for most of the database compounds. The failure rate of compounds that cannot be placed in the receptor site at all depends on the size of the compounds in the database and on the thresholds set for the maximum scores or the minimum matched features such as DOCK spheres. Typically, the number of unplaced compounds in a database is below 5%. The score can then be used to create an initial ranking of the compounds. Several postprocessing steps can be invoked including the optimization of placements with numerical techniques such as MD or the filtering of placements not fulfilling additional constraints that were omitted by the primary scoring function.[252] Due to the well-known limitations of individual scoring functions, it can be helpful to apply alternative scoring schemes such as the consensus scoring already discussed. If a set of inhibitors with known bioactivity is contained in the database, their docking results can be used to customize the scoring and to estimate how far in rank one should consider molecules for biological testing.

APPLICATIONS

Docking as a Virtual Screening Tool

Biological screening methods are commonly used for lead identification in a drug discovery program. High throughput screening (HTS) is usually employed to experimentally screen hundreds of thousands of compounds against a drug target.[41] Molecular databases can be extended via combinatorial chemistry[39,40] or compound acquisition. Today, proprietary databases often contain about 10^6 compounds. However, this is only a tiny fraction of the conceivable molecular space for which estimates range between 10^{60}

and 10^{100} compounds.[43,44] Computational screening methods can complement HTS because they are fast, cheap, and applicable to even virtual compounds. The calculations can help in deciding which compounds out of the vast pool of possibilities should be made in a combinatorial or medicinal chemistry program. The realm of virtual screening includes techniques that evaluate the drug-likeness[261] of compounds, development parameters,[262] and the 2D and 3D diversity[263,264] of designed libraries.[43,44,265] Virtual screening can even evaluate compound libraries based on the compounds' potential for favorable characteristics in regard to absorption, distribution, metabolism, excretion, and toxicity (ADME/TOX). Here, we want to comment on the role docking plays as usually the last filter in library designs.

Automated docking is often used to rank a library of putative ligand–protein complexes according to their binding affinities. One of the main problems in virtual screening is the large number of false positives, that is, inactive compounds that nevertheless have high scores (bioactive compounds with low scores are called false negatives). To illustrate the problem, we quote Jain[192] who gives the following example: "Suppose that in a database of 100,000 compounds just 20 will bind above some threshold affinity to the target in question. Even if the false negative rate is high (say 50%) one would likely be happy with finding 10 true hits. However, depending on how many compounds may be practically obtained (through purchase or synthesis) and assayed, the false positive rate has a significant impact on the likelihood of finding the true hits. If the false positive rate is just 1%, there will be 1000 false positive 'hits' in the screen. If there is the capacity to obtain and assay 50 of the 1010 hits, with no principled way of distinguishing true positives from false positives, there is a 60% chance of not finding a single true hit. In order to achieve a 90% chance of finding at least one true hit, the absolute false positive rate must be 0.2%."

Fortunately, many virtual screening test studies can identify highly potent, known inhibitors[32] or derivatives thereof[101] when seeded in large databases of nonbinders. However, more realistic applications of virtual screening include the identification of novel micromolar lead structures from a chemical database—a task that is much harder to accomplish due to the small difference in binding affinities between active and inactive compounds. (It should be pointed out by way of comparison that actual compound databases used at most large pharmaceutical companies easily give many micromolar hits for some target receptors, but few or none for other receptors.)

Early examples of academically successful identification of potential lead compounds by virtual screening with DOCK include drug targets such as HIV-1 protease, RNase H, thymidylate synthase, hemagglutinin, and malaria protease.[27] Shoichet et al.[61] discovered a new class of thymidylate synthase inhibitors by screening the Available Chemicals Directory (ACD) and also identified a binding mode different from that of the natural substrate. Grootenhuis et al.[266] identified a potent antitumor agent by virtually screening

10,000 compounds from the Cambridge Structural Database[95] against the crystal structure of the DNA dodecamer–netropsin complex using DOCK. DeJarlais et al.[267] used the same procedure to find haloperidol as a weak HIV-1 protease inhibitor. Antimalaria agents were identified by Ring et al.[268] again using DOCK.

More recent examples of identifying known potent inhibitors by virtual screening include Welch et al.,[32] who identified biotin as the highest scoring entity complexed with streptavidin in a pool of 80,000 compounds from the ACD. Seven of the 13 top ranked hits out of 16,000 ACD compounds were found to be dihydrofolate or methotrexate derivatives in a flexible DOCK screen against dihydrofolate reductase.[101] Kramer et al.[269] screened a library of 556 drug-like compounds from the CSD against 10 different proteins using FlexX. In eight cases, the original inhibitors were identified in the upper 7% of the hit list. Baxter at al.[270] achieved a good separation between known ligands of thrombin, factor Xa, and the estrogen receptor in a virtual screening experiment involving 10,000 drug-like molecules. A comprehensive evaluation of flexible docking was recently performed by Stahl et al.[271] A data subset of ca. 8000 compounds from the World Drug Index[272] was selected containing known ligands for seven different protein targets. The experiment showed that success was highly dependent on the complexity of the docking problem. Complexity was determined by several aspects such as the rigidity and the overall shape of the receptor site, the amount of hydrophobic versus hydrophilic interactions responsible for binding, and the influence of single water molecules to the ligand binding modes. In well-behaved cases like thrombin, between 55% and 90% of ligands known to bind could be identified within the top 10% of the ranked database. This value drops to about 15–40% if protein flexibility and water plays a significant role in binding as for HIV protease and HIV reverse transcriptase.

More recent practical successes in virtual screening include the work of Burkhard et al.[215] who identified a weak thrombin ligand whose binding mode was later verified by crystallography. Toney et al.[273] identified a series of metallo-β-lactamase inhibitors by using FLOG. Different representations of the receptor site hosting two zinc-bound water molecules were used in the docking of 1.6 million compounds with FLOG. Enzyme inhibition was measured for the best scoring 100–500 compounds. A total of 71 compounds of different structural classes with $IC_{50} \leq 20$ μM were identified. In a study of Horvath,[274] rigid docking of 2500 virtual compounds into trypanothione reductase revealed a new class of inhibitors. Recently, Perola et al.[275] reported a virtual screening experiment on the ACD (219,390 compounds) identifying micromolar farnesyltransferase inhibitors.

A virtual screening study was recently conducted to study the ability of DOCK/PMF scoring to find weak ligands to the FK506 binding proteins in a pool of nonbinders.[45] The binding affinities of 3247 small molecules (with an average molecular weight of 191 Da) were measured by NMR. The database

contained 28 active compounds with K_i's ranging between 60 and 2000 μM. Docking/PMF scoring showed a remarkable enrichment of the weakly active compounds in the upper ranks of the computational hit list (Figure 6).

Docking as a Ligand Design Tool

Docking techniques can help suggest binding modes and subsequently lead optimization when lead compounds cannot be cocrystallized with a target protein. Lead optimization based on a protein crystal structure is also often referred to as structure-based design. It is aimed at reaching tighter binding that hopefully results in better efficacy of the putative drug candidate. Recent reviews on how docking in combination with combinatorial chemistry is used in structure-based drug design was given by Kubinyi[158] and Kirkpatrick et al.[276]

The ability of docking programs to work as structure-based ligand design tools is often tested on small sets of known protein–ligand complexes. This evaluation bears the risk of a hidden bias in the data set and that only a small fraction of problem classes is covered. It is therefore necessary to evaluate the method on larger sets containing at least about a hundred test cases. Jones et al.[21] first published an evaluation of the genetic docking algorithm GOLD on more than 100 test cases selected from the PDB. They showed that GOLD was able to predict about 71% of the cases in agreement with the crystal structure. For FlexX, a similar evaluation was published for about 200 protein–ligand complexes achieving about 70% correct predictions within 2 Å.[269] In 46% of the cases, the correct predictions ranked highest among

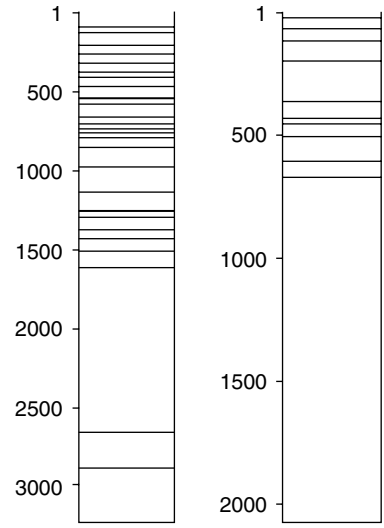

Figure 6 Ranking list of 3247 compounds with an average molecular weight (MW) of 191 Da (left) and a subset of 2077 compounds with MW $\leq$210 Da (right) according to their best scores in a DOCK/PMF run against FK506 binding protein. Computational ranks for the active compounds in the database are marked with horizontal bars. Data are taken from Muegge et al. (Ref. 45). Bold bars depict close neighboring bars.

sampled binding modes. When interpreting these good numbers, one has to keep in mind that the docking calculations were performed with the respective cocrystallized protein structure for each of the ligand–protein complexes. Depending on the protein flexibility, the performance is often lower for practical applications.[20] One method for analyzing this effect is the cross-docking experiment. Here, the ligands of several complexes with the same protein are docked in turn in each cocrystallized protein structures.[20,233,269]

The ultimate test for docking calculations is a blind prediction. A series of blind docking/scoring experiments called "Critical Assessment of Techniques for Free Energy Evaluation" (CATFEE) has been devised recently by Roitberg, Cachau, and Fidelis.[277] A meeting to analyze the results is planned for 2001. In the fourth experiment of the Second Meeting on the Critical Assessment of Techniques for Protein Structure Prediction (CASP2)[278], protein–ligand complexes for seven test cases with unpublished complex structures were predicted with several different docking algorithms. The experiment showed that the docking methods available today are not very robust.[206] For a fast docking method developed for screening purposes like FlexX, only about half the cases were predicted at least partially correct (about 2.5 Å rmsd).[30] Looking at the geometrical problem alone, a placement within 2.5 Å rmsd from the crystallographic ligand placement in the protein structure could be predicted for five of six cases (one case was a covalently bound ligand); however, the ranks of the crystallographic placements were in the range from 2 to 81 among scored binding modes. Since scoring functions are not consistent in identifying the correct binding mode of a ligand–protein complex scoring highest among sampled placements during a fully automated docking run, structure-based lead optimization often includes manual steps. The choice of the appropriate binding mode is usually made by an educated guess backed by energy considerations or topographical preferences calculated (e.g., using the program GRID).[116]

Small fragments can be computationally placed in different binding pockets and later linked together following the idea of "SAR by NMR."[14] For instance, Maduskule et al.[279] discovered a 34-nM inhibitor for factor Xa by manually docking fragments in the *S*1 and *S*4 subsites and connecting them with a tether. Automated docking can help, however, to overcome the human bias in placing the ligand fragments in the receptor site and to provide new suggestions including unanticipated binding modes of some ligands. Investigation of unanticipated docking orientations is an important capability of docking programs because not all ligands that bind to the same receptor align the same way. Cocrystallization experiments of related compounds have shown placements of similar ligands that may not be obvious. To illustrate the capabilities and pitfalls of manual and automated docking in lead optimization, we examine here the case of influenza virus neuraminidase.

Rational drug design led to the design of new inhibitors of neuraminidase—an enzyme important to replication of influenza virus.[280] The

glycoprotein cleaves sialic acid residues from the surface of progeny virus particles. Otherwise, virus aggregation occurs and spread of viral particles is stopped. Inhibiting cleavage thus interferes with the viral replication cycle. Eleven residues form the highly conserved binding site that is also highly polar containing three arginine residues and four glutamic acids. Design was started from the substrate (sialic acid) itself: the pyranose ring was first replaced by a dihydropyran (the compound labeled DANA in Figure 7). This ring modification led to a 133-fold increase in binding affinity. Replacing the 4-OH group by a guanidino group yielded an additional 100,000-fold increase in binding affinity (GANA). Due to rapid clearance of these neuraminic acid compounds from the body, new lead classes had to be considered. Structural and mechanistic studies suggested that the core structure of the neuraminic acids did not contribute to binding.[281] Therefore, a new class of benzoic acid compounds containing the same R groups as the neuraminic acid series was investigated.[282] To the surprise of the investigators, the replacement of the hydroxyl group by a guanidino group in the benzoic acids series (BANA113 in Figure 7) resulted in only a 1000-fold increase in binding affinity. Subsequent crystallographic investigations revealed a completely unexpected binding mode for BANA113.[23] Although the N-acetyl group and the carboxylic acid were setting in the anticipated sites, the benzoic acid ring was flipped by 180° placing the guanidino group into a pocket different from that taken by the guanidino group of the highly potent GANA. Electrostatic calculations of binding energies using the discretized continuum method implemented in DelPhi,[283,284]

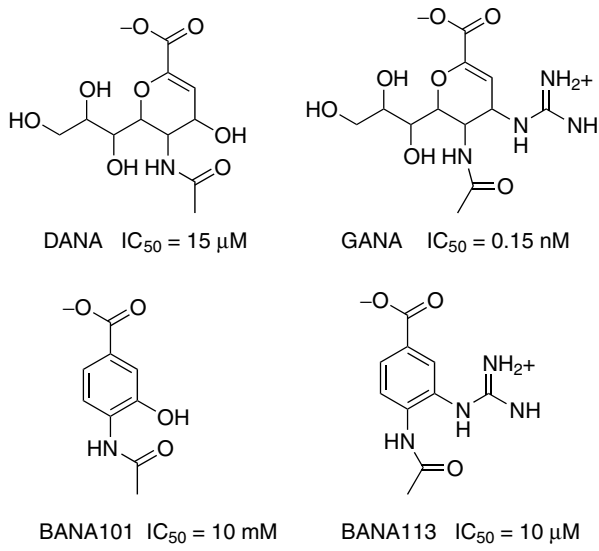

Figure 7 Chemical structures of neuraminidase inhibitors.

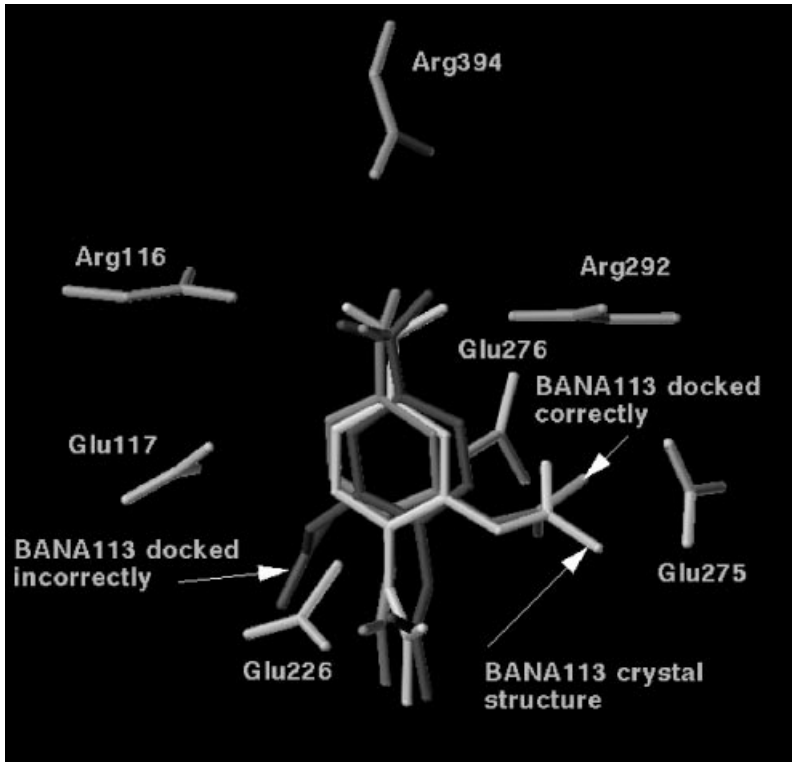

Figure 8 Crystal structure of BANA113 bound to influenza virus neuraminidase. Docked conformations ranking first (correct) and eleventh (incorrect) in a DOCK/PMF score docking experiment are compared to the crystallographic binding mode.

predicted the incorrect (unflipped) binding mode more than two orders of magnitude (3.2 kcal/mol) more favorable than the crystallographic one.[285]

In contrast, automated docking experiments using DOCK with energy score and PMF scoring showed that the correct binding mode of the neuraminidase inhibitor could have been predicted based on the crystal structures available at the time (Figure 8).[20] This example shows how a docking program unbiased from human guessing could provide insight into a lead optimization problem. Although the reliability of docking methods including automated docking approaches is not so high, docking methods can provide new suggestions for protein–ligand interactions that otherwise may have been overlooked.

Although probabilistic predictions are acceptable in virtual screening experiments for lead identification, in lead optimization studies, where only a limited number of compounds is examined, confidence in docked binding modes must be very high. Therefore, the conclusion from the neuraminidase

example is that one should use all techniques available at the time, including automated docking, in order to gather ideas about possible protein–ligand interactions and then examine all the results before making design decisions. This comprehensive strategy is particularly important in light of the fact that it is not uncommon for similar ligands to bind in different binding modes to the same protein.[173]

CONCLUDING REMARKS

A wide range of different methods for protein–ligand docking is available today. Although the docking of small and rigid ligands in their bound conformations is often successful, the docking of flexible drug-like ligands into a rigid binding site of a protein already presents a problem that is not generally solved. Protein flexibility, solvent molecules, and increasing ligand size introduce additional difficulties. Fast methods for the conformational sampling of binding modes of protein–ligand complexes have been developed. Programs typically sample large numbers of different low-energy configurations in a few minutes of computer time. The main challenge of docking programs today is to recognize the correct binding mode once it is sampled. Scoring functions are designed to facilitate this task. They are the actual Achilles' heel of docking programs. Scoring functions are not reliable enough today. However, progress has been made recently in the development of empirical and knowledge-based scoring functions as well as by the addition of solvation terms into force field scores.

Docking is the essence of in silico (virtual) screening tools. Fast and simple functions are employed that filter compounds out of a database that have no chance to fit the desired target. This filtering can be done, for example, by pharmacophore-like feature matching in the binding site. For the survivors of the initial screening, a more elaborate docking scheme is applied. The relative ranking of different protein–ligand structures according to their binding affinities is the crucial problem in a virtual screening application. The scoring problem is even more challenging in ranking protein–ligand complexes than in finding the correct binding mode of a single ligand–protein complex. Typically, scoring results in large numbers of false positives in a virtual screen. False positive rates can be reduced by using several scoring functions in a consensus-scoring strategy.[37] The ultimate goal of virtual screening is to mimic experimental high throughput screening results. For virtual screening to be competitive, the ratio between real hits and false positives must be pushed above 10%. If a database of 10^6 compounds contains 10 actives, the number of false positives must be below 100 (i.e., 0.01% of the total database). Although this goal is not reached today, docking as a virtual screening tool can still be useful in drug design projects. Virtual screening is fast and cheap because there is no need to develop a screening assay as well as no need to

purchase, synthesize, or store the compounds to be screened. Therefore, virtual screening can be beneficial as an initial filter in cases where a full high-throughput screening cannot be conducted, for either technical or cost reasons. On the other hand, virtual screening requires the existence of an appropriate 3D structure of the target molecule.

In ligand design, the strength of protein–ligand docking methods lies in the quick creation of a variety of plausible binding modes for a putative protein–ligand complex. Computationally proposed binding modes can be of value in suggesting modifications of a lead structure. These suggestions can then be tested experimentally. Sometimes, unexpected binding modes of protein–ligand complexes are predicted that help to solve lead optimization puzzles and guide medicinal chemistry efforts in a drug discovery program.

The technical revolution in high-speed chemical synthesis and biological high throughput screening techniques has somewhat outpaced computational approaches in recent years. One could argue that it does not make much sense to elaborately calculate binding affinities for a small number of compounds if they can be made and tested experimentally much faster. Therefore, the main opportunity for docking techniques today is to suggest new compounds that would not have been identified in HTS or synthesized as part of an ongoing SAR exploration. The need to generate novel chemical matter for a drug target is apparent if one looks at the surprisingly low success rate of HTS throughout the pharmaceutical industry. Only one in about ten HTS runs produces an acceptable lead that is optimized in a drug discovery program.[286] Moreover, for many proteins (e.g., orphan receptors), HTS cannot be developed due to lack of reference compounds. It is now recognized that HTS cannot be the sole source of lead compounds in the pharmaceutical industry. Rational design techniques, which did not enjoy undivided support in the pharmaceutical industry in prior decades, have flourished recently. Small molecule docking has contributed much to the recent successes of rational drug design. Today virtual screening belongs to the standard repertoire of many pharmaceutical research operations. Some pharmaceutical companies such as Vertex almost exclusively prosecute drug targets for which they can generate crystal structures.[287] Virtual screening and docking for lead optimization are used as main tools. As a result, the average number of compounds made in a drug optimization program at Vertex is an order of magnitude smaller than at other pharmaceutical companies. The fact that today only one in about 29,000 small molecules synthesized in the pharmaceutical industry results in a profitable drug[288] illustrates the drastic potential that rational drug design can have on cost reduction in the drug discovery.

Current efforts are directed toward improving the reliability of scoring functions as well as increased docking speed.[289] High throughput crystallography in combination with fast and reliable docking and scoring algorithms would cause a revolution in drug design. From our perspective, there are high expectations of docking/scoring approaches for the future.

ACKNOWLEDGMENTS

We would like to thank the Editors Kenny Lipkowitz and Don Boyd for carefully editing the manuscript. Also, we would like to thank Thomas Lengauer (GMD) and Martin Stahl (F. Hoffmann-La Roche AG, Basel) for helpful suggestions and discussions.

REFERENCES

1. A. T. Brünger, J. Kuryan, and M. Karplus, *Science,* **235,** 458 (1987). Crystallographic R Factor Refinement by Molecular Dynamics.
2. S. Subbiah, *Science,* **252,** 128 (1991). Low-Resolution Real Space Envelopes: An Approach to the Ab Initio Macromolecular Phase Problem.
3. W. Kühlbrandt, *Q. Rev. Biophys.,* **25,** 1 (1992). Two-Dimensional Crystallization of Membrane Proteins.
4. A. T. Brünger, *Nature* (*London*), **355,** 472 (1992). A Novel Statistical Quantity for Assessing the Accuracy of Crystal Structures.
5. J. L. Markley, *Methods Enzymol.,* **176,** 12 (1989). Two-Dimensional Nuclear Magnetic Resonance Spectroscopy of Proteins: An Overview.
6. G. M. Clore and A. M. Gronenborn, *Science,* **252,** 1390 (1991). Structures of Larger Proteins in Solution: Three- and Four-Dimensional Heteronuclear NMR Spectroscopy.
7. S. W. Fesik, *J. Med. Chem.,* **34,** 2937 (1991). NMR Studies of Molecular Complexes as a Tool in Drug Design.
8. A. Bax and S. Grzesiek, *Acc. Chem. Res.,* **26,** 131 (1993). Methodological Advances in Protein NMR.
9. 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, *Eur. J. Biochem.,* **80,** 319 (1977). The Protein Data Bank. A Computer-Based Archival File for Macromolecular Structures.
10. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, *Nucleic Acids Res.,* **28,** 235 (2000). The Protein Data Bank.
11. T. P. Hubbard, A. M. Lesk, and A. Tramontano, *Nature Struct. Biol.,* **3,** 313 (1996). Gathering Them in to the Fold.
12. URL http://www.rcsb.org/pdb.
13. V. L. Nienaber, P. L. Richardson, V. Klighofer, J. J. Bouska, V. L. Giranda, and J. Greer, *Nature Biotech.,* **18,** 1105 (2000). Discovering Novel Ligands for Macromolecules Using X-Ray Crystallography Screening.
14. S. B. Shuker, P. J. Hajduk, R. P. Meadows, and S. W. Fesik, *Science,* **274,** 1531 (1996). Discovering High-Affinity Ligands for Protein SAR by NMR.
15. J. M. Blaney and J. S. Dixon, *Perspect. Drug Disc. Des.,* **15,** 301 (1993). A Good Ligand Is Hard to Find: Automatic Docking Methods.
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. C. Sander and R. Schneider, *Proteins: Struct., Funct., Genet.,* **9,** 56 (1991). Database of Homology-Derived Protein Structures and the Structural Meaning of Sequence Alignment.
18. T. L. Blundell, B. L. Sibanda, M. J. E. Sternberg, and J. M. Thornton, *Nature* (*London*), **326,** 347 (1987). Knowledge-Based Prediction of Protein Structures and the Design of Novel Molecules.
19. A. Vedani, P. Zbinden, J. P. Snyder, and P. A. Greenidge, *J. Am. Chem. Soc.,* **117,** 4987 (1995). Pseudoreceptor Modeling: The Construction of Three-Dimensional Receptor Surrogates.

20. 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.
21. G. Jones, P. Willett, R. C. Glen, and A. R. Leach, *J. Mol. Biol.,* **267,** 727 (1997). Development and Validation of a Genetic Algorithm for Flexible Docking.
22. K. Sekar, S. Eswaramoorthy, M. K. Jain, and M. Sundaralingam, *Biochemistry,* **36,** 14186 (1997). Crystal Structure of the Complex of Bovine Pancreatic Phospholipase A_2 with the Inhibitor 1-Hexadecyl-3-(trifluoroethyl)-sn-glycero-2-phosphomethanol.
23. E. A. Sudbeck, M. J. Jedrzejas, S. Singh, W. J. Brouillette, G. M. Air, W. G. Laver, Y. S. Babu, S. Bantia, P. Chand, N. Chu, J. A. Montgomery, D. A. Walsh, and M. Luo, *J. Mol. Biol.,* **267,** 584 (1997). Guanidinobenzoic Acid Inhibitors of Influenza Virus Neuraminidase.
24. L. P. Ehrlich and R. C. Wade, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., Wiley-VCH, New York, 2001, Vol. 17, pp. 61–97. Protein–Protein Docking.
25. W. L. Jorgensen, *Science,* **254,** 954 (1991). Rusting of the Lock and Key Model for Protein–Ligand Binding.
26. I. D. Kuntz, J. M. Blaney, S. J. Oatley, and R. L. Langridge, *J. Mol. Biol.,* **161,** 269 (1982). A Geometric Approach to Macromolecule–Ligand Interactions.
27. I. D. Kuntz, *Science,* **257,** 1078 (1992). Structure-Based Strategies for Drug Design and Discovery.
28. 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.
29. 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.
30. B. Kramer, M. Rarey, and T. Lengauer, *Proteins: Struct., Funct., Genet.,* **1,** 221 (1997). CASP2 Experiences with Docking Flexible Ligands Using FlexX.
31. B. Kramer, G. Metz, M. Rarey, and T. Lengauer, *Med. Chem. Res.,* **9,** 463 (1999). Ligand Docking and Screening with FlexX.
32. 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.
33. D. S. Goodsell and A. J. Olson, *Proteins: Struct., Funct., Genet.,* **8,** 195 (1990). Automated Docking of Substrates to Proteins by Simulated Annealing.
34. M. D. Miller, S. K. Kearsley, D. J. Underwood, and R. P. Sheridan, *J. Comput.-Aided Mol. Des.,* **8,** 153 (1994). FLOG: A System to Select Quasi-flexible Ligands Complementary to a Receptor of Known 3-Dimensional Structure.
35. T. I. Oprea and G. R. Marshall, *Perspect. Drug Discovery Des.,* **9/10/11,** 35 (1998). Receptor-Based Prediction of Binding Affinities.
36. H. J. Böhm and M. Stahl, *Med. Chem. Res.,* **9,** 445 (1999). Rapid Empirical Scoring Functions in Virtual Screening Applications.
37. 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.
38. J. R. H. Tame, *J. Comput.-Aided Mol. Des.,* **13,** 99 (1999). Scoring Functions: A View from the Bench.
39. M. A. Gallop, R. W. Barrett, W. J. Dower, S. P. A. Fodor, and A. M. Gordon, *J. Med. Chem.,* **37,** 1233 (1994). Applications of Combinatorial Technologies to Drug Discovery. 1. Background and Peptide Combinatorial Libraries.
40. E. M. Gordon, R. W. Barrett, W. J. Dower, S. P. A. Fodor, and M. A. Gallop, *J. Med. Chem.,* **37,** 1385 (1994). Applications of Combinatorial Technologies to Drug Discovery 2. Combinatorial Organic Synthesis, Library Screening Strategies, and Future Directions.
41. M. W. Lutz, J. A. Menius, T. D. Choi, R. G. Laskody, P. L. Domanico, A. S. Goetz, and D. L. Saussy, *Drug Discovery Today,* **1,** 277 (1996). Experimental Design for High Throughput Screening.

42. R. F. Burns, R. M. A. Simmons, J. J. Howbert, D. C. Waters, P. G. Threlkeld, and B. D. Gitter, in *Exploiting Molecular Diversity,* Symposium Proceedings, Cambridge Healthtech Institute, San Diego, CA, Vol. 2, p. 6, January 23–25, 1995. Virtual Screening as a Tool for Evaluating Chemical Libraries.
43. W. P. Walters, M. T. Stahl, and M. A. Murcko, *Drug Discovery Today,* **3,** 160 (1998). Virtual Screening—an Overview.
44. Y. C. Martin, *Perspect. Drug Discovery Des.,* **7/8,** 159 (1997). Challenges and Prospects for Computational Aids to Molecular Diversity.
45. 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.
46. T. J. Marrone, B. A. Luty, and P. W. Rose, *Perspect. Drug Discovery Des.,* **20,** 209 (2000). Discovering High-Affinity Ligands from the Computationally Predicted Structures and Affinities of Small Molecules Bound to a Target: A Virtual Screening Approach.
47. K. E. B. Platzer, F. A. Momany, and H. A. Scheraga, *Int. J. Peptide Protein Res.,* **4,** 187 (1972). Conformational Energy Calculations of Enzyme–Substrate Interactions. I. Computation of Preferred Conformation of Some Substrates of α-Chymotrypsin.
48. K. E. B. Platzer, F. A. Momany, and H. A. Scheraga, *Int. J. Peptide Protein Res.,* **4,** 201 (1972). Conformational Energy Calculations of Enzyme–Substrate Interactions. II. Computation of the Binding Energy for Substrates in the Active Site of α-Chymotrypsin.
49. W. C. Guida, *Curr. Opin. Struct. Biol.,* **4,** 777 (1994). Software for Structure-Based Drug Design.
50. P. M. Colman, *Curr. Opin. Struct. Biol.,* **4,** 868 (1994). Structure-Based Drug Design.
51. T. P. Lybrand, *Curr. Opin. Struct. Biol.,* **5,** 224 (1995). Ligand–Protein Docking and Rational Drug Design.
52. R. Rosenfeld, S. Vajda, and C. DeLisi, *Ann. Rev. Biophys. Biomol. Struct.,* **24,** 677 (1995). Flexible Docking and Design.
53. H. J. Böhm, *Curr. Opin. Biotech.,* **7,** 433 (1996). Current Computational Tools for De Novo Ligand Design.
54. T. Lengauer and M. Rarey, *Curr. Opin. Struct. Biol.,* **6,** 402 (1996). Computational Methods for Biomolecular Docking.
55. F. S. Kuhl, G. M. Crippen, and D. K. Friesen, *J. Comput. Chem.,* **5,** 24 (1984). A Combinatorial Algorithm for Calculating Ligand Binding.
56. C. Bron and J. Kerbosch, *Commun. Assoc. Computing Machinery,* **16,** 575 (1973). Finding All Cliques of an Undirected Graph.
57. F. M. Richards, *Ann. Rev. Biophys. Bioeng.,* **6,** 151 (1977). Areas, Volumes, Packing, and Protein Structure.
58. M. L. Connolly, *J. Appl. Crystallogr.,* **16,** 548 (1983). Analytical Molecular Surface Calculation.
59. M. L. Connolly, *J. Appl. Crystallogr.,* **18,** 499 (1985). Molecular Surface Triangulation.
60. E. E. Kim, C. T. Baker, M. D. Dwyer, M. A. Murcko, B. G. Rao, R. D. Tung, and M. A. Navia, *J. Am. Chem. Soc.,* **117,** 1181 (1995). Crystal Structure of HIV-1 Protease in Complex with VX-478, A Potent and Orally Bioavailable Inhibitor of the Enzyme.
61. B. K. Shoichet, R. M. Stroud, D. V. Santi, I. D. Kuntz, and K. M. Perry, *Science,* **259,** 1445 (1993). Structure-Based Discovery of Inhibitors of Thymidylate Synthase.
62. B. K. Shoichet, D. L. Bodian, and I. D. Kuntz, *J. Comput. Chem.,* **13,** 380 (1992). Molecular Docking Using Shape Descriptors.
63. 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.
64. E. C. Meng, B. K. Shoichet, and I. D. Kuntz, *J. Comput. Chem.,* **13,** 505 (1992). Automated Docking with Grid-Based Energy Evaluation.

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

66. D. A. Gschwend and I. D. Kuntz, *J. Comput.-Aided Mol. Des.,* **10,** 123 (1996). Orientational Sampling and Rigid-Body Minimization in Molecular Docking Revisited—On-the-Fly Optimization and Degeneracy Removal.

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

68. X. Q. Zou, Y. X. 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.

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

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

71. M. C. Lawrence and P. C. Davis, *Proteins: Struct., Funct., Genet.,* **12,** 31 (1992). CLIX: A Search Algorithm for Finding Novel Ligands Capable of Binding Proteins of Known Three-Dimensional Structure.

72. M. Y. Mizutani, N. Tomioka, and A. Itai, *J. Mol. Biol.,* **243,** 310 (1994). Rational Automatic Search Method for Stable Docking Models of Protein and Ligand.

73. Y. Lamdan and H. J. Wolfson, in *Proceedings of the IEEE International Conference on Computer Vision,* 1988, pp. 238–249. Geometric Hashing: A General and Efficient Model-Based Recognition Scheme.

74. D. Fischer, R. Norel, H. Wolfson, and R. Nussinov, *Proteins: Struct., Funct., Genet.,* **16,** 278 (1993). Surface Motifs by a Computer Vision Technique: Searches, Detection and Implications for Protein–Ligand Recognition.

75. D. Fischer, S. L. Lin, H. L. Wolfson, and R. Nussinov, *J. Mol. Biol.,* **248,** 459 (1995). A Geometry-Based Suite of Molecular Docking Processes.

76. S. Linnainmaa, D. Harwood, and L. S. Davis, *IEEE, Transactions on Pattern Analysis and Machine Intelligence,* **10,** 634 (1988). Pose Determination of a Three-Dimensional Object Using Triangle Pairs.

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

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

79. T. I. Oprea, *J. Comput.-Aided Mol. Des.,* **14,** 251 (2000). Property Distribution of Drug-Related Chemical Databases.

80. S. K. Kearsley, D. J. Underwood, R. P. Sheridan, and M. D. Miller, *J. Comput.-Aided Mol. Des.,* **8,** 153 (1994). Flexibases: A Way to Enhance the Use of Molecular Docking Methods.

81. T. F. Havel, I. D. Kuntz, and G. M. Crippen, *J. Theoret. Biol.,* **104,** 359 (1983). The Combinatorial Distance Geometry Approach to the Calculation of Molecular Conformation. 1. A New Approach to an Old Problem.

82. T. F. Havel, I. D. Kuntz, and G. M. Crippen, *Bull. Math. Biol.,* **45,** 665 (1983). The Theory and Practice of Distance Geometry.

83. G. M. Crippen and T. F. Havel, *Distance Geometry and Molecular Conformation,* Research Studies Press, Taunton, UK, 1988.

84. D. M. Lorber and B. K. Shoichet, *Protein Sci.,* **7,** 938 (1998). Flexible Ligand Docking Using Conformational Ensembles.

85. 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.

86. R. L. DesJarlais, R. P. Sheridan, J. S. Dixon, I. D. Kuntz, and R. Venkataraghavan, *J. Med. Chem.,* **29,** 2149 (1986). Docking Flexible Ligands to Macromolecular Receptors by Molecular Shape.
87. B. Sandak, R. Nussinov, and H. J. Wolfson, *IEEE Workshop on Shape and Pattern Matching in Computational Biology,* 41 (1994). 3-D Flexible Docking of Molecules.
88. B. Sandak, R. Nussinov, and H. J. Wolfson, *Computer Applications in the Biosciences,* **11,** 87 (1995). An Automated Computer Vision and Robotics-Based Technique For 3-D Flexible Biomolecular Docking and Matching.
89. B. Sandak, R. Nussinov, and H. J. Wolfson, *J. Comput. Biol.,* **5,** 631 (1998). A Method for Biomolecular Structural Recognition and Docking Allowing Conformational Flexibility.
90. J. B. Moon and W. J. Howe, *Proteins: Struct., Funct., Genet.,* **11,** 314 (1991). Computer Design of Bioactive Molecules: A Method for Receptor-Based De Novo Ligand Design.
91. A. R. Leach and I. D. Kuntz, *J. Comput. Chem.,* **13,** 730 (1992). Conformational Analysis of Flexible Ligands in Macromolecular Receptor Sites.
92. M. Rarey, B. Kramer, and T. Lengauer, in *Proceedings of the Third International Conference on Intelligent Systems in Molecular Biology,* C. Rawlings, Ed., AAAI Press, Menlo Park, CA, 1995, pp. 300–308. Time-Efficient Docking of Flexible Ligands into Active Sites of Proteins.
93. M. Rarey, B. Kramer, and T. Lengauer, *J. Comput.-Aided Mol. Des.,* **11,** 369 (1997). Multiple Automatic Base Selection—Protein–Ligand Docking Based on Incremental Construction Without Manual Intervention.
94. G. Klebe and T. Mietzner, *J. Comput.-Aided Mol. Des.,* **8,** 583 (1994). A Fast and Efficient Method to Generate Biologically Relevant Conformations.
95. F. H. Allen, S. Bellard, M. D. Brice, B. A. Cartwright, A. Doubleday, H. Higgs, T. Hummelink-Peters, O. Kennard, W. D. S. Motherwell, J. R. Rodgers, and D. G. Watson, *Acta Crystallogr.,* **B35,** 2331 (1979). The Cambridge Crystallographic Data Center: Computer-Based Search, Retrieval, Analysis and Display of Information.
96. J. Gasteiger, C. Rudolph, and J. Sadowski, *Tetrahedron Computer Methodology,* **3,** 537 (1990). Automatic Generation of 3D-Atomic Coordinates for Organic Molecules.
97. J. Sadowski, J. Gasteiger, and G. Klebe, *J. Chem. Inf. Comput. Sci.,* **34,** 1000 (1994). Comparison of Automatic Three-Dimensional Model Builders Using 639 X-Ray Structures.
98. H.-J. Böhm, *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.
99. M. Rarey, B. Kramer, and T. Lengauer, *Bioinformatics,* **15,** 243 (1999). Docking of Hydrophobic Ligands with Interaction-Based Matching Algorithms.
100. M. Rarey, B. Kramer, and T. Lengauer, *Proteins: Struct., Funct., Genet.,* **34,** 17 (1999). The Particle Concept: Placing Discrete Water Molecules During Protein–Ligand Docking Predictions.
101. S. Makino and I. D. Kuntz, *J. Comput. Chem.,* **18,** 1812 (1997). Automated Flexible Ligand Docking Method and Its Application for Database Search.
102. 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.
103. S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, *J. Comput. Chem.,* **7,** 230 (1986). New Force Field for Simulations of Proteins and Nucleic Acids.
104. D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman, *Comput. Phys. Commun.,* **91,** 1 (1995). AMBER, a Package of Computer Programs for Applying Molecular Mechanics, Normal Mode Analysis, Molecular Dynamics and Free Energy Calculations to Simulate the Structural and Energetic Properties of Molecules.
105. G. Jones, P. Willett, and R. C. Glen, *J. Comput.-Aided Mol. Des.,* **9,** 532 (1995). A Genetic Algorithm for Flexible Molecular Overlay and Pharmacophore Elucidation.

106. C. M. Oshiro, I. D. Kuntz, and J. S. Dixon, *J. Comput.-Aided Mol. Des.,* **9,** 113 (1995). Flexible Ligand Docking Using a Genetic Algorithm.

107. K. P. Clark and Ajay, *J. Comput. Chem.,* **16,** 1210 (1995). Flexible Ligand Docking Without Parameter Adjustment Across Four Ligand–Receptor Complexes.

108. D. K. Gehlhaar, G. M. Verkhivker, P. A. Rejto, D. B. Fogel, L. J. Fogel, and S. T. Freer, in *Proceedings of the Fourth Annual Conference on Evolutionary Programming,* J. R. McDonnell, R. G. Reynolds, and D. B. Fogel, Eds., MIT Press, Cambridge, MA, 1995, pp. 615–627. Docking Conformationally Flexible Small Molecules into a Protein Binding Site Through Evolutionary Programming.

109. G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, and A. J. Olson, *J. Comput. Chem.,* **19,** 1639 (1998). Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function.

110. D. E. Goldberg, *Genetic Algorithms in Search Optimization and Machine Learning,* Addison-Wesley, Reading, MA, 1989.

111. R. Judson, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1997, Vol. 10, pp. 1–73. Genetic Algorithms and Their Use in Chemistry.

112. 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.

113. S. Kirkpatrik, C. D. J. Gelatt, and M. P. Vecchi, *Science,* **220,** 671 (1983). Optimization by Simulated Annealing.

114. G. M. Morris, D. S. Goodsell, R. Huey, and A. J. Olson, *J. Comput.-Aided Mol. Des.,* **10,** 293 (1996). Distributed Automated Docking of Flexible Ligands to Proteins: Parallel Applications of AutoDock 2.4.

115. D. S. Goodsell, G. M. Morris, and A. J. Olson, *J. Mol. Recognit.,* **9,** 1 (1996). Automated Docking of Flexible Ligands: Applications of AutoDock.

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

117. S. Yue, *Protein Eng.,* **4,** 177 (1990). Distance-Constrained Molecular Docking by Simulated Annealing.

118. T. P. Lybrand, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1990, Vol. 1, pp. 295–320. Computer Simulation of Biomolecular Systems Using Molecular Dynamics and Free Energy Perturbation Methods.

119. A. R. Leach, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1991, Vol. 2, pp. 1–55. A Survey of Methods for Searching the Conformational Space of Small and Medium-Sized Molecules.

120. U. Dinur and A. T. Hagler, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1991, Vol. 2, pp. 99–164. New Approaches to Empirical Force Fields.

121. T. P. Straatsma, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1996, Vol. 9, pp. 81–127. Free Energy by Molecular Simulation.

122. I. Pettersson and T. Liljefors, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1996, Vol. 9, pp. 167–189. Molecular Mechanics Calculated Conformational Energies of Organic Molecules: A Comparison of Force Fields.

123. 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 Simulation.

124. A. Di Nola, D. Roccatano, and H. Berendsen, *Proteins: Struct., Funct., Genet.,* **19,** 174 (1994). Molecular Dynamics Simulation of the Docking of Substrates to Proteins.

125. R. Mangoni, D. Roccatano, and A. Di Nola, *Proteins: Struct., Funct., Genet.,* **35,** 153 (1999). Docking of Flexible Ligands to Flexible Receptors in Solution by Molecular Dynamics Simulation.

126. J. A. Given and M. K. Gilson, *Proteins: Struct., Funct., Genet.*, **33,** 475 (1998). A Hierarchical Method for Generating Low-Energy Conformers of a Protein–Ligand Complex.

127. T. N. Hart and R. J. Read, *Proteins: Struct., Funct., Genet.*, **13,** 206 (1992). A Multiple-Start Monte Carlo Docking Method.

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

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

130. R. Abagyan, M. Totrov, 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.

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

132. J. D. Madura, M. E. Davis, M. K. Gilson, R. C. Wade, B. A. Luty, and J. A. McCammon, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1994, Vol. 5, pp. 229–267. Biological Applications of Electrostatic Calculations and Brownian Dynamics Simulations.

133. J. Y. Trosset and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA,* **95,** 8011 (1998). Reaching the Global Minimum in Docking Simulations: A Monte Carlo Energy Minimization Approach Using Bezier Splines.

134. J. Y. Trosset and H. A. Scheraga, *J. Comput. Chem.,* **20,** 412 (1999). PRODOCK: Software Package for Protein Modeling and Docking.

135. J. Y. Trosset and H. A. Scheraga, *J. Comput. Chem.,* **20,** 244 (1999). Flexible Docking Simulations: Scaled Collective Variable Monte Carlo Minimization Approach Using Bezier Splines, and Comparison with a Standard Monte Carlo Algorithm.

136. J. Y. Trosset and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.* **95,** 8011 (1998). Reaching the Global Minimum in Docking Simulations: A Monte Carlo Energy Minimization Approach Using Bezier Splines.

137. A. E. Torda and W. F. van Gunsteren, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1992, Vol. 3, pp. 143–172. Molecular Modeling Using NMR Data.

138. J. M. Blaney and J. S. Dixon, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1994, Vol. 5, pp. 299–335. Distance Geometry in Molecular Modeling.

139. A. Dress and T. F. Havel, *Discrete Applied Mathematics,* **19,** 129 (1988). Shortest Path Problems and Molecular Conformation.

140. P. L. Easthope and T. F. Havel, *Bull. Math. Biol.,* **51,** 173 (1989). Computational Experience with an Algorithm for Tetrangle Inequality Bound Smoothing.

141. A. K. Ghose and G. M. Crippen, *J. Comput. Chem.,* **6,** 350 (1985). Geometrically Feasible Binding Modes of a Flexible Ligand Molecule at the Receptor Site.

142. M. Billeter, T. F. Havel, and I. D. Kuntz, *Biopolymers,* **26,** 777 (1987). A New Approach to the Problem of Docking Two Molecules: The Ellipsoid Algorithm.

143. A. S. Smellie, G. M. Crippen, and W. G. Richards, *J. Chem. Inf. Comput. Sci.,* **31,** 386 (1991). Fast Drug-Receptor Mapping by Site-Directed Distances: A Novel Method of Predicting New Pharmacological Leads.

144. V. Sobolev, R. C. Wade, G. Vriend, and M. Edelman, *Proteins: Struct., Funct., Genet.,* **25,** 120 (1996). Molecular Docking Using Surface Complementarity.

145. 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.

146. J. Wang, P. A. Kollman, and I. D. Kuntz, *Proteins: Struct., Funct., Genet.,* **36,** 1 (1999). Flexible Ligand Docking: A Multistep Strategy Approach.

147. D. Hoffmann, B. Kramer, T. Washio, T. Steinmetzer, M. Rarey, and T. Lengauer, *J. Med. Chem.,* **42,** 4422 (1999). Two-Stage Method for Protein–Ligand Docking.

148. CHARMm22, Molecular Simulations Inc./Accelrys, San Diego, CA. URL www.msi.com.

149. 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.

150. D. Hoffmann, T. Washio, K. Gessler, and J. Jacob, in *Proceedings of the Workshop on Monte Carlo Approach to Biopolymers and Protein Folding,* P. Grassberger, G. Barkema, and W. Nadler, Eds., World Scientific, Singapore, 1998, pp. 153–170. Tackling Concrete Problems in Molecular Biophysics Using Monte Carlo and Related Methods: Glycosylation, Folding, Solvation.

151. B. A. Luty, Z. R. Wasserman, P. Stouten, C. N. Hodge, M. Zacharias, and J. A. McCammon, *J. Comput. Chem.,* **16,** 454 (1995). A Molecular Mechanics/Grid Method for Evaluation of Ligand-Receptor Interactions.

152. Z. R. Wasserman and C. N. Hodge, *Proteins: Struct., Funct., Genet.,* **24,** 227 (1996). Fitting an Inhibitor into the Active Site of Thermolysin: A Molecular Dynamics Case Study.

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

154. N. J. Nielsson, *Principles of Artificial Intelligence,* Springer Verlag, Berlin, 1982.

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

156. 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.

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

158. H. Kubinyi, *Curr. Opin. Drug Discov. Devel.,* **1,** 16 (1998). Combinatorial and Computational Approaches in Structure-Based Drug Design.

159. M. Rarey, B. Kramer, C. Bernd, and T. Lengauer, in *Biocomputing: Proceedings of the 1996 Pacific Symposium,* L. Hunter and T. Klein, Eds., World Scientific Publishing Co, Singapore, 1996, Time-Efficient Docking of Similar Flexible Ligands. URL http://www.cgl.ucsf.edu/psb/psb96/proceedings/eproceedings.html.

160. S. Makino and I. D. Kuntz, *J. Comput. Chem.,* **19,** 1834 (1998). ELECT++: Faster Conformational Search Method for Docking Flexible Molecules Using Molecular Similarity.

161. C. W. Murray, D. E. Clark, T. R. Auton, M. A. Firth, J. Li, R. A. Sykes, B. Waszkowycz, D. R. Westhead, and S. C. Young, *J. Comput.-Aided Mol. Des.,* **11,** 193 (1997). PRO_SELECT: Combining Structure-Based Drug Design and Combinatorial Chemistry for Rapid Lead Discovery. 1. Technology.

162. Y. Sun, T. J. A. Ewing, A. G. Skillman, and I. D. Kuntz, *J. Comput.-Aided Mol. Des.,* **12,** 597 (1998). CombiDOCK: Structure-Based Combinatorial Docking and Library Design.

163. M. Rarey and T. Lengauer, *Perspect. Drug Disc. Des.,* **20,** 63 (2000). A Recursive Algorithm for Efficient Combinatorial Library Docking.

164. E. K. Kick, D. C. Roe, A. G. Skillman, G. C. 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.

165. D. C. Roe and I. D. Kuntz, *J. Comput.-Aided Mol. Des.,* **9,** 269 (1995). Builder V2—Improving the Chemistry of a De Novo Design Strategy.

166. H. J. Böhm, *J. Comput.-Aided Mol. Des.,* **8,** 623 (1994). On the Use of LUDI to Search the Fine Chemicals Directory for Ligands of Proteins of Known Three-Dimensional Structure.

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

168. A. Caflisch, *J. Comput.-Aided Mol. Des.,* **10,** 372 (1996). Computational Combinatorial Ligand Design: Application to Human Alpha-Thrombin.

169. A. Miranker and M. Karplus, *Proteins: Struct., Funct., Genet.,* **11,** 29 (1991). Functionality Maps of Binding Sites: A Multiple Copy Simultaneous Search Method.

170. S. Makino, T. J. A. Ewing, and I. D. Kuntz, *J. Comput.-Aided Mol. Des.,* **13,** 513 (1999). DREAM++: Flexible Docking Program for Virtual Combinatorial Libraries.

171. J. E. Ladbury and B. Z. Chowdhry, *Chem. Biol.,* **3,** 791 (1996). Sensing the Heat: The Application of Isothermal Titration Calorimetry to Thermodynamic Studies of Biomolecular Interactions.

172. P. B. Garland, *Q. Rev. Biophys.,* **29,** 91 (1996). Optical Evanescent Wave Methods for the Study of Biomolecular Interactions.

173. G. Farber, *Pharmacol. Ther.,* **84,** 327 (1999). New Approaches to Rational Drug Design.

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

175. 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. Implications for Theoretical and Experimental Studies.

176. A. Warshel, A. Papazyan, and I. Muegge, *J. Biol. Inorg. Chem.,* **2,** 143 (1997). Microscopic and Semimacroscopic Redox Calculations—What Can and Cannot Be Learned from Continuum Models?

177. A. Warshel and J. Aqvist, *Annu. Rev. Biophys. Biophys. Chem.,* **20,** 267 (1991). Electrostatic Energy and Macromolecular Function.

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

179. P. Kollman, *Chem. Rev.,* **7,** 2395 (1993). Free Energy Calculations-Applications to Chemical and Biological Phenomena.

180. I. Muegge, U. Ermler, G. Fritzsch, and E. W. Knapp, *J. Phys. Chem.,* **99,** 17917 (1995). Free Energy of Cofactors at the Quinone Q(a) Site of the Photosynthetic Reaction Center of *Rhodobacter sphaeroides* Calculated by Minimizing the Statistical Error.

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

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

183. F. S. Lee, Z. T. Chu, and A. Warshel, *J. Comput. Chem.,* **14,** 161 (1993). Microscopic and Semimicroscopic Calculations of Electrostatic Energies in Proteins by the POLARIS and ENZYMIX Programs.

184. J. Aqvist, C. Medina, and J. E. Samuelsson, *Protein Eng.,* **7,** 386 (1994). New Method for Predicting Binding Affinity in Computer-Aided Drug Design.

185. M. K. Holloway, *Perspect. Drug Disc. Des.,* **9/10/11,** 63 (1998). A Priori Prediction of Ligand Affinity by Energy Minimization.

186. 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, A. K. Ghosh, E. A. Giuliani, S. L. Graham, J. P. Guare, R. W. Hungate, T. A. Lyle, W. M. 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-1 Protease Inhibitors Employing Energy Minimization in the Active Site.

187. C. Zhang, V. Vasmatzis, J. Cornette, and C. DeLisi, *J. Mol. Biol.,* **267,** 707 (1997). Determination of Atomic Desolvation Energies from the Structures of Crystallized Proteins.

188. N. Horton and M. Lewis, *Protein Sci.,* **1,** 169 (1992). Calculation of the Free Energy of Association for Protein Complexes.

189. H. J. Böhm, *J. Comput.-Aided Mol. Des.,* **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.

190. M. D. Eldridge, 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.

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

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

193. A. Alex and P. Finn, *J. Mol. Struct.* (*THEOCHEM*), **398–399,** 551 (1997). Fast and Accurate Prediction of Relative Binding Energies.

194. 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.

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

196. R. L. Jernigan and I. Bahar, *Curr. Opin. Struct. Biol.,* **6,** 195 (1996). Strucure-Derived Potentials and Protein Simulations.

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

198. 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.

199. 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.

200. A. Wallqvist, R. L. Jernigan, and D. G. Covell, *Protein Sci.,* **4,** 1881 (1995). A Preference-Based Free Energy Parameterization of Enzyme–Inhibitor Binding. Application to HIV-1–Protease Inhibitor Design.

201. T. R. Stouch and P. C. Jurs, *J. Chem. Inf. Comput. Sci.,* **26,** 4 (1986). A Simple Method for the Representation, Quantification, and Comparison of the Volumes and Shapes of Chemical Compounds.

202. M. C. Nicklaus, S. Wang, J. S. Driscoll, and G. W. A. Milne, *Bioorg. Med. Chem.,* **3,** 411 (1995). Conformational Changes of Small Molecules Binding to Proteins.

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

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

205. D. R. Flower, *J. Mol. Graphics Modell.,* **15,** 238 (1998). SERF: A Program for Accessible Surface Area Calculations.

206. J. S. Dixon, *Proteins: Struct., Funct., Genet., Suppl.,* **1,** 198 (1997). Evaluation of the CASP2 Docking Section.

207. B. Lee and F. M. Richards, *J. Mol. Biol.,* **55,** 379 (1971). Interpretation of Protein Structure—Estimation of Static Accessibility.

208. 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.

209. B. K. Shoichet and I. D. Kuntz, *Protein Eng.,* **6,** 723 (1993). Matching Chemistry and Shape in Molecular Docking.
210. P. H. Walls and M. J. E. Sternberg, *J. Mol. Biol.,* **228,** 277 (1992). New Algorithm to Model Protein–Protein Recognition Based on Surface Complementarity.
211. D. Eisenberg and A. D. McLachlan, *Nature* (*London*), **319,** 199 (1986). Solvation Energy in Protein Folding and Binding.
212. V. N. Viswanadhan, M. R. Reddy, A. Wlodawer, M. D. Varney, and J. N. Weinstein, *J. Med. Chem.,* **39,** 705 (1996). An Approach to Rapid Estimation of Relative Binding Affinities of Enzyme Inhibitors: Application to Peptidomimetic Inhibitors of the Human Immunodeficiency Virus Type 1 Protease.
213. V. N. Viswanadhan, A. K. Ghose, G. R. Revankar, and R. K. Robins, *J. Chem. Inf. Comput. Sci.,* **29,** 163 (1989). Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure–Activity Relationships. 4. Additional Parameters for Hydrophobic and Dispersive Interactions and Their Application for an Automated Superposition of Certain Naturally Occurring Nucleoside Antibiotics.
214. L. A. Kuhn, C. A. Swanson, M. E. Pique, J. A. Tainer, and E. D. Getzoff, *Proteins: Struct., Funct., Genet.,* **23,** 536 (1995). Atomic and Residue Hydrophilicity in the Context of Folded Protein Structures.
215. 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.
216. B. B. Goldman and W. T. Wipke, *Proteins: Struct., Funct., Genet.,* **38,** 79 (2000). QSD Quadratic Shape Descriptors. 2. Molecular Docking Using Quadratic Shape Descriptors (QSDock).
217. I. Muegge, T. Schweins, R. Langen, and A. Warshel, *Structure,* **4,** 475 (1996). Electrostatic Control of GTP and GDP Binding in the Oncoprotein P21(Ras).
218. I. Muegge, H. Tao, and A. Warshel, *Protein Eng.,* **10,** 1363 (1997). A Fast Estimate of Electrostatic Group Contributions to the Free Energy of Protein–Inhibitor Binding.
219. P. D. J. Grootenhuis and P. J. M. van Galen, *Acta Crystallogr.,* **D51,** 560 (1995). Correlation of Binding Affinities with Non-Bonded Interaction Energies of Thrombin–Inhibitor Complexes.
220. N. S. Blom and J. Sygusch, *Proteins: Struct., Funct., Genet.,* **27,** 493 (1997). High Resolution Fast Quantitative Docking Using Fourier Domain Correlation Techniques.
221. N. Tomioka and A. Itai, *J. Comput.-Aided Mol. Des.,* **8,** 347 (1994). GREEN: a Program Package for Docking Studies in Rational Drug Design.
222. K. A. Sharp and B. Honig, *Annu. Rev. Biophys. Biophys. Chem.,* **19,** 301 (1990). Electrostatic Interactions in Macromolecules.
223. G. Jones, P. Willett, and R. C. Glen, *J. Mol. Biol.,* **254,** 43 (1995). Molecular Recognition of Receptor Sites Using a Genetic Algorithm with a Description of Solvation.
224. M. Liu and S. M. Wang, *J. Comput.-Aided Mol. Des.,* **13,** 435 (1999). MCDOCK: A Monte Carlo Simulation Approach to the Molecular Docking Problem.
225. 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.
226. 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.
227. 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.

228. D. Qui, P. S. Shenkin, E. P. Hollinger, and W. C. Still, *J. Phys. Chem.,* **101,** 3005 (1997). The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii.

229. P. F. W. Stouten, C. Frommel, H. Nakamura, and C. Sander, *Mol. Simul.,* **10,** 97 (1993). An Effective Solvation Term Based on Atomic Occupancies for Use in Protein Simulations.

230. 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.

231. M. Vieth, J. D. Hirst, A. Kolinski, and C. L. Brooks III, *J. Comput. Chem.,* **14,** 1612 (1998). Assessing Energy Functions for Flexible Docking.

232. D. R. Westhead, D. E. Clark, and C. W. Murray, *J. Comput.-Aided Mol. Des.,* **11,** 209 (1997). A Comparison of Heuristic Search Algorithms for Molecular Docking.

233. C. W. Murray, C. A. Baxter, and A. D. Frenkel, *J. Comput.-Aided Mol. Des.,* **13,** 547 (1999). The Sensitivity of the Results of Molecular Docking to Induced Fit Effects: Application to Thrombin, Thermolysin and Neuraminidase.

234. C. W. Murray, T. R. Auton, and M. D. Eldridge, *J. Comput.-Aided Mol. Des.,* **12,** 503 (1998). 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 Model.

235. Available Chemicals Directory (ACD) is available from MDL Information Systems Inc., San Leandro, CA 94577, and contains specialty bulk chemicals from commercial sources. URL www.mdli.com.

236. M. J. Sippl, *J. Mol. Biol.,* **213,** 859 (1990). Calculation of Conformational Ensembles from Potentials of Mean Force.

237. 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.

238. M. J. Sippl, M. Ortner, M. Jaritz, P. Lackner, and H. Flöckner, *Folding Design,* **1,** 289 (1996). Helmholtz Free Energies of Atom Pair Interactions in Proteins.

239. D. T. Jones, W. R. Taylor, and J. M. Thornton, *Nature* (*London*), **358,** 86 (1992). A New Approach to Protein Fold Recognition.

240. A. Godzik, A. Kolinski, and J. Skolnick, *J. Mol. Biol.,* **227,** 227 (1992). Topology Fingerprint Approach to the Inverse Protein Folding Problem.

241. C. Ouzounis, C. Sander, M. Scharf, and R. Schneider, *J. Mol. Biol.,* **232,** 805 (1993). Prediction of Protein Structure by Evaluation of Sequence–Structure Fitness.

242. I. Bahar and R. L. Jernigan, *J. Mol. Biol.,* **266,** 195 (1997). Interresidue Potentials in Globular Proteins and the Dominance of Highly Specific Hydrophilic Interactions at Close Separation.

243. A. Ben-Naim, *J. Chem. Phys.,* **107,** 3698 (1997). Statistical Potentials Extracted from Protein Structures: Are these Meaningful Potentials?

244. P. D. Thomas and K. Dill, *J. Mol. Biol.,* **257,** 457 (1996). Statistical Potentials Extracted from Protein Structures: What Is Wrong with Them?

245. W. A. Koppensteiner and M. J. Sippl, *Biochemistry* (*Moscow*), **63,** 247 (1998). Knowledge-Based Potentials—Back to the Roots.

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

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

248. S. Ha, R. Andreani, A. Robbins, and I. Muegge, *J. Comput.-Aided Mol. Des.,* **14,** 435 (2000). Evaluation of Docking/Scoring Approaches: A Comparative Study Based on MMP-3 Inhibitors.

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

250. 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. Des.,* **13,** 167 (1999). Comparison of Two Implementations of the Incremental Construction Algorithm in Flexible Docking of Thrombin Inhibitors.

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

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

253. T. A. Halgren, *J. Comput. Chem.,* **17,** 520 (1996). Merck Molecular Force Field. II. MMFF94 Van der Waals and Electrostatic Parameters for Intermolecular Interactions.

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

255. SYBYL, version 6.6, Tripos Associates, St. Louis, MO. URL www.tripos.com.

256. M. Hendlich, F. Rippman, and G. Barnickel, *J. Mol. Graphics Modell.,* **15,** 359 (1997). LIGSITE: Automatic and Efficient Detection of Potential Small Molecule Binding Sites in Proteins.

257. MDL Information Systems, Inc., San Leandro, CA. URL www.mdli.com.

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

259. R. S. Pearlman, *Chem. Design Automation News,* **2** (*1*) (1987). Rapid Generation of High Quality Approximate 3D Molecular Structures. Information about CONCORD is available at www.tripos.com.

260. URL www.epm.ornl.gov/pvm/.

261. W. P. Walters, Ajay, and M. A. Murcko, *Curr. Opin. Chem. Biol.,* **3,** 384 (1999). Recognizing Molecules with Drug-Like Properties.

262. B. L. Podlogar, I. Muegge, and L. J. Brice, *Curr. Opin. Drug Discovery Devel.,* **4,** 102 (2001). Current Computational Methods to Estimate Drug Development Parameters for Use in Drug Discovery: A Zero Infrastructure Approach.

263. E. J. Martin, J. M. Blaney, M. A. Siani, D. C. Spellmeyer, A. K. Wong, and W. H. Moos, *J. Med. Chem.,* **38,** 1431 (1995). Measuring Diversity. Experimental Design of Combinatorial Libraries for Drug Discovery.

264. 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.

265. E. J. Martin, D. C. Spellmeyer, R. E. Critchlow Jr., and J. M. Blaney, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1997, Vol. 10, pp. 75–100. Does Combinatorial Chemistry Obviate Computer-Aided Drug Design?

266. P. D. Grootenhuis, P. A. Kollman, G. L. Seibel, R. L. DesJarlais, and I. D. Kuntz, *Anticancer Drug Des.,* **5,** 237 (1990). Computerized Selection of Potential DNA Binding Compounds.

267. R. L. DeJarlais, G. L. Seibel, I. D. Kuntz, P. S. Furth, J. C. Alvarez, P. R. Ortiz DeMontellano, D. L. DeCamp, L. M. Babe, and C. S. Craik, *Proc. Natl. Acad. Sci. U.S.A.* **87,** 6644 (1990). Structure-Based Design of Nonpeptide Inhibitors Specific for the Human Immunodeficiency Virus 1 Protease.

268. C. S. Ring, E. Sun, J. H. McKerrow, G. K. Lee, P. J. Rosenthal, I. D. Kuntz, and F. E. Cohen, *Proc. Natl. Acad. Sci. U.S.A.* **90,** 3538 (1993). Structure-Based Inhibitor Design by Using Protein Models for the Development of Antiparasitic Agents.

269. B. Kramer, M. Rarey, and T. Lengauer, *Proteins: Struct., Funct., Genet.,* **37,** 228 (1999). Evaluation of the FlexX Incremental Construction Algorithm for Protein–Ligand Docking.

270. C. A. Baxter, C. W. Murray, B. Waszkowycz, J. Li, R. A. Sykes, R. G. A. Bone, T. D. J. Perkins, and W. Wylie, *J. Chem. Inf. Comput. Sci.,* **40,** 254 (2000). New Approach to Molecular Docking and Its Application to Virtual Screening of Chemical Databases.

271. M. Stahl, M. Rarey, and G. Klebe, in *Bioinformatics,* T. Lengauer, Ed., Wiley-VCH, New York, 2001, Vol. 2, in press. Screening of Drug Databases.

272. World Drug Index is available from Derwent Information, London, UK. URL www.derwent.com.

273. 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 Sensitization Using Biphenyl Tetrazoles as Potent Inhibitors of *Bacteroides fragilis* Metallo-β-lactamase.

274. D. Horvath, *J. Med. Chem.,* **40,** 2412 (1997). A Virtual Screening Approach Applied to the Search for Trypanothione Reductase Inhibitors.

275. E. Perola, K. Xu, T. M. Kollmeyer, S. H. Kaufmann, F. G. Prendergast, and Y. P. Pang, *J. Med. Chem.,* **43,** 401 (2000). Successful Virtual Screening of a Chemical Database for Farnesyltransferase Inhibitor Leads.

276. D. L. Kirkpatrick, S. Watson, and S. Ulhaq, *Comb. Chem. High Throughput Screening,* **2,** 211 (1999). Structure-Based Drug Design: Combinatorial Chemistry and Molecular Modeling.

277. R. E. Cachau, K. Fidelis, and A. E. Roitberg, a Web page formerly at URL http://uqbar.ncifcrf.gov/~catfee/index.html.

278. URL http://predict.sanger.ac.uk/casp2/.

279. T. P. Maduskule Jr., K. J. McNamara, Y. Ru, R. M. Knabb, and P. F. W. Stouten, *J. Med. Chem.,* **41,** 53 (1998). Rational Design and Synthesis of Novel, Potent Bis-phenylamidine Carboxylate Factor Xa Inhibitors.

280. M. von Itzstein, W. Wu, G. B. Kok, M. S. Pegg, J. C. Dyason, B. Jin, T. V. Phan, M. L. Smythe, H. F. White, 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.

281. M. N. Janakiraman, C. L. White, W. G. Laver, G. M. Air, and M. Luo, *Biochemistry,* **33,** 8172 (1994). Structure of Influenza Virus Neuraminidase B/Lee/40 Complexed with Sialic Acid and a Dehydro Analog at 1.8 Å Resolution: Implications for the Catalytic Mechanism.

282. S. Singh, M. J. Jedrzejas, G. M. Air, M. Luo, W. G. Laver, and W. J. Brouillette, *J. Med. Chem.,* **38,** 3217 (1995). Structure-Based Inhibitors of Influenza Virus Sialidase. A Benzoic Acid Lead with Novel Interaction.

283. M. K. Gilson, A. Rashin, R. Fine, and B. Honig, *J. Mol. Biol.,* **184,** 503 (1985). On the Calculation of Electrostatic Interactions in Proteins.

284. I. Klapper, R. Hagstrom, R. Fine, K. Sharp, and B. Honig, *Proteins: Struct., Funct., Genet.,* **1,** 47 (1986). Focusing of Electric Fields in the Active Site of Cu-Zn Superoxide Dismutase: Effects of Ionic Strength and Amino-Acid Modification.

285. M. J. Jedrzejas, S. Singh, W. J. Brouillette, G. M. Air, and M. Luo, *Proteins: Struct., Funct., Genet.,* **23,** 264 (1995). A Strategy for Theoretical Binding Constant, K_i, Calculations for Neuraminidase Aromatic Inhibitors Designed on the Basis of the Active Site Structure of Influenza Virus Neuraminidase.

286. J. A. Landro, I. C. A. Taylor, W. G. Stirtan, D. G. Osterman, J. Kristie, E. J. Hunnicutt, P. M. M. Rae, and P. M. Sweetnam, *J. Pharmacol. Toxicol. Meth.*, **44**, 273 (2000). HTS in the New Millennium. The Role of Pharmacology and Flexibility.

287. W. P. Walters (Vertex Pharmaceuticals), personal communication, 2000.

288. C. Vose, Lecture given at the Clinical Pharmacology, Bayer North America 10th Anniversary Symposium "Clinical Pharmacology in the Next Millennium", Bayer Research Center, West Haven, CT, December 21, 2000. Role of Clinical Pharmacology in Go/No Go Decision-Making.

289. Note added after this chapter entered production: In regard to our section on consensus scoring, a number of additional references have appeared. See, e.g., C. Bissantz, G. Folkers, D. Rognan, *J. Med. Chem.,* **43,** 4759 (2000). Protein-Based Virtual Screening of Chemical Databases. 1. Evaluation of Different Docking/Scoring Combinations. The ETH Zürich

group compares DOCK4.01, FlexX1.8, and GOLD1.1 for docking and ChemScore, DOCK, FlexX, GOLD, PMF, Fresno, and SCORE for consensus scoring. The program Fresno is described by D. Rognan, S. L. Laumoeller, A. Holm, S. Buus, and V. Tschinke, *J. Med. Chem.*, **42,** 4650 (1999). Predicting Binding Affinities of Protein Ligands from Three-Dimensional Coordinates: Application to Peptide Binding to Class I Major Histocompatibility Proteins. SCORE is described in Ref. 204. In the section of our chapter on Docking Calculations, the use of workstation clusters is mentioned. Recently, groups of Oxford University and GlaxoSmithKline disclosed independently developed systems for distributing docking and related drug design calculations on a potentially vast array of personal computers. See: W. G. Richards and D. D. Robinson, Abstracts of 221nd National Meeting of the American Chemical Society, San Diego, CA, April 1-5, 2001, COMP 27. Pattern Recognition Techniques in Drug Discovery; C. E. Keefer, E. C. Bigham, S. S. Young, Abstracts of 222nd National Meeting of the American Chemical Society, Chicago, IL, August 26-30, 2001, COMP 97. Pharmacophore Identification Using Massive Distributed Computing. With respect to our section on MD simulation techniques for molecular docking, we would like to point out recent efforts to decrease simulation time using generalized effective potentials described by Y. Pak and S. Wang, *J. Phys. Chem. B*, **104**, 354 (2000). Application of a Molecular Dynamics Simulation Method with a Generalized Effective Potential to the Flexible Molecular Docking Problems.

Reviews in Computational Chemistry, Volume 17. Edited by Kenny B. Lipkowitz, Donald B. Boyd

ISBNs: 0-471-39845-4 (Hardcover); 0-471-22441-3 (Electronic)

CHAPTER 2

Protein–Protein Docking

Lutz P. Ehrlich* **and Rebecca C. Wade†**

*European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69012 Heidelberg, Germany, (present address): *Lion Bioscience AG, D-69123 Heidelberg, Germany, and (present address): †European Media Laboratory, Schloss-Wolfsbrunnenweg 33, D-69118 Heidelberg, Germany*

INTRODUCTION

In this review,[1] we discuss current methods to predict the structure of a protein–protein complex, given the structures of the respective proteins in their unbound conformations. We begin by putting the protein–protein docking problem in context with respect to other biomolecular docking problems and with respect to its role in the current genomic era. We then highlight the crucial points that need to be considered when embarking on a docking project. We review the anatomy of prototypical protein–protein interfaces and describe structural changes observed upon binding in experimentally determined complexes. Then, after giving a brief overview of manual docking, we present automated methods for docking. We describe current algorithms, their limitations, and the situations in which they are applicable. After these descriptive review sections, we guide the reader through a case study that illustrates some of the important points that should be considered in protein–protein docking.

Why This Topic?

The study of protein–protein interactions is at a turning point. With the sequences of a significant number of genomes revealed, comparative bioinformatics studies[2–4] can be used to reveal protein interaction partners. On the experimental side, large-scale automation of the yeast two-hybrid technique is making it possible to identify all possible protein interaction partners in entire genomes,[5] and mass spectrometry (MS) has recently become a powerful tool for identifying the proteins that constitute protein complexes.[6] All of these techniques give researchers valuable information about which proteins in an organism interact with each other. They do not, however, define exactly how the proteins interact. This information is vital for the understanding of cellular function and disease-related processes. A full understanding of these cellular processes and the rational design of molecules to modulate them, notably for therapeutic benefit, requires knowledge of the atomic interactions of the proteins involved. Only a small fraction of the proteins in the human genome and other genomes have structures that have been solved by nuclear magnetic resonance (NMR) or X-ray crystallographic methods. However, techniques combining structural genomics[7] with advanced homology modeling[8] and ab initio fold prediction[9] tools give rise to the hope that workable models for all proteins involved in disease-related processes will become available. With these protein three-dimensional (3D) structures in hand, automated protein–protein docking methods will be vital for structure-based approaches to drug design and biotechnology.[10]

Protein–Protein Binding Data

Given the multitude of possible protein–protein binding modes, it is important to define the goals of any particular docking project at the outset. Is it to derive the interactions of a large set of relatively rigid proteins by screening at low resolution? Or does the researcher want a detailed model of the binding process of two proteins that exhibit significant flexibility in the binding interface? To make a suitable choice, it is necessary to consider the properties of protein–protein complexes. Therefore, in the following sections, we recapitulate current knowledge from experimental data on the structure, energetics, and dynamics of protein–protein complexes.

Structural Properties

In the most comprehensive structural analysis to date, Lo Conte et al.[11] studied 75 protein–protein complexes comprising 24 protease–inhibitor, 19 antibody–antigen, and 32 other complexes (including 9 further enzyme–inhibitor and 11 signal transduction complexes). The authors found that protein–protein interfaces typically have a size of 1600 ± 400 Å^2 with a few complexes exhibiting very large (2000–4660 Å^2) or very small (less than 1000 Å^2) interfaces. With respect to their chemical nature, the interfaces were found to

resemble protein surfaces more than protein interiors. Nevertheless, Jones and Thornton[12] found, by analysis of six parameters (solvation potential, residue interface propensity, hydrophobicity, planarity, protrusion, and accessible surface area) that the interface surface patches are distinguishable from other surface patches on proteins. On average, Lo Conte et al.[11] found one hydrogen bond per 170-Å^2 interface area. The interfaces of protease–inhibitor complexes were found to exhibit more polar residues, and those of antibody–antigen complexes fewer polar residues than the rest of the complexes studied. In general, protein–protein interfaces are enriched in aromatic residues.

Water molecules are frequently present at protein–protein interfaces and, on average, one water molecule per 100 Å^2 is found in high-resolution structures.[11] The presence of waters in crystal structures of protein–protein complexes first became apparent in antibody–lysozyme complexes,[13–14] and they have since been identified at the interfaces of many protein–protein complexes. The role of these water molecules seems to be twofold. First, they contribute to the proper packing of the protein–protein interface region.[11] Second, they mediate intermolecular hydrogen-bond formation. On average, there are as many water mediated as direct hydrogen bonds across protein–protein interfaces.[15] Figure 1 illustrates, as an example, part of the network of water-mediated hydrogen bonds in the barnase–barstar interface. Apart from playing a structural role, water molecules also make important energetic contributions to protein–protein binding. Mutational and computational analyses have shown their energetic contribution. Covell and Wallqvist[16] computed that in crystal structures of three selected protein–protein complexes, interfacial crystallographically observed water molecules contribute around 25% of the total calculated binding strength.

Kinetic and Thermodynamic Properties

The measured dissociation constants of biomolecular protein–protein complexes cover a range from roughly 10^{-3} to 10^{-13} M, corresponding to binding affinities up to about 18 kcal/mol. Association rates fall in the range 10^3 to 10^{10} M^{-1} s^{-1}, and dissociation rates range from 10^3 to 10^{-8} s^{-1}.

The energetics of protein association can be studied by a variety of experimental techniques,[17] each of which permits measurements of equilibrium or kinetic values in a certain range. Widely used techniques include isothermal titration calorimetry, surface plasmon resonance measurement, stopped flow kinetics, optical spectroscopy, MS, and analytical ultracentrifugation. The techniques differ in their requirements (e.g., amount of protein, labeling with fluorophores, attachment to sensor surfaces, and the environment provided by the experimental set up) and therefore in their applicability to individual cases. Different techniques can also give quite different values for what might be expected to be the same quantity. For example, association rates measured by surface plasmon resonance, with one protein immobilized on a surface, are usually different from those measured for the two proteins in solution and under otherwise similar conditions.

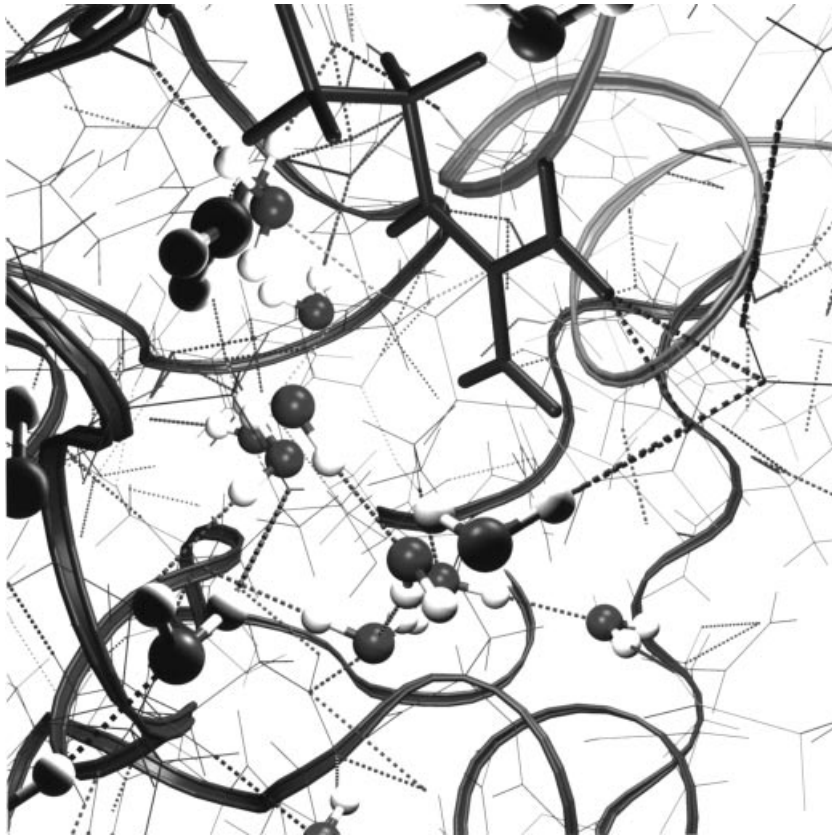

Figure 1 Interfacial hydrogen-bonding network in the complex of barnase and barstar. Water molecules are depicted in ball-and-stick representation. Barnase is shown on the left by the dark ribbon with arginine-59 in a stick representation. Barstar on the right is represented by the light gray ribbon. Hydrogen bonds are depicted by dashed lines. Hydrogen atoms were added to the A and D chains and water molecules of PDB file 1brs by optimizing the hydrogen-bond network using WHAT IF (Refs. 101,102) and energy minimization using the CHARMM22 force field (Ref. 103) implemented in X-PLOR (Ref. 104).

Protein engineering is a powerful tool to use in conjunction with binding measurements to investigate which residues are most important for binding. It has been shown that not all residues at protein–protein interfaces contribute equally to binding affinity. Rather, there often exist "hot spots" containing a few residues that account for most of the binding affinity. The existence of hot spots has been demonstrated for binding of the human growth hormone to its receptor[18] and subsequently for other complexes including barnase and barstar[19] and interleukin-4 and the interleukin-4α receptor.[20] Moreover, it has been shown for a number of examples, including the interleukin-4–interleukin-4α receptor complex,[20] that the binding epitope (in this context, the part of the protein determining recognition and binding) can consist of

two distinct but overlapping epitopes: one containing residues of primary importance for the dissociation rate, the other containing residues of most importance for the association rate. The association rate will be most dependent on mutation of residues making contacts in the diffusional encounter complex,[21] whereas the dissociation rate will be dependent on the structure of the bound complex. This difference can be exploited for the design of mutants in which the association rate, but not the dissociation rate, is altered.[22]

The binding affinity of two proteins is frequently the result of enthalpy–entropy compensation. Ideally, then, the estimation of the binding affinity during protein–protein docking should include the computation of both enthalpies and entropies of binding,[23] thus enabling estimation of the temperature dependence of the binding affinity. However, enthalpy–entropy compensation makes it difficult to account for these two terms explicitly[24] while providing fast, accurate estimates of binding affinity during docking. Further, there is relatively little calorimetric data available on protein–protein binding for calibrating the empirical functions used to estimate enthalpies and entropies. However, some systematic studies have been carried out to perform coordinated calorimetric experiments and computations.[25] Good agreement between experiment and the structural energetic calculations (in which thermodynamic parameters are estimated based upon changes that occur in polar and apolar solvent-accessible surface area) have been obtained.[26,27]

Protein–protein binding energetics are influenced not only by the protein sequence and structure, but also by the conditions under which binding takes place. The pH is an important parameter, and pH dependence has been measured for protein–protein binding affinity.[19,28] Sensitivity to pH arises from changes in the ionization state of residues important for binding. For example, Schreiber and Fersht[29] could attribute the pH dependence of barnase–barstar unbinding to a single barnase histidine residue. Ionic strength can also affiect binding, particularly when electrostatic interactions are strong. There can be specific ion-binding efects that can affiect the stability of the proteins and their binding thermodynamics, as exemplified by phosphate ion binding.[25] Finally, the viscosity of the solvent can affect the binding kinetics; viscogens exert osmotic pressure and thus alter binding affinity.[30]

There are several energetic contributions to protein–protein binding affinity. The relative importance of hydrophobic and electrostatic interactions varies in diffierent complexes. So too does the contribution of individual water molecules and specific hydrogen bonds. Protein–protein interfaces do not generally show charge complementarity, but, due to the long-range nature of electrostatic interactions and the heterogeneity of the protein–solvent dielectric environment, they do show complementarity of their surface electrostatic potentials.[31] Therefore, it is not a straightforward task to derive simple expressions that are universally applicable to estimating binding affinity. However, correlations between experimental binding affinities and relatively simple additive energy functions that depend on interfacial properties have been obtained.[32,33]

Dynamic Properties

One of the major difficulties in predicting the structure of the bound complex from two unbound protein conformations is the structural changes that take place upon binding. These changes can arise from:

- The need to establish or improve specific interactions between the two binding partners.
- The need to improve geometric fit and avoid steric clashes.
- The inherent flexibility of the proteins involved.
- Functional reasons, such as conformational changes that can trigger events in signaling or are important in allosteric effiects.

Local structural changes, such as surface side-chain rotations, occur on time scales[34] of 10^{-11} to 10^{-10} s and length scales of several angstroms. A typical example of this kind of motion is illustrated in Figure 2, which shows the

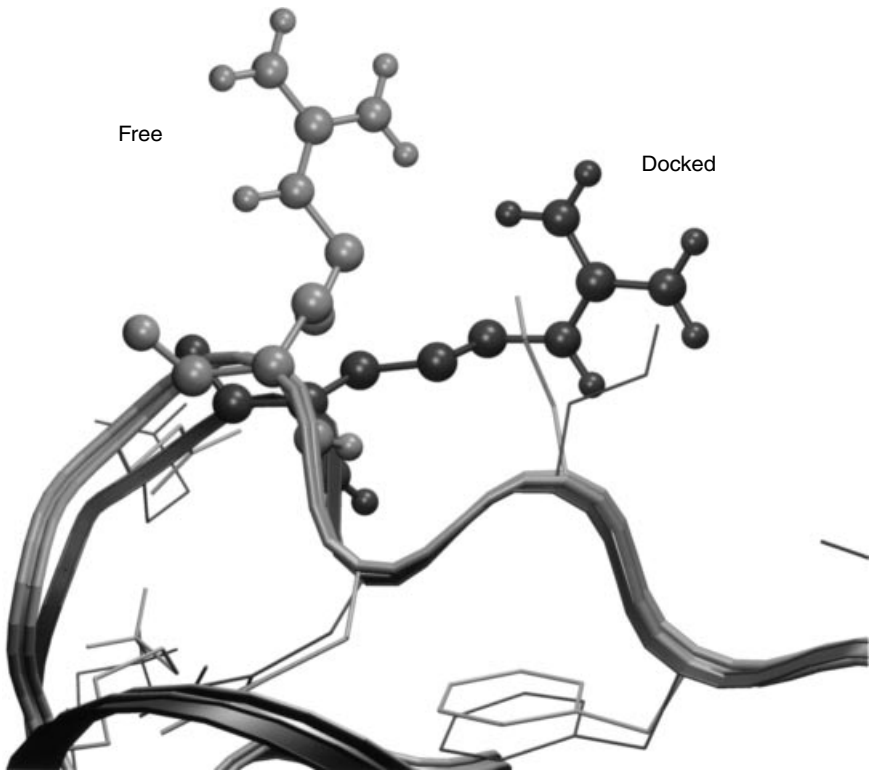

Figure 2 Rearrangement of interfacial side chains in barnase upon binding barstar. Dark (light) color represents the bound (unbound) conformation. Arg-59 is highlighted in ball-and-stick representation, using a lighter color for the free and a darker for the bound conformation. The viewpoint is chosen from barnase across its guanine-binding loop toward barstar.

structural change in an interfacial loop of barnase upon binding to barstar. The structural reorganization involves small-scale backbone adjustments along with changes of side-chain rotamers.

At the other end of the spectrum, large-scale structural changes occur on time scales[34] of 10^{-11} to 10^{-3} s and can involve motions over several tens of angstroms. To illustrate the diffierent types of large-scale conformational changes upon docking, we have chosen to focus on the complex of cyclin and cyclin-dependent kinase 2 (CDK2). A C_α representation of the complex with CDK2 superposed in bound and unbound forms is illustrated in Figure 3(**a**). A large-scale loop rearrangement upon binding is illustrated in Figure 3(**b**). A disorder–order transition is illustrated in Figure 3(**c**). For docking in such cases, the ordered bound structure of the unordered region's conformation must be predicted. Finally, a large-scale domain motion upon binding is shown in Figure 3(**d**).

In their recent review of the conformational changes upon formation of 31 heteroprotein complexes (18 enzyme–inhibitor, 7 antibody–antigen, and 6 other), Betts and Sternberg[35] conclude that conformational change is considerable for just over half of the complexes studied. In these cases, the observed structural changes were larger than those between different experimental structures of the same protein. From these comparisons, the thresholds above which a structural change upon binding can be considered significant are: 0.6 Å for C_α atoms and 1.7 Å for side chains of exposed residues. Conformational changes involved both side-chain and main-chain contributions. In their data set, Betts and Sternberg found no large-scale conformation changes upon protein–protein binding arising from correlated shear or hinge-bending motion.

Challenges for Computational Docking Studies

Given the multitude of approaches now used to dock small molecules and peptides to proteins, why should the protein–protein docking problem require its own approaches? Could one make do with the software packages developed for these related problems? Unfortunately, the answer is usually no.

A fundamental difference between protein–protein and protein–small-molecule docking is the number of degrees of freedom that must be considered when generating putative complexes. With the average protein–protein interface[11] spanning an area of 1600 Å^2, the number of protein atoms directly involved in contact formation between receptor and ligand is large compared to the number of atoms involved in the binding of small molecules or peptides. Thus, the generation of new putative binding conformations is computationally more complex and may require new classes of algorithms to be applied. Related to the larger number of degrees of freedom and the size of the interface is the extent of conformational changes of the receptor. As mentioned in the previous sections, large-scale structural changes involving backbone motions on the order of ca. 10 Å are observed for some complexes. Thus, approaches

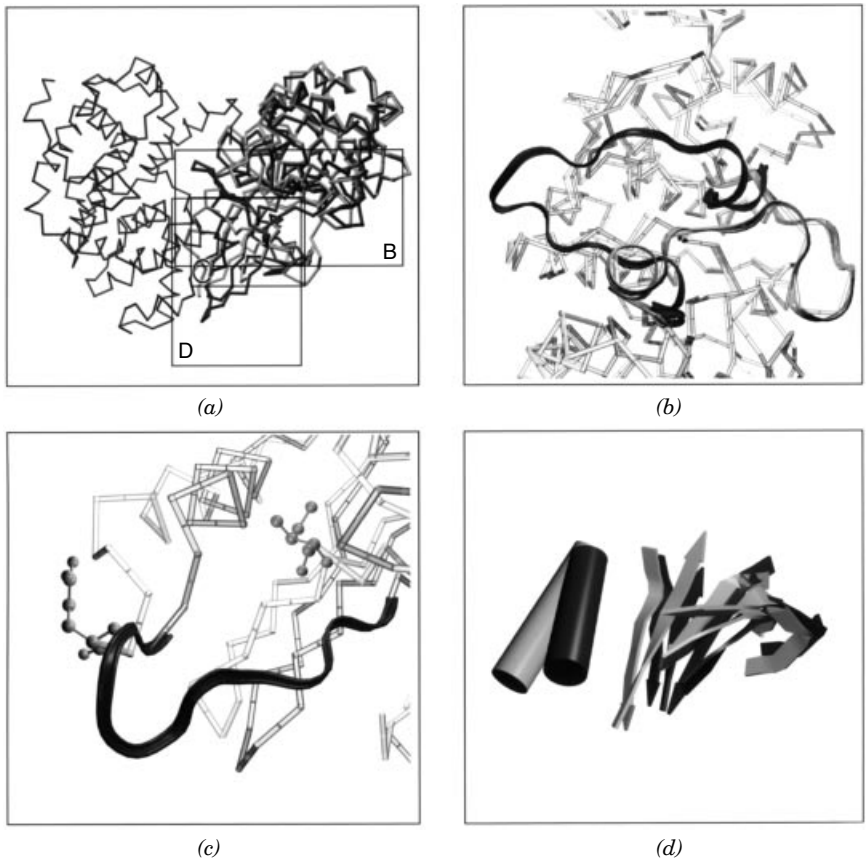

Figure 3 Large-scale conformational changes upon protein–protein binding. (**a**) Cyclin bound to cyclin-dependent kinase 2 (CDK2) (PDB code 1fin, chains A and B). In the structure of the complex, the unbound CDK2 structure (PDB code 1hck, chain A) is superposed onto the bound structure. The all-C_α rmsd between bound and unbound CDK2 conformations is 3.9 Å. Superposition of the relevant CDK2 chains was performed by minimizing the rmsd of all C_α atoms (Ref. 105). The rectangular regions B, C, and D are shown magnified in (**b**), (**c**), and (**d**), respectively. For PDB, see Ref. 106. (**b**) Large-scale loop rearrangement upon binding. Residues 144–167 of CDK2 change their backbone conformation up to ca. 12 Å. (**c**) Disorder–order transition of residues 37–40 in CDK2. The disordered residues are not resolved in the experimental structure; the flanking residues are shown in a ball-and-stick representation. (**d**) Large-scale domain motion upon binding. Residues 1–80 of CDK2 display a tilt motion leading to a local backbone rmsd of 3.5 Å upon binding.

to generate putative complexes will ideally have to be able to tackle large-scale receptor conformational changes.

Another difference between protein–protein and protein–small-molecule docking involves the treatment of nonbonded interactions. For the docking of rigid low molecular weight compounds, the receptor's molecular interaction

fields, such as the electrostatic potential, can often be computed and stored prior to the docking simulation. On the other hand, the more extensive conformational changes occurring during protein–protein binding require updating of these computationally expensive quantities during docking. Thus, scoring functions for protein–protein docking either have to treat these interactions in a heuristic way or they must invoke more efficient techniques for the evaluation of the expensive nonbonded interactions.

In the following section, we describe computational approaches used in the protein–protein docking field. We focus mostly on the aspect of generating putative complex structures. Although some small molecule docking programs have been extended for application to protein–protein docking as well, most approaches represent new efforts targeted specifically at protein–protein docking.

COMPUTATIONAL APPROACHES TO THE DOCKING PROBLEM

In this section, we present an overview of the methods currently in use for docking binary protein complexes. Although docking of ternary protein complexes has been reported,[36] the majority of computational studies to date deal with the case of two interacting proteins.

Manual docking approaches have long been used for generating protein–protein complexes. These approaches require the researcher to dock the two proteins interactively and determine the chemical and geometric fit by means of computer graphics tools. Preliminary structures of complexes generated manually can be refined to remove bad contacts by energy minimization. Manual docking is often done by using experimental data on the effects of mutations as guides for generating plausible structures of complexes. Anderson and Weng[37] described an advanced approach for manual protein–protein docking utilizing immersive virtual reality environments such as haptic devices that create a sense of touch via controlled force feedback to the user's hand. Although they could demonstrate successful protein–protein docking in two cases using virtual reality, no large-scale applications have been reported so far.

Two general categories of approaches to automated protein–protein docking can be distinguished. In the first, rigid-body docking, both proteins are considered to be completely rigid, thereby reducing the docking problem to a search through six degrees of freedom for the optimal orientation. Rigid-body docking approaches are usually geared toward optimizing the geometric and/or chemical fit of the two proteins. Typical execution times on current computer hardware are on the order of minutes for representative enzyme–inhibitor complexes. This computational efficiency enables the user to automatically screen databases of proteins for mutual interactions. On the other hand, conformational rearrangements of the proteins and interfacial solvation are not treated in any of the rigid-body docking approaches.

In the second approach, flexible docking, protein conformational flexibility is treated explicitly. The extent to which protein flexibility is modeled varies. Some approaches account for the relative motions of protein domains, others account for rotation of amino acid side chains, and others treat all atomic degrees of freedom. These approaches are based on minimizing energy or on optimizing shape-based criteria, or on performing atomic-detail simulations that allow a time interpretation of the binding process. Naturally, execution speeds can be extensive, ranging from hours to days for a single docking on current computing hardware. Typically, the value of fine-grained flexible docking approaches is to understand an interesting binding event particularly well. Screening databases with thousands of putative interaction partners is out of the reach of current flexible docking methods due to their computational demands.

Docking = Sampling + Scoring

Throughout this chapter, we roughly separate docking approaches by their treatment of protein flexibility. Within the two broad categories (rigid body and flexible docking), the docking problem can be separated into two subproblems:

1. The sampling problem: Given two independent protein structures, generate putative structures of the complex.
2. The scoring problem: Given a set of putative complexes, rank them by estimating their free energy of binding.

A plethora of techniques has been employed to address the sampling problem, ranging from complete enumeration of the relative orientations of the rigid binding partners to methods based on effective energy gradients to determine new atomic positions.

For the scoring problem, the free energy of binding for a certain putative complex has to be estimated. These estimates should take into account both the enthalpic and the entropic contributions to the free energy of binding. Although free energy perturbation[38–41] techniques employing molecular dynamics simulations for sampling can be applied to protein–protein binding,[42] they are too demanding in terms of computational resources to be of practical value for docking problems at present. Instead, a common strategy is to use heuristic techniques to estimate the entropic component of the free energy of binding (e.g., desolvation energies are parameterized by atomic solvation parameters[43]), and simplified potential energy functions are used to estimate the enthalpic component. A simple scoring function might be given by the number of intermolecular contacts shorter than a given threshold distance or the number of van der Waals (vdW) clashes in the protein–protein interface. A slightly more complex scoring function would account for the chemical nature of the interacting atoms of the proteins as well as their spatial

arrangement. One common formulation of such a scoring function is based on atomic solvation parameters (ASP):

$$V_{\text{asp}} = \sum_{i=1}^{N} \sigma_i A_i, \qquad [1]$$

where the sum over all atoms i involves the atomic solvation parameter σ_i and the solvent-accessible surface A_i of that atom.

A more detailed approach to scoring is to compute a molecular mechanics (MM) potential energy for the enthalpic part of the binding free energy; typically, such a function contains the following terms:

$$V_{\text{MM}} = V_{\text{bonds}} + V_{\text{angles}} + V_{\text{dihedrals}} + V_{\text{impropers}} + V_{\text{nonbonded}} \qquad [2]$$

The bond-stretching term V_{bonds} has the following form:

$$V_{\text{bonds}} = \sum_{\text{bonds}} K_b(b - b_0)^2 \qquad [3]$$

where K_b is the force constant of the harmonic spring representing a covalent bond, b_0 is the bond's equilibrium length, and b is the bond's length as derived from the Cartesian coordinates of the atoms involved.

The bond angle term V_{angles}

$$V_{\text{angles}} = \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 \qquad [4]$$

describes bending of the angles between covalently bonded atoms. The parameters θ and θ_0 are the bond angle and its equilibrium value, and K_θ is the bending force constant.

The dihedral term $V_{\text{dihedrals}}$

$$V_{\text{dihedrals}} = \sum_{\text{dihedrals}} K_\chi[1 + \cos(n\chi - \delta)] \qquad [5]$$

describes torsion about the bond between the second and third of four sequential, covalently bonded atoms. The number of local energy minima is determined by n, and their corresponding internal dihedral angles χ are determined by the offset δ. The associated force constant is K_χ.

A heuristic potential term to keep peptide bonds and aromatic ring structures planar is given by the improper dihedral term

$$V_{\text{impropers}} = \sum_{\text{impropers}} K_\phi(\phi - \phi_0)^2 \qquad [6]$$

Here, a pseudo-dihedral angle is defined similarly to a dihedral angle and is denoted by ϕ. The respective force constant is K_ϕ.

The Coulombic and vdW interactions are represented by the nonbonded potential term

$$V_{\text{nonbonded}} = \sum_{\text{nonbonded}} \left(\varepsilon \left[\left(\frac{R_{\min_{ij}}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\min_{ij}}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\mathrm{D} r_{ij}} \right) \qquad [7]$$

Here, the interatomic distance between atoms i and j is r_{ij}. The depth of the Lennard-Jones well describing the vdW interaction is given by ε, and the location of its minimum is given by $R_{\min_{ij}}$. The electrostatic interaction is represented by charges q_i, q_j, the interaction of which is modified by an effective dielectric constant D.

Electrostatic desolvation effects upon binding can be scored by invoking a Poisson–Boltzmann (PB) formulation of electrostatic binding energies[44] of the form

$$V_{\text{PB}} = \frac{1}{2} \int \rho \boldsymbol{\phi} \, d^3 r \qquad [8]$$

where ρ is the total charge density and the electrostatic potential $\boldsymbol{\phi}$ is given by solution of the PB equation

$$-\nabla \cdot \varepsilon \nabla \boldsymbol{\phi} = \rho^f + \lambda \sum_i c_i q_i \exp\left(-q_i \boldsymbol{\phi} / k_{\text{B}} T\right) \qquad [9]$$

Here, ρ^f denotes the fixed charge distribution of the proteins, and ε the position-dependent permittivity. The term $-\nabla \cdot \varepsilon \nabla \boldsymbol{\phi}$ represents the divergence of the displacement field and, according to Gauss's law, is equal to the charge density. The ionic charge density is given by $\lambda \Sigma_i c_i q_i \exp\left(-q_i \boldsymbol{\phi} / \kappa_{\text{B}} T\right)$, where ions of charge q_i have a bulk concentration of c_i, and $\lambda = 1$ in ion accessible regions and $\lambda = 0$ elsewhere.

Although the sampling and the scoring problems are conceptually different in nature, their solutions are not independent. The complexity of the energy landscape, which is dependent on factors such as the smoothing and resolution of the scoring function, will influence the extent and type of sampling required. Although efforts to solve the sampling and scoring problems are intertwined, we will describe current approaches to solve the two problems separately. Summaries of the methods discussed and their respective technical components are given in Tables 1 and 2. Because the run times given in these tables for the individual methods were obtained for computations on different computer hardware and example protein–protein complexes, they should be considered only as rough relative estimates.

Table 1 Rigid-Body Docking Methods[a]

Reference	Sampling Method	Scoring Method	DOF	Resolution	Runtime[b]
Cherfils et al.[45]	MCSA/EM	vdW + ES	6	One sphere per residue	10 h
Gabdoulline and Wade[46]	BD	Excluded volume and effective charge PB ES	6	Atomic	10 h
Ouporov et al.[47]	BD	Excluded volume and PB ES	6	Atomic	100 h
Gabb et al.[48]	Complete enumeration (FFT)	vdW + ES and filtering	6	1–2 Å	10 h
Meyer et al.[49]	Complete enumeration (FFT)	Hydrogen-bond filtering and shape matching	6	ca. 1.5 Å	1 h
Vakser[50]	Complete enumeration (FFT)	Shape matching	6	3–7 Å	1 min
Ritchie and Kemp[51]	Complete enumeration (polynomial expansion)	Electrostatic and shape matching	6	Depends on degree of polynomial used	1 h
Fischer et al.[52]	Geometric hashing	Shape matching	6	ca. 1 Å	1 h
Hendrix and Kuntz[53]	Geometric hashing/ site-point matching + EM	Shape matching + MM	6	ca. 1 Å	1 h

[a]Abbreviations of technical terms can be found in Ref. 1.
[b]The indicated approximate runtimes cannot be compared quantitatively, as the authors used different computing equipment. We present only orders of magnitude to give the reader a rough idea of what time scales are involved.

In the following two sections, we report current methods for rigid- and flexible-body docking. After describing the algorithmic ideas, we make a rough classification and highlight the limits and capabilities of the respective methods.

Rigid-Body Docking

The key assumption of rigid-body docking approaches is that any conformational changes in the interaction partners upon binding can be neglected or modeled implicitly. Implicit representations, such as reduction of atomic radii or lowering of the resolution of the model by discretization of the proteins on to grids with spacings of several angstroms, may be used to account for limited motions such as side-chain rearrangements at the interface. Although rearrangements upon binding are indeed limited to side chains in many of the protein–protein complexes found in the Protein Data Bank, even for such complexes, rigid-body docking of unbound structures at high resolution can result in the correctly docked complex being missed.[48]

Table 2 Flexible Docking Methods[a]

Reference	Sampling Methods	Scoring Methods	DOF	Resolution	Runtime[b]
Sandak et al.[54]	Geometric hashing	Shape matching	$6 + 3N_{\text{hinge}}^{\text{discrete}\ c}$	≈2 Å	10 min
Jackson et al.[55]	Mean-field equation	vdW + ES and desolvation energy	$6 + N_{\text{rotamers}}^{\text{discrete}\ d}$	ca. 1 Å	10 min
Althaus et al.[56]	Optimization (DEE, LP) and EM	MM and MM + desolvation	$N_{\text{rotamers}}^{\text{discrete}\ d}$, $N_{\text{atoms}}{}^{f}$	Atomic	1 h
Weng et al.[57]	Complete enumeration (ASP/PB) and EM	Effective free energy or MM	$N_{\text{torsions}}^{\text{discrete}\ e}$, $N_{\text{atoms}}{}^{f}$	Atomic	Not given
Abagyan et al.[58]	MCSA/EM	MM + desolvation or MM	Arbitrary	Atomic	10 h
Ullmann et al.[59]	MCSA + MD/EM	Colulombic + MM + PB/SA	6, then $3N_{\text{atoms}}{}^{f}$	Atomic	Not given

[a] Abbreviations of technical terms can be found in Ref. 1.
[b] The indicated approximate runtimes cannot be compared quantitatively, as the authors used different computing equipment. We present only orders of magnitude to give the reader a rough idea of what time scales are involved.
[c] The total number of hinges assumed for the flexible protein.
[d] The total number of rotameric states for all residues considered.
[e] The total number of variable torsional angles.
[f] The number of atoms.

Low-resolution rigid-body docking can be applied to a wide range of protein–protein complexes. These must display low-resolution recognition, a property found by Vakser et al.[60] to be present in more than half of the complexes in a set of ca. 500 nonredundant protein–protein complex crystal structures.

Gradient-Based Techniques

A variety of techniques has been developed to explore the conformational space spanned by the six degrees of freedom considered in rigid-body docking. In early work, Cherfils et al.[45] parameterized protein–protein orientations by the protein separation distance along the line connecting the centers of mass and five variables for the respective angular orientations. For each set of these five angular variables, the optimal separation distance was computed by a simple scheme resulting in the smallest mutual separation without steric clashes. Steric clashes were computed with a low-resolution model in which each residue is replaced by a single sphere. For each given set of angular parameters, the closest separation distance was determined, and an approximate interaction energy comprising simple repulsive vdW-like and surface burial terms was computed. The five-dimensional space of angular variables was

searched by performing Monte Carlo simulated annealing and energy minimization in random subspaces. Sets of candidate complexes generated by repeated application of this searching technique were further energy minimized using an atomic level molecular mechanics force field. With this method, complexes close to the experimental structure could be generated when starting from the respective bound conformations. However, docking of the unbound protein structures in two test cases resulted in significant loss of interprotein hydrogen bonds, thus indicating the shortcomings of the method in dealing with minor induced fit. Recently, this approach was used by Janin[61] to generate transition state complexes for the barnase–barstar system in order to study the kinetics of this protein–protein binding event.

An alternative to the energy minimization approach of Cherfils et al.,[45] is stochastic or Brownian dynamics (BD)[62] in which bimolecular association trajectories are simulated, thus allowing for a time interpretation of the binding process. In the work of Gabdoulline and Wade,[46,63] BD simulations are used to determine the association rates of proteins, given a known structure of the complex of the two proteins. Instead of generating a single bound complex, clusters of diffusional encounter complexes[21] are obtained by simulating many trajectories. Intermolecular excluded volume and electrostatic interactions are used to derive the forces. An effective charge model[64] is used to compute approximate Poisson–Boltzmann electrostatic forces in an efficient manner. Brownian dynamics can also be used to predict the structures of electrostatically favorable docked protein–protein complexes, as was recently demonstrated by Ouporov et al.[47]

Complete Enumeration

Instead of finding the optimal six-dimensional coordinate set by gradient-based methods, an alternative approach is to discretize all dimensions and locate the approximate optimum complex by complete enumeration. With the number of points in each direction being N, a naive enumeration would have a computational complexity of $O(N^6)$. This approach is illustrated by recent work of Palma et al.,[65] who use a binary representation of each protein on a cubic grid. For a given grid, each grid point is assigned a value of 0 when no atom is close, and 1 otherwise. For a given relative orientation of two proteins, their geometric fit can then be determined by efficient binary computations of the representative grids. The best preliminary orientations are then scored using electrostatic and surface matching, and pairwise residue potential terms. These scores are combined by means of a neural network to derive a single score for a given orientation.

An alternative for identifying the optimal rigid-body orientation is provided by the convolution-based strategy pioneered by Katchalski-Katzir et al.[66] If the scoring function is cast into the shape of a mathematical correlation function, the complexity can be reduced to $O(N^3 \log(N^3))$. Figure 4 illustrates the main idea of the algorithm. First, the three angular variables are separated

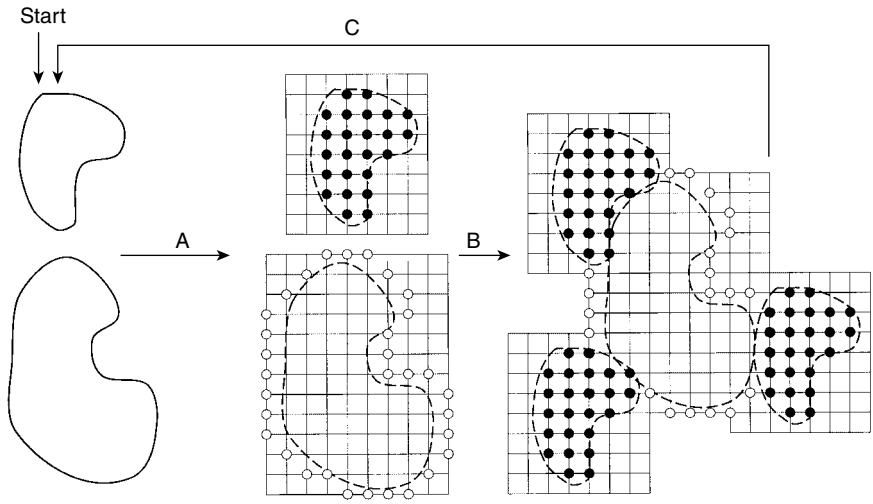

Figure 4 Illustration of the FFT-based rigid body docking approach. Initially, the two proteins in a particular relative orientation are discretized during step **A** of the algorithm. Then, all possible translations are scanned by superposing the smaller protein's grid onto the larger one's without changing the relative orientation of the proteins (this scan is performed very efficiently by means of a FFT-based convolution formalism). For each such superposition, a shape-based score is computed. In step **B**, a small number of the highest scoring translations are kept and collected for later evaluation. During step **C**, a new relative orientation of the proteins is chosen with a given discretization in the angular degrees of freedom. After scanning all relative orientations, a certain number of the highest scoring rotations/translations is screened with more elaborate scoring methods.

from the three translational variables. For each given rotation, the two proteins are discretized onto two cubic grids. The grid points are assigned integer values according to their proximity to the protein atoms. Loosely speaking, the two proteins in a given relative orientation are surrounded by a cubic lattice. For every single possible translation of one protein lattice with respect to the other in this orientation, the scoring function is evaluated by performing a pointwise multiplication of the overlapping lattices and summing up the single contributions. To scan all possible translations for a given rotation, this correlation function can be computed very efficiently by means of fast Fourier transforms (FFTs).[67] By Fourier transforming the geometric or electrostatic properties of both proteins on a grid and computing the product of one transform with the conjugate of the other, a matrix will be obtained. Reverse Fourier transforming this matrix results in the scoring function values for all possible translations at this particular orientation. Only a set of the best scores and their associated translations is retained and compared to the respective sets from other orientations. Besides reducing the complexity of the algorithm by using discrete Fourier transforms for the correlation computation, these

methods offer the potential for computational efficiency. Whereas the FTDOCK package of Gabb et al.[48] has a parallel implementation of the FFTs on a shared-memory architecture, the DOT program[68,69] utilizes a farmer/worker approach to distribute sets of orientations between worker nodes in a distributed memory environment. In this approach, one control processor (the farmer) distributes tasks to identical worker processes, which compute and return the result.

Different types of scoring function can be applied in combination with convolution-based docking. In FTDOCK, a geometric matching score is used to generate complexes, whereas electrostatics is applied only as a posterior binary filter. In DOT, electrostatic and geometric scores are used in conjunction during generation of complexes. Electrostatic computations are performed in FTDOCK with a Coulombic formalism, a distance-dependent dielectric constant, and point charges interpolated on a grid. In contrast, DOT can utilize the output of software suites such as UHBD[70] or DelPhi[71] that compute electrostatic potentials as solutions to the Poisson-Boltzmann equation.[44] Both FTDOCK and DOT use an atomic representation of the proteins involved to derive the discretized protein grid. All fine-grained convolution-based docking approaches are demanding with respect to memory usage: to perform the FFTs efficiently, grids of N^3 complex numbers must be kept in memory. For large proteins, this necessitates minimum grid spacings larger than 1 Å, and therefore the protein might not be represented at the resolution desired.

An efficient implementation, useful for low-resolution docking, has been provided by Vakser and Aflalo in their GRAMM package.[72] This program provides an implementation of the approach of Katchalski-Katzir et al.,[66] employing a coarse grid with a grid spacing up to 7 Å to represent the interacting proteins.[50] Although a grid spacing of 7 Å might not be useful for detecting the impact of single-residue mutations on protein–protein association, this approach can help to detect the proper binding hemispheres for large sets of proteins.[60] A modified version of the algorithm considers only hydrophobic contacts for the geometric surface matching;[72] this modification was shown to improve the scores of the correct positive results compared to the false positives. Low-resolution docking, as implemented in GRAMM, appears strongly influenced by the choice of the parameters describing the protein shape, including the grid spacing;[73] care must be taken to find the optimal parameter set for a given protein–protein docking problem.

In all of the correlation-based docking approaches described above, a complete enumeration of the rotational angle space is performed. Meyer et al.[49] modified the basic technique by filtering to select those relative rotations that lead to formation of at least two simultaneous intermolecular hydrogen bonds. This filtering is achieved by storing the directionality of the hydrogen-bond donor and acceptor sites in a translation-independent way. After performing the translational scans in the conventional manner, the

rotational parameters of the complexes, filtered to satisfy the hydrogen-bonding criteria, are further refined using a derivative-free minimization technique to optimize the geometric fit of the two proteins.

While retaining the formulation of the geometric and chemical fit as a mathematical correlation, Ritchie and Kemp[51] use a series expansion to describe the shapes of the proteins. This mathematical description uses generalized Laguerre polynomials and spherical harmonics.[67] Ritchie and Kemp formulate the scoring function in one distance and five angular variables. With their mathematical framework, translations and rotations of the proteins can be reduced to a recomputation of these expansion coefficients, which can be carried out efficiently. After a coarse-grained global scan of all the variables and evaluation of the peaks in the geometric part of the scoring function, a fine-grained scan of the interesting parts of the configurational space is performed using the full geometric and electrostatic contributions. In contrast to the FFT-based docking schemes described above, the memory requirements of the Ritchie–Kemp algorithm are relatively modest and do not depend on the size of the proteins. As this technique is based on scanning the entire configurational space, it offers the potential for parallelization as well. Moreover, it becomes very efficient when the binding site on one of the proteins is known. In the spherical polar approach, this can be specified by a simple constraint in one or two angular degrees of freedom prior to evaluating correlations.

Geometric Hashing Techniques

As an alternative to completely enumerating all relative orientations, Fischer et al.[52] reported the development of a computer-vision-based technique called geometric hashing. This scheme is conceptually similar to that utilized in the DOCK program[53,74] which tries to find matches in the internal distances between spheres[74] or shape-based site-points,[53] which represent the surface features of the proteins.

By construction, the geometric hashing method generates only complexes with at least partial geometric fit, which is achieved by storing critical point information (position or surface normals) with respect to locally defined coordinate systems (derived from sets of critical points themselves). Utilizing efficient data structures, matches of these local coordinate systems are ranked by the number of ligand and receptor critical points that are in close proximity. For the highest scoring pairs of local coordinate systems, the geometric transforms that bring them into overlap are computed in an iterative fashion. After filtering the resulting set of possible ligand transformations by a scoring scheme (e.g., based on buried surface area), a set of candidate complexes is generated. The complexity of the algorithm is $O(N^3)$, with N being the number of receptor critical points. In the implementation of Fischer et al.,[52] the number of receptor critical points is about 0.3–1 times the number of surface atoms on the receptor (and ranges from ca. 60 to 600 for the set of representative proteins that they studied).

Flexible Docking

In this section, we describe some of the current methods for treating protein flexibility explicitly during docking. We order the methods according to the level of flexibility covered, starting with large-scale motion between rigid domains and ending with full atomic flexibility.

Geometric Hashing Techniques

One particular type of large-scale motion, hinge bending, is treated in a modified[75] version of the geometric hashing algorithm of Nussinov and coworkers.[52,54] In the modified approach (illustrated in Figure 5), the user must specify the position of the hinge(s) (several hinges are possible) in the "ligand" protein prior to starting the docking procedure. Triangles spanned by triples of critical points of both proteins are precomputed; each such triangle is characterized by its side lengths. The location and orientation of such a ligand triangle with respect to a hinge is stored as a translation vector and a rotation transform during the preprocessing phase of the algorithm. In the recognition (i.e., sampling) phase of the algorithm, the triangles spanned by critical points on the surface of the receptor protein are matched to those computed for the flexible ligand protein. When a match is found, application of the stored transformation to the current trial ligand protein coordinate frame results in the putative hinge position and orientation. This hinge position and orientation is stored and receives a "vote." When different parts of the ligand protein vote for the same hinge location, the entire protein can be reconstructed in a bent conformation. After such a reconstruction, candidates are filtered with respect to steric clashes and scored by the percentage of the ligand protein's atoms in contact with the receptor protein. The algorithm naturally extends to proteins with several hinges. Sandak et al.[75] docked the M13 peptide to calmodulin. Although the algorithm can recover the docked complex when starting from the single-bound conformations, the results for the unbound protein conformation are significantly worse. One reason could be that the hinge motion implemented in the algorithm does not describe well the actual large-scale conformational changes of calmodulin.

Conformational Search Techniques

As part of the suite of codes for flexible docking developed by Sternberg et al.,[76] the MULTIDOCK code[55] optimizes the conformations of the interfacial side chains after obtaining the predicted docked orientation from a rigid-body docking procedure. The approach entails iterations until convergence of the interaction energy is attained. Each iteration comprises two steps: determination of the best rotamers for interfacial side chains, and then small-scale rigid-body energy minimization. The energy terms used include mean-field terms for the bonded, electrostatic, van der Waals, and solvation contributions. Whereas bonded interactions are only modeled approximately, by

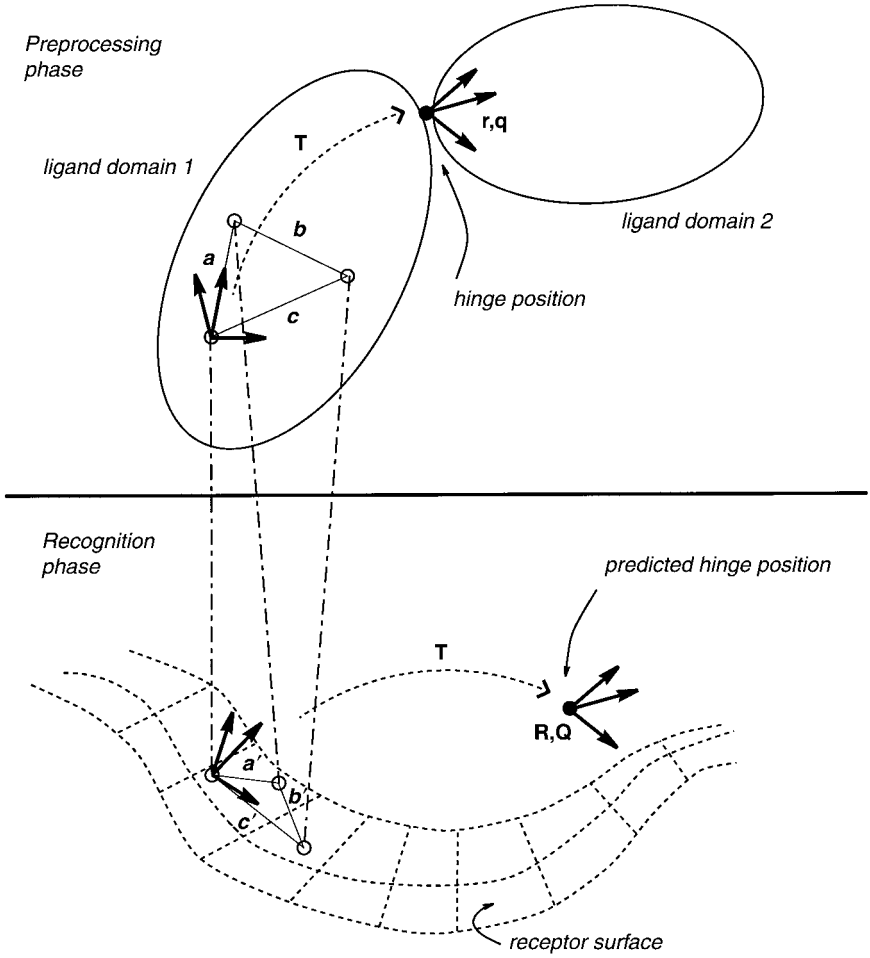

Figure 5 Illustration of the geometric hashing approach to flexible docking. During the preprocessing phase, the so-called ligand (i.e., one of the proteins) surface is covered by triangles spanned by connecting all interest points. For each triangle, such as that shown with sides of lengths a, b, and c, a mathematical transform **T** is computed. This transform superposes the triangle's coordinate frame (represented by the bold arrows) onto the hinge coordinate frame located at position **r** and orientation **q**. In the recognition phase, a search is performed for similar receptor triangles (i.e., spanned by the interest points on the receptor surface). When a similar triangle is found on the receptor, such as that shown with sides a', b', and c', the ligand's interest point's local coordinate system is constructed at the corresponding point on the receptor surface. By means of the transform **T,** a putative hinge location **R** and orientation **Q** are predicted. When a similar prediction is made for a hinge location and orientation from the ligand's second domain, the ligand can be constructed in a bent conformation close to the receptor.

assigning each rotamer an energy derived from observed side-chain rotameric preferences in a rotamer library, the solvation term takes into account detailed interactions with water molecules which are treated as soft sphere Langevin dipoles on a grid that make vdW, electrostatic, and hydrophobic interactions with the protein. The interaction with bulk solvent beyond the Langevin dipoles grid around the protein is modeled by the Kirkwood equation (which gives an expression for the electrostatic interaction energy of a solute's monopole and dipole moments with a surrounding solvent dielectric continuum). The approach[55] shows good improvement of the ranking of the complexes proposed by rigid-body docking approaches, but encounters difficulties when the complexes studied exhibit extensive side-chain rearrangements or backbone motion. In addition, charged residues in nonrotameric states cannot be modeled by the algorithm.

Althaus et al.[56] use combinatorial optimization techniques to optimize the discrete rotamers of interfacial side chains after generating candidate complexes with a Fourier-correlation based rigid-body docking approach. During side-chain optimization, the AMBER[77] molecular mechanics energy function is minimized. In the final scoring of the set of putative optimal complexes, a more elaborate energy function is used. This energy function includes continuum electrostatics, surface burial, protein conformational entropy, and vdW terms.

Two further energy functions based on Poisson-Boltzmann electrostatics and atomic solvation parameter (ASP)-based parameterizations of the solvation free energy changes[43] were evaluated by Weng et al.[57] in the context of side-chain optimization after rigid-body docking.

Gradient-Based Techniques

Whereas the side-chain optimization techniques described above confine the side-chain conformations to discrete tabulated rotameric states, the approach of Abagyan et al.[58,78] considers the side-chain torsions as degrees of freedom in an optimization procedure. In their procedure, named "internal coordinate mechanics," a two-tier energy function is minimized with respect to arbitrary internal coordinates of the interacting proteins. The minimization consists of biased distortions of the conformations, followed by local energy minimization with respect to a molecular mechanics potential energy function. The new conformation is scored by an elaborate energy function including surface area dependent terms and a non-Coulombic electrostatic term accounting for dielectric heterogeneity. The resulting score is used to accept the new conformation on the basis of a Metropolis criterion (i.e., the probability of acceptance is given by $\min[1,\exp(-\Delta E)]$, where ΔE is the change in energy on going from the current to the new conformation). The approach conceptually allows for limited backbone flexibility as well. The approach has been used successfully for prediction of the lysozyme–HyHel5 antibody complex starting with lysozyme in the unbound conformation.[78]

Ullmann et al.[59] developed a simulation-based approach that they used to dock plastocyanin and cytochrome f. The method consisted of three stages. First, the proteins were held rigid in their unbound conformations and their Coulombic interaction energy was minimized by running thousands of Monte Carlo trajectories in which the temperature was gradually cooled from 300 to 0 K. Second, complexes with low Coulombic energy were refined by molecular dynamics simulation using the CHARMM force field[79] with a spherical shell of explicit water molecules around plastocyanin and all atoms within the shell treated as mobile. The system was simulated at 300 K for 200 ps and then gradually annealed to lower temperatures and energy minimized. Final low-energy conformations were rescored with an energy function containing Poisson-Boltzmann electrostatic and solvent accessible surface area dependent terms. The configurations with the most favorable Coulombic interactions did not have the most favorable total energies, thus indicating the importance of considering hydration and conformational flexibility. Several docked complexes were obtained by this procedure and could be used to interpret data on electron transfer between the proteins. Notably, the configuration with the most efficient electron transfer pathway did not have the lowest energy.

EXAMPLE

In this section, we illustrate some of the methods described above (along with their limitations) when trying to dock two proteins from their respective unbound structures. This description does not aim at providing the optimal workflow of such a project; we believe that right now, there is no one optimal methodology for all classes of proteins. Instead, the type of docking experiment, the availability of experimental data, and the expected occurrence of large-scale motion for a particular protein will determine which selection of the approaches described above is likely to be feasible. In our example, we describe attempts at docking barnase, an experimentally well-characterized extracellular ribonuclease, to its intracellular inhibitor, barstar.[80–82] The limited degree of induced fit and wealth of experimental data, both kinetic and mutational,[24,29,83] might suggest that this particular docking problem would be easy to solve with a rigid-body docking approach. However, we show that even such limited flexibility has an important impact on the results of rigid-body docking. In our example, we aim to give the reader an idea of how the applicability of a certain method could be assessed. All in all, the scenario we have in mind is that of a researcher with plenty of time to dock a single protein–protein complex. In this example, we do not address the problem of detecting binding modes in large databases of proteins in an automated fashion.

Estimating the Extent of Conformational Change upon Binding

Before applying a certain published method, it is worthwhile to estimate the extent to which large-scale motion will occur for the proteins involved. One approach is to decompose correlated motions of atoms into so-called modes. This terminology stems from normal mode analysis[84] of mechanical systems, which is based on evaluation of potential energy derivatives for mechanical systems. Approximate techniques such as essential dynamics[85] determine these modes from the positional fluctuations of selected atoms around their mean in an ensemble of conformations by solving an eigenvalue problem. Here, the eigenvectors of the position fluctuation matrix represent independent correlated motions of sets of atoms. As this technique does not require evaluations of potential energy derivatives or extensive energy minimization, it is computationally much more applicable to an approximate prediction of the large-scale motions of a protein than normal mode analysis. Whereas the conformational ensembles in the original essential dynamics technique are generated by MD, the Concoord method developed by de Groot et al.[86] perturbs a given protein structure on the basis of experimentally observed positional fluctuations. Thus, a conformational ensemble for use in an essential dynamics study can be generated at negligible computational cost. In Figure 6, the first collective mode of barnase is predicted by analyzing such a Concoord ensemble of conformations. Alternative techniques, such as the detection of motion in a Gaussian network model[87] or of independent domain motions,[88] differ in how they represent the protein and its potential energy, but represent efficient tools for predicting the large-scale motions of the associating proteins.

In Figure 6, the first essential mode of barnase's is displayed. One particularly flexible region is located around barnase's arginine-59 which, from mutational studies,[19,89] is known to contribute a large part of the free energy of binding. As this particular region is part of the binding interface, it is not clear whether reasonable predictions for the complex can be obtained by a rigid-body docking approach starting from the unbound structures. Methods to assess qualitatively the extent of large-scale flexibility can help in choosing an appropriate docking approach. In our barnase–barstar demonstration case, the superposition of bound and unbound structures reveals minor overall backbone changes upon binding; the C_α rmsd is 0.5 Å, although local rmsds of up to 2 Å are observed in the barnase–R59 region.

Rigid-Body Docking

When performing rigid-body docking with FTDOCK using the unbound protein structures, without filtering by means of biochemical data, several clusters are proposed by the program as illustrated in Figure 7. The major

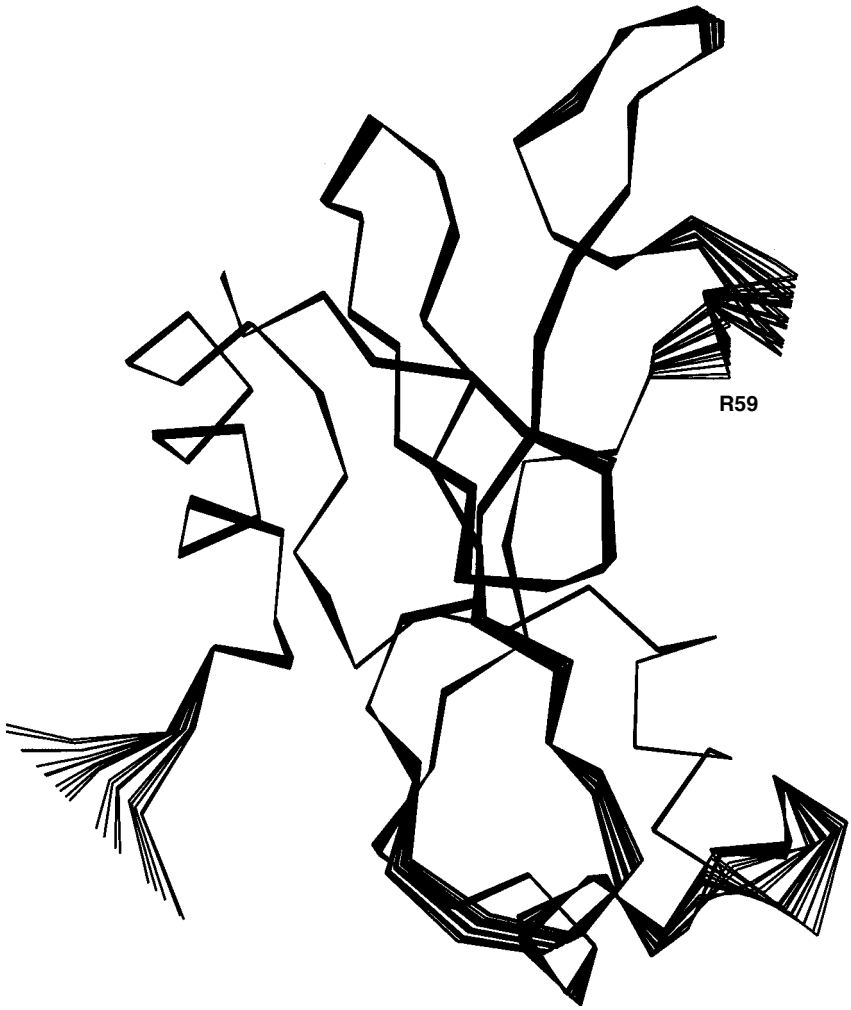

Figure 6 Barnase deformed along the first collective mode computed from perturbed conformations generated by Concoord; see text for details. The location of Arg-59 is indicated.

cluster shown in Figure 7 closely resembles the encounter complex cluster obtained by extensive Brownian dynamics simulation performed by Gabdoulline and Wade.[46,63] These diffusional encounter complex structures satisfy the requirement for formation of two native polar contacts at a distance at ca. 6 Å. Formation of these encounter complexes shows that rigid-body docking can provide starting configurations for in-depth studies of the short-range postdiffusional association process.

The best candidate structure from the FTDOCK computations only established 13% of the native contacts, as opposed to 51% when starting

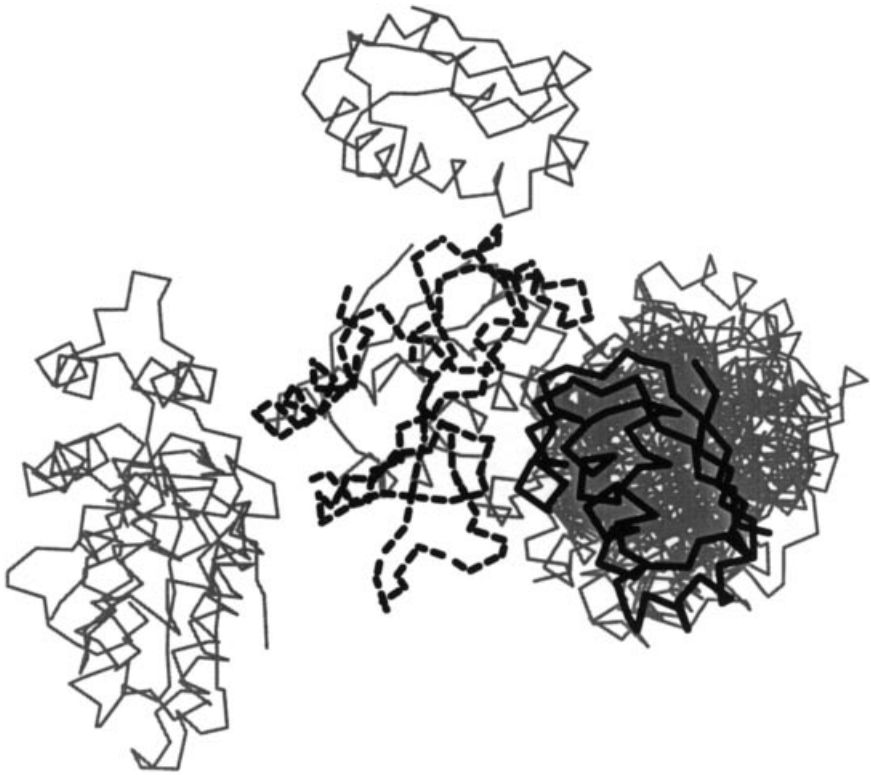

Figure 7 The 20 highest ranking putative complexes from the best 100 FTDOCK candidates obtained after fine-grained rigid-body docking of the unbound barnase and barstar structures. Barnase is shown in the middle by dashed lines, surrounded by putative clusters of barstar orientations. The dense cluster to the right contains the correct docking orientation (shown in bold).

from the separated bound conformations (data not shown). Compared to simulations starting from the unbound conformation, the percentage of established native contacts could be improved to 22% when the local loop region around barnase-R59 was deformed to its respective bound conformation while the rest of the structure was held fixed in its unbound conformation.

Although Gabb et al.[48] pointed out the shortcomings of their FTDOCK methodology when performing docking experiments with unbound protein structures, the result above shows that backbone structural variability is more important in the interface of the associating proteins. Consequently, a strategy separating backbone and side-chain flexibility might encounter significant difficulties in prediction of a bound complex from the free structures because the relevant small-scale backbone rearrangements should be taken into account explicitly.

Flexible Docking with Side-Chain Flexibility

To test the impact of localized backbone and side-chain conformational changes when using an interface refinement strategy, we performed constrained MD simulations in dihedral angle space[90] (so-called torsion dynamics) of barnase–barstar association from an artificial starting arrangement similar to those resulting from a rigid-body docking approach. In this starting conformation, the two proteins were pulled apart along a line connecting their centers of mass, leading to an rmsd of the interfacial barstar C_α atoms of 5 Å. Using different backbone conformations, only the side chains were allowed to move in this all-atom, molecular dynamics simulation.

Whereas most side-chain optimization techniques described above rely on choosing one of a set of rotamers for a given side chain, the MD-based simulations can sample the full range of side-chain torsional angles. Treating these degrees of freedom as continuous is particularly relevant for so-called hot spot residues[88] such as barnase-R59, which contribute a large part to the free energy of binding and do not occupy a standard rotameric state in the bound complex. The favorable energy arising from the nonbonded interactions outweighs the unfavorable torsional energy of the nonstandard rotameric state.

Impact of Side-Chain Conformational Changes

Barnase-R59 has been shown to be involved in an intricate network of water-mediated hydrogen-bonding interactions involving barstar and at least five ordered water molecules buried in the interface. Water molecules are rarely included explicitly in protein–protein docking studies; most approaches utilize implicit solvent approaches. In our torsion dynamics approach, the mechanical properties of the solvent were taken into account by stochastic forces acting on surface protein atoms, and the solvent dielectric properties were represented by a distance-dependent dielectric constant. Thus, we can expect that the use of this approximate model for the solvent will limit the accuracy of predictions of the conformations and interactions of side-chains involved in salt bridges or hydrogen bonds in the solvated interfacial region. As has been pointed out by Bogan and Thorn[89] and Xu et al.,[33] these residues are often the ones that contribute most to the free energy of binding.

In our simulations, we performed association studies to elucidate their role by selecting pairs of those hot spot residues that established a barnase–barstar contact in the bound complex. In these simulations, we studied what effect the side-chain conformation had on contact formation for certain hot spot residue pairs. A typical result is shown in Figure 8. Starting from the correct rotameric state for certain hot spot residue pairs with the rest of the proteins in their unbound conformations leads to improved contact formation in the initial docking phase. This effect is not observed when performing equivalent simulations with non-hot-spot interfacial residue pairs.

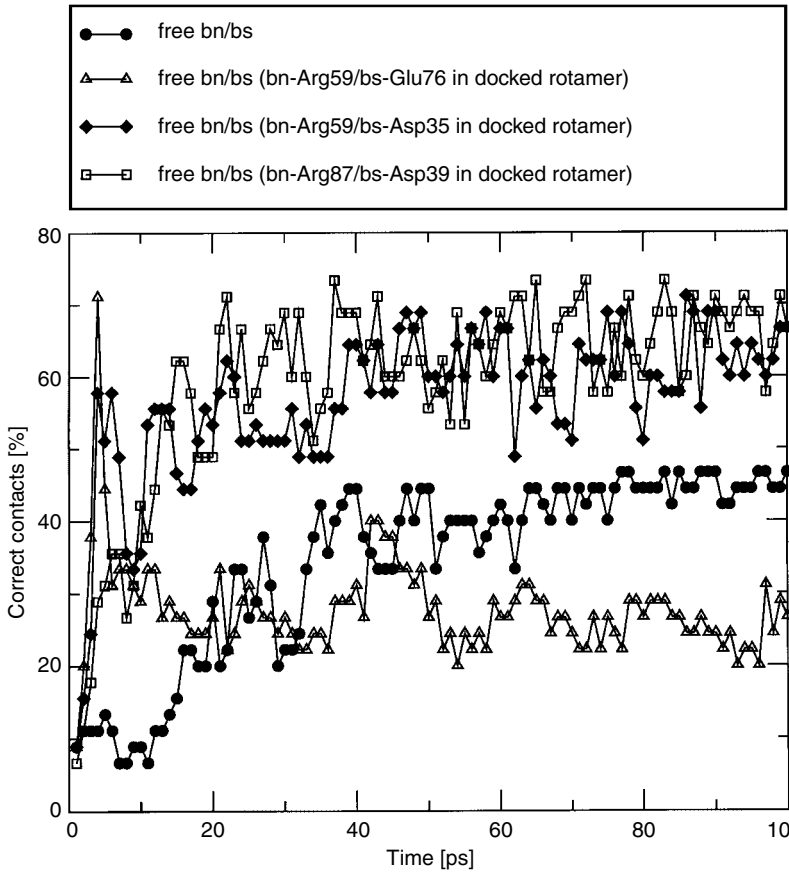

Figure 8 The impact of hot-spot residue pair conformations on contact formation during docking of barnase (bn) and barstar (bs). Contact formation (as measured by the percentage of native contacts) during the first 100 ps of torsion dynamics simulations is shown; see text for details.

A series of simulations starting from reverse initial conditions (protein structures in bound conformations with hot-spot residue pairs in unbound rotameric states) leads to impaired contact formation during the initial docking phase. These findings indicate, first, that the hot-spot residues can actively hinder or support contact formation; and second, that a side-chain refinement protocol relying on specific side-chain rotameric states may well miss crucial structural detail in the interfacial region.

Impact of Backbone Conformational Changes

From the FTDOCK simulations described above, it is obvious that small localized backbone rearrangements in the protein–protein interface can

influence the association process. To clarify the extent of this backbone effect, we performed side-chain torsion dynamics simulations using a barnase structure with a deformed loop region surrounding barnase-R59. Starting from the unbound conformation, this loop region was deformed onto the C_α trace of the respective bound structure. This amounts to a local rmsd of 2 Å, compared to the 0.5-Å overall backbone rmsd for the entire structure. Compared to prior simulations with the unbound conformations, an improvement of up to 20% established contacts could be observed for the best in a series of simulations. The difference in docking progress between the bound and unbound starting conformation was almost compensated by introducing the proper backbone conformation in this binding-site loop.

Flexible Docking with Full Flexibility

Although side-chain torsion dynamics simulations can help in understanding the influence of single residues and static backbone conformations on the binding process, they still neglect the effects arising from backbone flexibility. Standard molecular dynamics simulation techniques can be used as a sampling tool to create putative complexes. However, unless simulations are performed with suitable starting configurations close to the docked configuration, a major part of the simulation time will be spent exploring uninteresting regions of conformational space. At the other end of the spectrum, global energy minimization methods can be used to generate complexes while abandoning the physical interpretation of the binding process.

Following an intermediate path, we developed a hybrid simulation method to study the peculiarities of the barnase–barstar binding process. This hybrid method was used to study the short-range association process. Simulations start with orientations resembling encounter complexes, obtained by rigid-body docking governed by long-range electrostatic interactions. This hybrid simulation method is inspired by the genetic algorithm class of algorithms used for global optimization,[91,92] in which entire ensembles of candidates are evaluated with respect to an optimality criterion and only promising individuals are propagated to the next generation. In our ''ensemble enriching'' simulations, multiple copies of barnase interact with the mean field arising from multiple copies of barstar and vice versa. No interaction is simulated between copies of the same protein. In typical simulations, short (10 ps) simulations of 10 barnase and 10 barstar copies are performed with full atomic mobility. After each such simulation, all 100 possible complexes are scored with a heuristic energy function comprising nonbonded interaction energy, pairwise distance potentials, as well as implicit desolvation energy terms. These scores are used to derive an average fitness score for each protein copy in the ensemble. Following this scoring stage, the five lowest scoring conformations in each ensemble are discarded, and the remaining five are duplicated. This effectively leads to an enrichment of the promising candidate

conformations while retaining full flexibility and a time interpretation of the binding process. A complete docking simulation consists of several sets of these simulation and enrichment stages.

The results of simulating barnase–barstar association after 100 ps of simple multiple copy MD without ensemble enriching are shown in Figure 9. Whereas contact formation in the simple multiple copy MD simulation has leveled off after the initial 100 ps, ensemble enriching with a linear scoring function is able to shift the ensemble population toward a higher number of established native contacts. Compared to the side-chain torsion dynamics

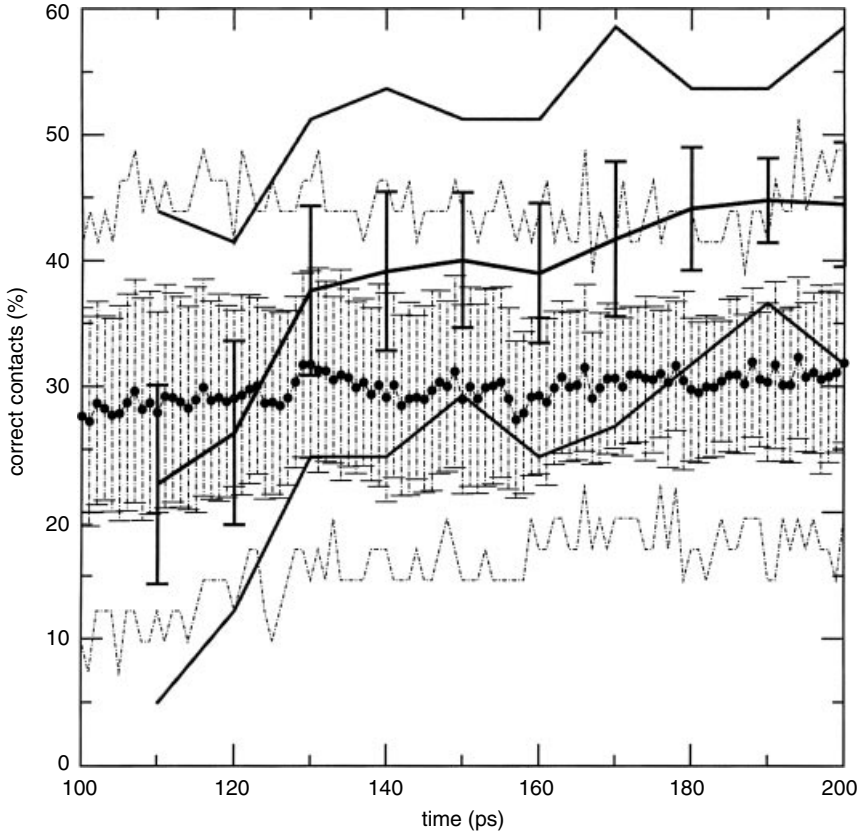

Figure 9 Formation of native contacts during a multiple copy MD simulation with and without ensemble enriching after an initial 100 ps multiple copy MD simulation; see text for details. Ensembles that each consist of 10 barnase and 10 barstar copies are simulated, giving rise to 100 putative complexes. The thick lines show the mean and standard deviations, along with the best/worst values for the ensemble enriching method. The superposed thin lines show the corresponding data for the plain multicopy MD simulations without ensemble enriching.

simulation starting from the unbound conformations of both proteins, an additional 20% of native contacts are gained when treating the simulated proteins as completely flexible. Some barnase copies with large numbers of established contacts exhibit a motion of the barnase–R59 loop region toward its bound conformation.

Although the ensemble enriching multiple copy MD approach is currently too demanding in terms of computational resources to be applied to a large set of protein–protein complexes, it shows that simulations with fully flexible proteins can be useful computational aids when no a priori limitations on flexibility should be made.

FUTURE DIRECTIONS

Computational protein–protein docking can be expected to be a very active field in the near future for several reasons. First, genomics and proteomics studies will provide large amounts of data on protein–protein binding. This binding data will often be of a qualitative kind (protein 1 interacts with protein 2), and computational studies will be necessary for detailed interpretation. Second, large-scale structural genomics studies will mean that 3D structures become available for many more protein targets of therapeutic interest. However, it is anticipated that the structure of a target protein will often be modeled by homology rather than be determined experimentally, and therefore may contain significant structural uncertainties and errors. These errors will be most apparent in loop regions, which can occur at the protein-binding interface (e.g., as in antibody–antigen complexes). Thus, these homology models are a challenge for use in docking studies, and their inaccuracies mean that either low-resolution rigid-body docking strategies or strategies that treat backbone and side-chain flexibility explicitly will be needed. Third, computational advances can be expected in both the sampling and the scoring aspects of docking methods.[93]

A range of computational docking methods will be required in this context and hierarchical approaches may be adopted so that computations are suited to the type of data available and the questions to be answered. For example, screening of large protein data sets for protein–protein binding requires computationally efficient approaches such as sequence-based methods and low-resolution rigid-body docking. These will be supported by techniques to perform knowledge-based 3D structural searches of protein–protein binding sites.[94] When more experimental evidence about a particular complex is available (such as binding free energy differences from mutagenesis studies), the researcher can use these as conditions to filter proposed complexes from a low-resolution rigid-body docking approach. Then, a detailed atomistic model can be constructed and molecular flexibility treated.

Computational developments in scoring functions can be expected to focus particularly on solvation. There is presently considerable research effort devoted to the development of new implicit solvation models (see, e.g., Ref. 95) and many of these models should be applicable to docking. Techniques will also be developed to include treatment of explicit water molecules in the docking procedures. Currently, this can only be done during refinement of docked structures. In the small-molecule docking field, ways are being developed to consider interfacial water molecules explicitly during docking.[96,97] Related approaches can be expected to be developed for protein–protein docking.

As more computational docking methods are developed, stringent comparisons of docking strategies, as in the Critical Assessment of Structure Prediction (CASP) project (http://predictioncenter.llnl.gov/),[98,99] will be needed to evaluate which method is better suited for which problem class. Although CASP does not include a docking section any longer, other forums such as the Critical Assessment of Techniques for Free Energy Evaluation (CATFEE) project (http://uqbar.ncifcrf.gov/~catfee/) may provide objective blind prediction tests for certain aspects of docking strategies. A complementary approach to compare and benchmark docking techniques is to establish a benchmark set of protein–protein docking problems against which new methods can be routinely tested. This benchmark set should contain structurally diverse and energetically well-characterized protein–protein complexes. To test scoring functions, the experimentally determined protein–protein complexes should be distinguished from decoy complexes. Sets of decoy complexes to test scoring functions for rigid-body docking are beginning to be made available.[100]

CONCLUSIONS

Current protein–protein docking approaches can generate structures exhibiting rmsds of ca. 1–2 Å compared with experimental structures in favorable cases, such as certain enzyme–inhibitor complexes that exhibit only minor structural changes upon binding. Results are usually far less encouraging for cases where structural changes are more prominent, such as in antibody–antigen complexes. So far, protein–protein docking programs have not successfully proven themselves able to cope with large-scale structural changes. Intense efforts are underway to integrate flexibility at the backbone level into the current approaches to improve the sampling procedures.

Fueled by the challenges of proteomics and structural genomics, the demand for reliable and fast computational approaches to solve the protein–protein docking problem is likely to increase. With increasing computing power and smarter algorithms, computer docking techniques should be able to meet this demand, although there is much to do.

REFERENCES

1. **ASP** atomic solvation parameters, **BD** Brownian dynamics, **DEE** dead end elimination, **DOF** degrees of freedom, **EM** energy minimization, **ES** electrostatics, **FFT** fast Fourier transform, **LP** linear programming, **MCSA** Monte Carlo simulated annealing, **MM** molecular mechanics, **PB** Poisson–Boltzmann, **vdW** van der Waals.
2. M. Pellegrini, E. M. Marcotte, M. J. Thompson, D. Eisenberg, and T. O. Yeates, *Proc. Natl. Acad. Sci. U.S.A.,* **96,** 4285 (1999). Assigning Protein Functions by Comparative Genome Analysis: Protein Phylogenetic Profiles.
3. E. M. Marcotte, M. Pellegrini, H. L. Ng, D. W. Rice, T. O. Yeates, and D. Eisenberg, *Science,* **285,** 751 (1999). Detecting Protein Function and Protein–Protein Interactions from Genome Sequences.
4. M. Huynen, W. L. B. Snel, and P. Bork, *Curr. Opin. Struct. Biol.,* **10,** 366 (2000). Exploitation of Gene Context.
5. P. Uetz, L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, and J. M. Rothberg, *Nature* (*London*), **403,** 623 (2000). Comprehensive Analysis of Protein–Protein Interactions in *Saccharomyces cerevisiae*.
6. B. Küster and M. Mann, *Curr. Opin. Struct. Biol.,* 8, 393 (1998). Identifying Proteins and Post-Translational Modifications by Mass Spectrometry.
7. B. Rost, *Structure,* **6,** 259 (1998). Marrying Structure and Genomics.
8. R. Sanchez and A. Sali, *Curr. Opin. Struct. Biol.,* **7,** 206 (1997). Advances in Comparative Protein-Structure Modelling.
9. J. Skolnick and J. S. Fetrow, *Trends Biotechnol.,* **18,** 34 (2000). From Genes to Protein Structure and Function: Novel Application of Computational Approaches in the Genomic Era.
10. M. J. E. Sternberg, H. A. Gabb, and R. M. Jackson, *Curr. Opin. Struct. Biol.,* **8,** 250 (1998). Predictive Docking of Protein–Protein and Protein–DNA Complexes.
11. L. Lo Conte, C. Chothia, and J. Janin, *J. Mol. Biol.,* **285,** 2177 (1999). The Atomic Structure of Protein–Protein Recognition Sites.
12. S. Jones and J. M. Thornton, *J. Mol. Biol.,* **272,** 133 (1997). Prediction of Protein–Protein Interaction Sites Using Patch Analysis.
13. T. O. Fischmann, G. A. Bentley, T. N. Bhat, G. Boulot, R. A. Mariuzza, S. E. Phillips, D. Tello, and R. J. Poljak, *J. Biol. Chem.,* **266,** 12915 (1991). Crystallographic Refinement of the Three–Dimensional Structure of the FabD1.3–Lysozyme Complex at 2.5 Å Resolution.
14. B. C. Braden, B. A. Fields, and R. J. Poljak, *J. Mol. Recognit.,* **8,** 317 (1995). Conservation of Water Molecules in an Antibody–Antigen Interaction.
15. J. Janin, *Struct. Fold. Des.,* **7,** R277 (1999). Wet And Dry Interfaces: The Role of Solvent in Protein–Protein and Protein Recognition.
16. D. G. Covell and A. Wallqvist, *J. Mol. Biol.,* **269,** 281 (1997). Analysis of Protein–Protein Interactions and the Effects of Amino Acid Mutations on Their Energetics. The Importance of Water Molecules in the Binding Epitope.
17. J. H. Lakey and E. M. Raggett, *Curr. Opin. Struct. Biol.,* **8,** 119 (1998). Measuring Protein–Protein Interactions.
18. T. Clackson and J. A. Wells, *Science,* **267,** 383 (1995). Hot Spot of Binding Energy in a Hormone–Receptor Interface.
19. G. Schreiber and A. R. Fersht, *J. Mol. Biol.,* **248,** 478 (1995). Energetics of Protein–Protein Interactions: Analysis of the Barnase–Barstar Interface by Single Mutations and Double Mutant Cycles.

20. Y. Wang, B. J. Shen, and W. Sebald, *Proc. Natl. Acad. Sci. U.S.A.,* **94,** 1657 (1997). A Mixed-Charge Pair in Human Interleukin 4 Dominates High-Affinity Interaction with the Receptor Alpha Chain.
21. R. R. Gabdoulline and R. C. Wade, *J. Mol. Recognit.,* **12,** 226 (1999). On The Protein–Protein Diffusional Encounter Complex.
22. T. Selzer, S. Albeck, and G. Schreiber, *Nat. Struct. Biol.,* **7,** 537 (2000). Rational Design of Faster Associating and Tighter Binding Protein Complexes.
23. K. P. Murphy, *Med. Res. Rev.,* **19,** 333 (1999). Predicting Binding Energetics from Structure: Looking Beyond ΔG Degrees.
24. C. Frisch, G. Schreiber, C. M. Johnson, and A. R. Fersht, *J. Mol. Biol.,* **267,** 696 (1997). Thermodynamics of the Interaction of Barnase and Barstar: Changes in Free Energy Versus Changes in Enthalpy on Mutation.
25. S. P. Edgecomb and K. P. Murphy, *Curr. Opin. Biotechnol.,* **11,** 62 (2000). Structural Energetics of Protein Folding and Binding.
26. B. M. Baker and K. P. Murphy, *J. Mol. Biol.,* **268,** 557 (1997). Dissecting the Energetics of a Protein–Protein Interaction: The Binding of Ovomucoid Third Domain to Elastase.
27. B. M. Baker and K. P. Murphy, *Methods Enzymol.,* **295,** 294 (1998). Prediction of Binding Energetics from Structure Using Empirical Parameterization.
28. B. M. Baker and K. P. Murphy, *Biophys. J.,* **71,** 2049 (1996). Evaluation of Linked Protonation Effects in Protein Binding Reactions Using Isothermal Titration Calorimetry.
29. G. Schreiber and A. R. Fersht, *Biochemistry,* **32,** 5145 (1993). Interaction of Barnase with Its Polypeptide Inhibitor Barstar Studied by Protein Engineering.
30. K. A. Xavier and R. C. Willson, *Biophys. J.,* **74,** 2036 (1998). Association and Dissociation Kinetics of Anti–Hen Egg Lysozyme Monoclonal Antibodies Hyhel-5 and Hyhel-10.
31. A. J. McCoy, C. V. Epa, and P. M. Colman, *J. Mol. Biol.,* **268,** 570 (1997). Electrostatic Complementarity at Protein/Protein Interfaces.
32. N. Horton and M. Lewis, *Protein Sci.,* **1,** 169 (1992). Calculation of the Free Energy of Association for Protein Complexes.
33. D. Xu, S. L. Lin, and R. Nussinov, *J. Mol. Biol.,* **265,** 68 (1997). Protein Binding Versus Protein Folding: The Role of Hydrophilic Bridges in Protein Associations.
34. J. A. McCammon and S. C. Harvey, *Dynamics of Proteins and Nucleic Acids,* Cambridge University Press, New York, 1987.
35. M. J. Betts and M. J. Sternberg, *Protein Eng.,* **12,** 271 (1999). An Analysis of Conformational Changes on Protein–Protein Association: Implications for Predictive Docking.
36. D. K. Hendrix, T. E. Klein, and I. D. Kuntz, *Protein Sci.,* **8,** 1010 (1999). Macromolecular Docking of a Three-Body System: The Recognition of Human Growth Hormone by Its Receptor.
37. A. Anderson and Z. Weng, *J. Mol. Graphics Modell.,* **17,** 180 (1999). VRDD: Applying Virtual Reality Visualization to Protein Docking and Design.
38. T. P. Lybrand, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1990, Vol. 1, pp. 295–320. Computer Simulation of Biomolecular Systems Using Molecular Dynamics and Free Energy Perturbation Methods.
39. T. P. Straatsma, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1996, Vol. 9, pp. 81–127. Free Energy by Molecular Simulation.
40. 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 Simulation.
41. 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.

42. B. O. Brandsdal and A. O. Smalas, *Protein Eng.*, **13,** 239 (2000). Evaluation of Protein–Protein Association Energies by Free Energy Perturbation Calculations.
43. L. Wesson and D. Eisenberg, *Protein Sci.,* **1,** 227 (1992). Atomic Solvation Parameters Applied to Molecular Dynamics of Proteins in Solution.
44. J. D. Madura, M. E. Davis, M. K. Gilson, R. C. Wade, B. A. Luty, and J. A. McCammon, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1994, Vol. 5, pp. 229–267. Biological Applications of Electrostatic Calculations and Brownian Dynamics Simulations.
45. J. Cherfils, S. Duquerroy, and J. Janin, *Proteins: Struct., Genet., Funct.,* **11,** 271 (1991). Protein–Protein Recognition Analyzed by Docking Simulation.
46. R. R. Gabdoulline and R. C. Wade, *Biophys. J.,* **72,** 1917 (1997). Simulation of the Diffusional Association of Barnase and Barstar.
47. I. V. Ouporov, H. R. Krull, and K. A. Thomasson, *Biophys. J.,* **76,** 17 (1999). Brownian Dynamics Simulations of Interactions between Aldolase and G- or F-Actin.
48. H. A. Gabb, R. M. Jackson, and M. J. Sternberg, *J. Mol. Biol.,* **272,** 106 (1997). Modelling Protein Docking Using Shape Complementarity, Electrostatics and Biochemical Information.
49. M. Meyer, P. Wilson, and D. Schomburg, *J. Mol. Biol.,* **264,** 199 (1996). Hydrogen Bonding and Molecular Surface Shape Complementarity as a Basis for Protein Docking.
50. I. A. Vakser, *Protein Eng.,* **8,** 371 (1995). Protein Docking for Low-Resolution Structures.
51. D. W. Ritchie and G. J. Kemp, *Proteins: Struct., Genet., Funct.,* **39,** 178 (2000). Protein Docking Using Spherical Polar Fourier Correlations.
52. D. Fischer, S. L. Lin, H. L. Wolfson, and R. Nussinov, *J. Mol. Biol.,* **248,** 459 (1995). A Geometry-Based Suite of Molecular Docking Processes.
53. D. K. Hendrix and I. D. Kuntz, *Pacific. Symp. Biocomput.,* 317 (1998). Surface Solid Angle-based Site Points for Molecular Docking.
54. B. Sandak, R. Nussinov, and H. J. Wolfson, *Comput. Appl. Biosci.,* **11,** 87 (1995). An Automated Computer Vision and Robotics-Based Technique for 3-D Flexible Biomolecular Docking and Matching.
55. R. M. Jackson, H. A. Gabb, and M. J. Sternberg, *J. Mol. Biol.,* **276,** 265 (1998). Rapid Refinement of Protein Interfaces Incorporating Solvation: Application to the Docking Problem.
56. E. Althaus, O. Kohlbacher, H.-P. Lenhof, and P. Müller, in *Recomb 2000—Proceedings of the Fourth Annual International Conference on Computational Molecular Biology,* R. Shamir, S. Miyano, S. Istrail, P. Pevzner, and M. Waterman, Eds., ACM Press, New York, 2000, pp. 15–24.
57. Z. Weng, S. Vajda, and C. DeLisi, *Protein Sci.,* **5,** 614 (1996). Prediction of Protein Complexes Using Empirical Free Energy Functions.
58. R. Abagyan, M. Totrov, and D. Kuznetsov, *J. Comput. Chem.,* **15,** 488 (1994). ICM—A New Method for Protein Modelling and Design: Applications to Docking and Structure Prediction from the Distorted Native Conformation.
59. G. M. Ullmann, E.-W. Knapp, and N. M. Kostic, *J. Am. Chem. Soc.,* **119,** 42 (1997). Computational Simulation and Analysis of Dynamic Association Between Plastocyanin and Cytochrome F. Consequences for the Electron-Transfer Reaction.
60. I. A. Vakser, O. G. Matar, and C. F. Lam, *Proc. Natl. Acad. Sci. U.S.A.,* **96,** 8477 (1999). Systematic Study of Low-Resolution Recognition in Protein–Protein Complexes.
61. J. Janin, *Proteins: Struct., Genet., Funct.,* **28,** 153 (1997). The Kinetics of Protein–Protein Recognition.
62. J. D. Madura, J. M. Briggs, R. C. Wade, and R. R. Gabdoulline, in *Encyclopedia of Computational Chemistry,* P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer, and P. R. Schreiner, Eds., Wiley, Chichester, UK, 1998, Vol. 1, pp. 141–154. Brownian Dynamics.

63. R. R. Gabdoulline and R. C. Wade, *Methods,* **14,** 329 (1998). Brownian Dynamics Simulation of Protein–Protein Diffusional Encounter.
64. R. R. Gabdoulline and R. C. Wade, *J. Phys. Chem.,* **100,** 3868 (1996). Effective Charges for Macromolecules in Solvent.
65. P. N. Palma, L. Krippahl, J. E. Wampler, and J. J. Moura, *Proteins: Struct., Genet., Funct.,* **39,** 372 (2000). BiGGER: A New (Soft) Docking Algorithm for Predicting Protein Interactions.
66. E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A. A. Friesem, C. Aflalo, and I. A. Vakser, *Proc. Natl. Acad. Sci. U.S.A.,* **89,** 2195 (1992). Molecular Surface Recognition: Determination of Geometric Fit Between Proteins and Their Ligands by Correlation Techniques.
67. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing,* Cambridge University Press, Cambridge, UK, 1993.
68. L. F. Ten Eyk, J. Mandell, V. A. Roberts, and M. E. Pique, in *Proceedings of the 1995 ACM/IEEE Supercomputing Conference,* A. Hayes and M. Simmons, Eds., ACM Press, New York, 1995.
69. V. A. Roberts and M. E. Pique, *J. Biol. Chem.,* **274,** 38051 (1999). Definition of the Interaction Domain for Cytochrome C on Cytochrome C Oxidase.
70. M. E. Davis, J. D. Madura, B. A. Luty, and J. A. McCammon, *Comput. Phys. Commun.,* **62,** 187 (1991). Electrostatics and Diffusion of Molecules in Solution: Simulations with the University of Houston Brownian Dynamics Program.
71. A. Nicholls and B. Honig, *J. Comput. Chem.,* **12,** 435 (1991). A Rapid Finite Difference Algorithm, Utilizing Successive Over-Relaxation to Solve the Poisson–Boltzmann Equation.
72. I. A. Vakser and C. Aflalo, *Proteins: Struct., Genet., Funct.,* **20,** 320 (1994). Hydrophobic Docking: A Proposed Enhancement to Molecular Recognition Techniques.
73. A. Bridges, L. Gruenke, Y. T. Chang, I. A. Vakser, G. Loew, and L. Waskell, *J. Biol. Chem.,* **273,** 17036 (1998). Identification of the Binding Site on Cytochrome P450 2b4 for Cytochrome B5 and Cytochrome P450 Reductase.
74. B. K. Shoichet and I. D. Kuntz, *J. Mol. Biol.,* **221,** 327 (1991). Protein Docking and Complementarity.
75. B. Sandak, H. J. Wolfson, and R. Nussinov, *Proteins: Struct., Genet., Funct.,* **32,** 159 (1998). Flexible Docking Allowing Induced Fit in Proteins: Insights from an Open to Closed Conformational Isomers.
76. M. J. E. Sternberg, P. Aloy, H. A. Gabb, R. M. Jackson, G. Moont, E. Querol, and F. X. Aviles, in *Sixth International Conference on Intelligent Systems in Molecular Biology,* J. Glasgow, T. Littlejohn, F. Major, R. Lathrop, D. Sankoff, and C. Sensen, Eds., AAAI Press, Menlo Park, CA, 1998. pp. 183–192. A Computational System for Modelling Flexible Protein–Protein and Protein–DNA Docking.
77. W. Cornell, P. Cieplak, C. I. Bayly, 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 (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids and Organic Molecules.
78. M. Totrov and R. Abagyan, *Nat. Struct. Biol.,* **1,** 259 (1994). Detailed Ab Initio Prediction of Lysozyme-Antibody Complex with 1.6 Å Accuracy.
79. 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.
80. Y. Mauguen, R. W. Hartley, E. J. Dodson, G. G. Dodson, G. Bricogne, C. Chothia, and A. Jack, *Nature, (London),* **297,** 162 (1982). Molecular Structure of a New Family of Ribonucleases.
81. M. J. Lubienski, M. Bycroft, S. M. Freund, and A. R. Fersht, *Biochemistry,* **33,** 8866 (1994). Three-Dimensional Solution Structure and ^{13}C Assignments of Barstar Using Nuclear Magnetic Resonance Spectroscopy.

82. A. M. Buckle, G. Schreiber, and A. R. Fersht, *Biochemistry,* **33,** 8878 (1994). Protein–Protein Recognition: Crystal Structural Analysis of a Barnase–Barstar Complex at 2.0-Å Resolution.

83. G. Schreiber and A. R. Fersht, *Nat. Struct. Biol.,* **3,** 427 (1996). Rapid, Electrostatically Assisted Association of Proteins.

84. H. Goldstein, *Classical Mechanics,* Addison-Wesley, Boston, 1977.

85. A. Amadei, A. B. Linssen, and H. J. Berendsen, *Proteins: Struct., Genet., Funct.,* **17,** 412 (1993). Essential Dynamics of Proteins.

86. B. L. de Groot, D. M. F. van Aalten, R. M. Scheek, A. Amadei, G. Vriend, and H. J. C. Berendsen, *Proteins: Struct., Genet., Funct.,* **29,** 240 (1997). Prediction of Protein Conformational Freedom from Distance Constraints.

87. T. Haliloglu, I. Bahar, and B. Erman, *Phys. Rev. Lett.,* **79,** 3090 (1997). Gaussian Dynamics of Folded Proteins.

88. K. Hinsen, *Proteins: Struct., Genet., Funct.,* **33,** 417 (1998). Analysis of Domain Motions by Approximate Normal Mode Calculations.

89. A. A. Bogan and K. S. Thorn, *J. Mol. Biol.,* **280,** 1 (1998). Anatomy of Hot Spots in Protein Interfaces.

90. L. M. Rice and A. T. Brünger, *Proteins: Struct., Genet., Funct.,* **19,** 277 (1994). Torsion Angle Dynamics: Reduced Variable Conformational Sampling Enhances Crystallographic Structure Refinement.

91. C. R. Reeves, Ed., *Modern Heuristic Techniques For Combinatorial Problems,* Blackwell Scientific Publications, Oxford, UK, 1993.

92. R. Judson in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1997, Vol. 10, pp. 1–73. Genetic Algorithms and Their Use in Chemistry.

93. 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.

94. R. M. Jackson and R. B. Russell, *J. Mol. Biol.,* **296,** 325 (2000). The Serine Protease Inhibitor Canonical Loop Conformation: Examples Found in Extracellular Hydrolases, Toxins, Cytokines and Viral Proteins.

95. B. Roux and T. Simonson, *Biophys. Chem.,* **78,** (1999). Implicit Solvent Models for Biomolecular Simulations.

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

97. M. Rarey, B. Kramer, and T. Lengauer, *Proteins: Struct., Genet., Funct.,* **34,** 17 (1999). The Particle Concept: Placing Discrete Water Molecules During Protein–Ligand Docking Predictions.

98. N. C. Strynadka, M. Eisenstein, E. Katchalski-Katzir, B. K. Shoichet, I. D. Kuntz, R. Abagyan, M. Totrov, J. Janin, J. Cherfils, F. Zimmerman, A. Olson, B. Duncan, M. Rao, R. Jackson, M. Sternberg, and M. N. James, *Nat. Struct. Biol.,* **3,** 233 (1996). Molecular Docking Programs Successfully Predict the Binding of a Beta-Lactamase Inhibitory Protein to Beta-Lactamase.

99. J. S. Dixon, *Proteins: Struct., Genet., Funct.,* **Suppl. 1,** 198 (1997). Evaluation of the CASP2 Docking Section.

100. C. J. Camacho, S. R. Kimura, C. DeLisi, and S. Vajda, *Biophys. J.,* **78,** 1094 (2000). Kinetics of Desolvation-Mediated Protein–Protein Binding.

101. G. Vriend, *J. Mol. Graphics,* **8,** 52 (1990). WHAT IF: A Molecular Modeling and Drug Design Program.

102. R. W. Hooft, C. Sander, and G. Vriend, *Proteins: Struct., Genet., Funct.,* **26,** 363 (1996). Positioning Hydrogen Atoms by Optimizing Hydrogen-Bond Networks in Protein Structures.

103. A. D. MacKerell Jr., D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, *J. Phys. Chem. B,* **102,** 3586 (1998). All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies.

104. A. T. Brünger, *X-PLOR. A System for X-Ray Crystallography and NMR,* Yale University Press, New Haven, CT, 1992.

105. K. Hinsen, *J. Comput. Chem.,* **21,** 79 (2000). The Molecular Modelling Toolkit: A New Approach to Molecular Simulations.

106. F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyers 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. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, *Nucleic Acids Res.*, **28,** 235 (2000). The Protein Data Bank. www.rcsb.org/pdb/.

Reviews in Computational Chemistry, Volume 17. Edited by Kenny B. Lipkowitz, Donald B. Boyd

ISBNs: 0-471-39845-4 (Hardcover); 0-471-22441-3 (Electronic)

CHAPTER 3

Spin–Orbit Coupling in Molecules

Christel M. Marian

German National Research Center for Information Technology (GMD), Scientific Computing and Algorithms Institute (SCAI), Schloss Birlinghoven, D-53754 St. Augustin, Germany

WHAT IT IS ALL ABOUT

Spin enters quantum chemistry for two main reasons. First, spin degrees of freedom imply a particular symmetry behavior with respect to the exchange of two identical particles; for electrons, this symmetry constraint is commonly known as the Pauli principle. Second, and more relevant in the present context, there is a magnetic moment associated with spin that can interact with all other kinds of magnetic moments.

In the absence of spin-magnetic interactions, electronic wave functions can be made eigenfunctions of the total electronic spin $\hat{S}^2$. Such an $\hat{S}^2$ eigenfunction with eigenvalue $s(s+1)$ is $(2s+1)$-fold degenerate with respect to the energy and is called a *spin multiplet*. A close look at high-resolution molecular electronic spectra reveals, however, that electronic spin multiplets are not properly degenerate. Actually, the sublevels are separated energetically by what is called the *fine-structure splitting*.

A *zero-field splitting* originates from internal magnetic interactions of the electrons. By far the most important among the mechanisms lifting the degeneracy of electronic spin multiplets—and the focus of this tutorial—is spin–orbit coupling. Spin–orbital coupling (SOC) is a relativistic effect; consequently, its impact on molecular properties increases with nuclear charge Z to an extent that molecules containing heavy elements cannot be described

correctly even in a qualitative manner, if spin–orbit coupling is not taken into account. The second relevant type of interaction contributing to the zero-field splitting is electronic spin–spin interaction. Unlike spin–orbit coupling, electronic spin–spin coupling does not scale with nuclear charge. Therefore, spin–orbit coupling effects tend to outweigh spin–spin interaction energies by at least one order of magnitude. On the other hand, in spatially nondegenerate electronic states of light molecules, spin–orbit coupling contributes to the multiplet splitting only in second or higher order. A well-known example is the $^3\Sigma^-$ ground state of O_2; here electronic spin–spin interaction is approximately of the same size as second-order spin–orbit interaction.

Also external magnetic fields may cause multiplet splittings. These are usually much smaller than zero-field splittings, but they can be tuned by the strength of the external field. Prominent historical milestones—which eventually led to the detection of spin—are the Stern–Gerlach experiment and Zeeman spectroscopy as will be discussed in the next section. Other well-known experiments that exploit the effects of external magnetic fields on molecular energies are electron spin resonance (ESR) and nuclear magnetic resonance (NMR) spectroscopy. In ESR experiments, transitions within an electronic spin multiplet are induced; in NMR the same applies to nuclear spin multiplets.

As we shall see later, the degeneracies of an electronic multiplet are not lifted completely by spin–orbit or spin–spin coupling. For example, each electronic state of a nonrotating diatomic molecule with a nonzero angular momentum remains doubly degenerate in a field-free surrounding. These degeneracies result from the invariance properties of the Hamiltonian with respect to time inversion. They persist for all types of magnetic interaction Hamiltonians that involve two angular momenta defined with respect to the same origin as, for example, the molecule-fixed coordinate system. Splitting of the parity sublevels requires the interaction of an internal angular momentum such as the angular momentum $\vec{L}$ or electron spin $\vec{S}$ with an external magnetic field $\vec{B}$ or with the angular momentum $\vec{R}$ brought about by the rotation of the nuclear frame. Their energetic separation is typically of the same size as hyperfine splittings. The latter are brought about by all interactions between electrons and nuclei apart from their strong mutual Coulomb attraction. We shall not address either of these effects here. Suffice it to say that they are several orders-of-magnitude smaller than fine-structure effects caused by electronic spin–orbit or spin–spin interactions.

In addition to causing fine-structure splitting, magnetic interactions may couple states of different spin multiplicities. As a consequence, so-called *spin-forbidden transitions* yield some intensity. Well-known examples for this phenomenon are phosphorescence and nonradiative transitions at intersystem crossings.

Because of the importance of spin-dependent effects in molecules, it might be interesting to learn something more about spin and its interactions

rather than just a few rules of thumb. On the one hand, this is not an easy task because spin is a quantum effect and has no classical analogue. Hamiltonians describing its interactions with internal and external magnetic fields can be derived from relativistic quantum theory which will, however, not be the center of interest here. Rather, we will use these results and focus on the quantum theory of angular momenta and the group theoretical machinery to describe their transformation properties and coupling. On the other hand, the classification of spin-dependent interactions according to their transformation properties will make things easy. Once we have understood the underlying concept, we can apply the machinery to all kinds of magnetic interactions.

THE FOURTH ELECTRONIC DEGREE OF FREEDOM

The Stern–Gerlach Experiment

In 1921, Stern and Gerlach performed an experiment that later turned out to be a milestone in quantum mechanics.[1,2] First, it provided an experimental basis for the concept of electron spin, introduced in 1925 by Goudsmit and Uhlenbeck.[3,4] Second, it evolved into *the* quantum mechanical experiment par excellence. From this experiment, we easily learn basic concepts of quantum mechanics such as the additivity of probability amplitudes, basis states, projection operators, and the resolution of the identity.[5] The latter concept relates to the fact that a complete set of basis states (i.e., the identity) can be inserted in any quantum mechanical equation without changing the result.

Stern and Gerlach set out to measure the magnetic moment of silver atoms by deflecting a beam of silver atoms in an inhomogeneous magnetic field as sketched in Figure 1. The idea behind this experiment was the following. If silver atoms possess a magnetic moment $\vec{\mu}$, their potential energy in a magnetic field $\vec{B}$ oriented along the z axis is given by

$$E = -\vec{\mu} \cdot \vec{B} = -|\mu||B|\cos\theta = -\mu_z B_z \qquad [1]$$

and the corresponding force leading to a deflection in the z direction by

$$F = -\frac{\partial E}{\partial z} = \frac{\partial(|\mu||B|\cos\theta)}{\partial z} = \mu_z \frac{\partial B_z}{\partial z} \qquad [2]$$

Here, θ is the acute angle between the orientation of the particle magnetic moment and the magnetic field vector. Before entering the magnet, the silver atoms are oriented randomly with respect to the magnetic field (i.e., $\cos(\theta)$ can adopt any value between -1 and 1). Classically, the interaction of

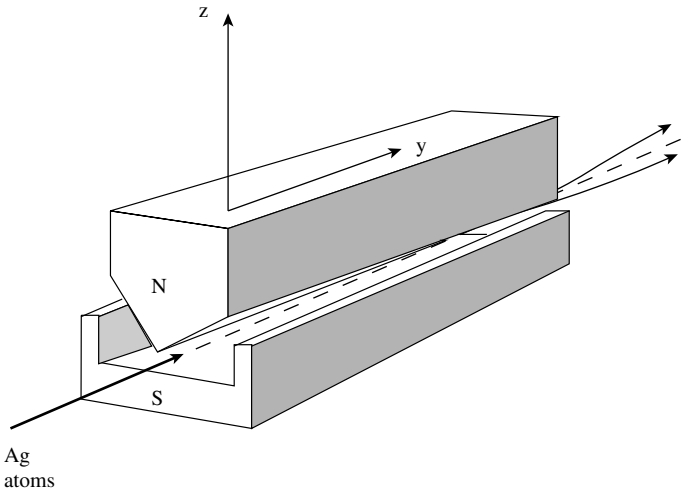

Figure 1 Schematic drawing of the Stern–Gerlach experiment.

randomly oriented magnetic dipoles with an inhomogenous magnetic field in the z direction is expected to produce a continuous distribution of deflected atoms, smeared out along a line in the z direction where the upper bound M_{up} corresponds to $\mu_z = |\mu|$ and the lower bound M_{down} to $\mu_z = -|\mu|$.

The results of the Stern–Gerlach experiment were in complete contradiction to the classical interpretation and its predictions. Silver atoms turned out to possess a magnetic moment, but instead of a single, smeared-out distribution, two spots centered around M_{up} and M_{down} were observed. Thus the magnetic moment of a silver atom is space-quantized by an inhomogeneous magnetic field, and this magnetic moment can adopt only two values, $\mu_z = \pm|\mu|$.

The origin of this magnetic moment was not clear in 1922. In its electronic ground state, a silver atom does not possess a spatial angular momentum, and the concept of an intrinsic electronic angular momentum (the electron spin) was yet to be created. In 1925, Goudsmit and Uhlenbeck introduced a fourth (spin) electron degree of freedom—in addition to the three spatial coordinates (x, y, z)—as a model to ease the explanation of the anomalous Zeeman effect.[3,4]

From our present standpoint, we know that the deflection of a silver atom in the Stern–Gerlach experiment is caused by the interaction of its electronic spin angular momentum $\vec{S}$ with an inhomogeneous magnetic field. The projection of $\vec{S}$ on the direction of this field, M_S, is quantized. For a silver atom, M_S can take two values: $+\frac{1}{2}\hbar$ and $-\frac{1}{2}\hbar$, where $\hbar$ is Planck's constant h over 2π and adopts a value of $1.054571596 \times 10^{-34}$ Js in cgi units.

Zeeman Spectroscopy

The Zeeman effect is the modification of an atomic or molecular spectrum by the application of a uniform magnetic field. Historically, scientists differentiated between the normal and the anomalous Zeeman effect. As a spectroscopic tool, only the latter is of importance nowadays. The *normal Zeeman* effect on an atomic spectrum yields three lines where there is one in the absence of the magnetic field. The interval between these lines is proportional to the applied magnetic field. This splitting pattern results from the interaction between the external magnetic field and the orbital angular momentum of the atom under investigation (see below) and was well understood in the early 1920s. While this theory is successful in many cases, it completely fails in accounting quantitatively for the phenomena in other cases, the so-called *anomalous Zeeman* effect. In particular, the Zeeman effect on the spectra of atoms with an odd number of electrons (e.g., the hydrogen and alkali atoms) puzzled physicists in the beginning of the twentieth century. Historically, the doublet structure of the alkali Zeeman spectra led to the postulation of half-integer angular momentum quantum numbers and eventually to the concept of an electron spin.

The Normal Zeeman Effect

Let us first consider the normal Zeeman effect, which applies to transitions between electronic states with zero total spin magnetic moment, so-called singlet states. Like the projection M_S of $\vec{S}$ in the Stern–Gerlach experiment, the projection M_L of the spatial angular momentum $\vec{L}$ is space quantized in the external magnetic field. We shall describe the quantization of the spatial angular momentum by means of quantum mechanical methods in detail later. Suffice it to say that each state with spatial angular momentum quantum number L splits into $2L + 1$ components, i.e., a P state ($L = 1$) splits into three components with

$$M_L = +1 \qquad M_L = 0 \qquad M_L = -1 \tag{3}$$

and Zeeman potential energies of

$$E_{+1} = \mu_B B_z \qquad E_0 = 0 \qquad E_{-1} = -\mu_B B_z \tag{4}$$

An electronic singlet S state ($L = 0$) does not interact at all with a magnetic field. In Figure 2, the Zeeman effect on an electronic transition between an atomic S state and a P state with zero spin is sketched. Radiative electric dipole transitions can occur between all three Zeeman sublevels of the P state and the S state, thus giving rise to three (closely spaced) spectral lines.

Let us compare the spectral pattern of a Zeeman-split ^1P–^1S transition with the Zeeman effect on an electronic transition between an atomic singlet

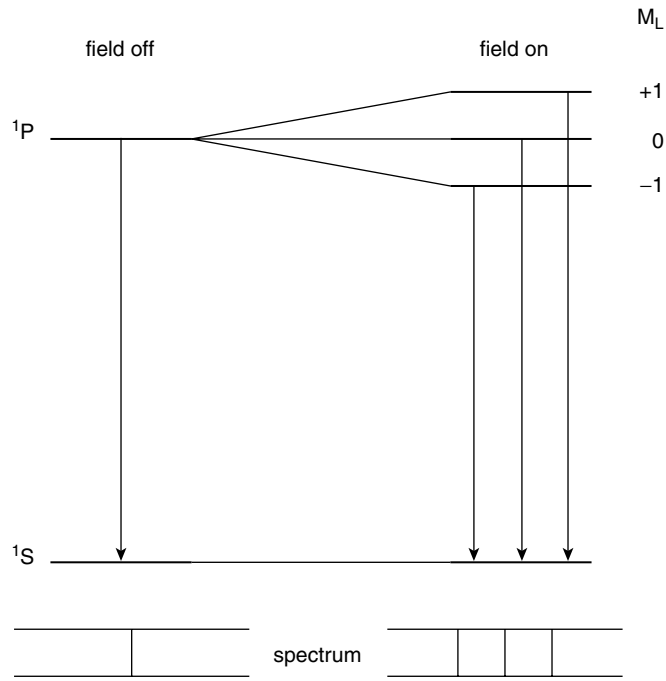

Figure 2 Normal Zeeman effect on a ^1P–^1S transition. The Zeeman splitting is not drawn to scale.

P and a D state, sketched in Figure 3. In an external magnetic field, an atomic D state splits into five components with

$$M_L = +2 \qquad M_L = +1 \qquad M_L = 0 \qquad M_L = -1 \qquad M_L = -2 \qquad [5]$$

and Zeeman potential energies of

$$E_{+2} = 2\mu_B B_z \qquad E_{+1} = \mu_B B_z \qquad E_0 = 0$$
$$E_{-1} = -\mu_B B_z \qquad E_{-2} = -2\mu_B B_z \qquad [6]$$

Since electric dipole transitions alter the M_L quantum number by at most one unit

$$\Delta M_L = 0, \pm 1 \qquad [7]$$

and the energy separations between two neighboring levels in P and D states are equal, the same pattern of three lines results as in the ^1P–^1S spectrum.

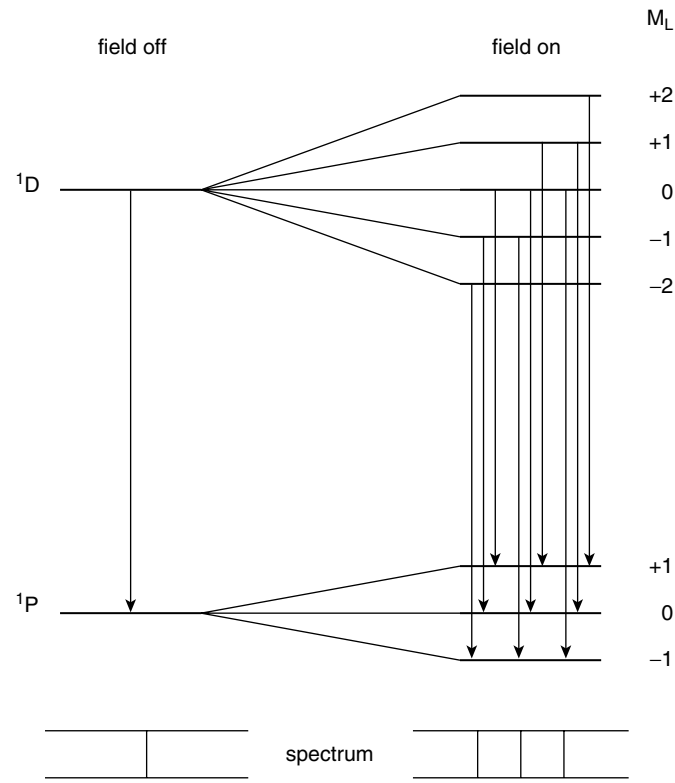

Figure 3 Normal Zeeman effect on a ^1D–^1P transition. The Zeeman splitting is not drawn to scale.

The Anomalous Zeeman Effect

Much more interesting and informative than Zeeman spectroscopy on atoms with zero electronic spin is the Zeeman effect on electric dipole transitions between states with a nonzero electronic spin moment. For historical reasons, this is called the anomalous Zeeman effect.

An example for the Zeeman effect on a spectrum of an alkali metal is sketched in Figure 4. Here is a brief modern interpretation of this term scheme. On the very left of Figure 4, a hypothetical level scheme is drawn that would result if there were no magnetic interactions at all for a single electron in an s orbital ($\ell = 0$, lower state) and in a p orbital ($\ell = 1$, upper state). In this case, the upper electronic state is sixfold degenerate, the lower one is twofold. However, even in the absence of an external magnetic field, two transitions can be observed—in sodium these are the famous yellow D lines. In the alkali metal atoms, this zero-field splitting is caused by spin–orbit coupling. The state labeled $^2P_{3/2}$ is electronically fourfold degenerate; $^2P_{1/2}$ and $^2S_{1/2}$ are twofold each. If we switch on an external magnetic field, the degeneracies are lifted

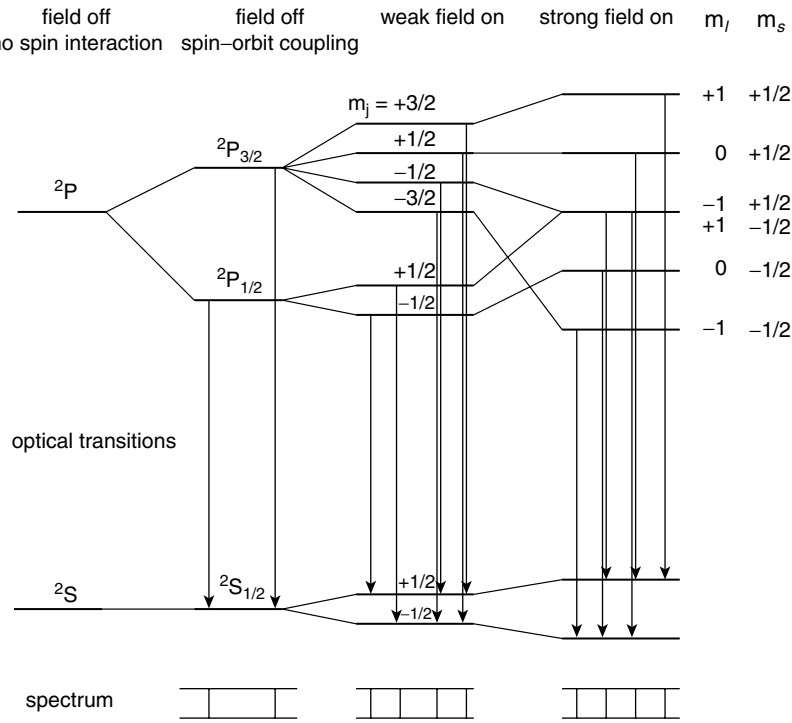

Figure 4 Anomalous Zeeman effect on a ^2P–^2S transition (weak field). The strong-field limit is called the Paschen–Back effect. The splitting of the levels is not drawn to scale.

completely. In a weak or medium field, the internal magnetic interactions between the electronic spin and orbital angular momenta are still larger than their interactions with the external magnetic field. Because of the spin–orbit coupling, it is not possible to assign separate quantum numbers for the projections m_ℓ and m_s of the orbital and spin angular momentum, respectively; only their sum $m_\ell + m_s = m_j$ is conserved. If the external field is increased to a strength such that its interaction with the electronic spin and orbital angular momenta exceeds the internal coupling between ℓ and s, m_ℓ and m_s become space-quantized separately with respect to the external field (Paschen–Back effect). Figure 4 shows that the splitting pattern of the upper electronic state has changed (i.e., five equidistant levels are obtained). In particular, it is noteworthy that the middle one (shown on the far right in Figure 4) is doubly degenerate and has exactly the same energy as the hypothetical ^2P state without magnetic interaction (shown to the very left in Figure 4). Because of the same energies, it follows that the total magnetic moment of these levels on the far right must be zero. Since neither the orbital angular momentum nor the spin magnetic momentum are zero, their separate interactions with the external field must compensate. Let us for a moment concentrate on just one of these levels: its m_ℓ and m_s values amount to -1 and $+\frac{1}{2}$, respectively.

For their interactions with the external field to be equal, the proportionality constant between the spin angular momentum and the associated magnetic moment must be twice as large as for the angular momentum. More precisely, this ratio amounts to 2.0023; the deviation from the value of 2, predicted by Dirac theory, is mainly caused by radiative corrections explainable in the framework of quantum field theory.

Now, let us forget about spin and go back in history again. In 1921, Landé had been successful in deducing the energy separations of the atomic term levels from the complicated splitting of spectral lines.[6–8] The most fundamental of his results was that he introduced half-integer magnetic quantum numbers for the description of the doublet structure of the alkali spectra. [An integer value of the angular momentum quantum number j always gives rise to an odd number of Zeeman sublevels $(2j + 1)$; since the electronic states of the alkali atoms split into an even number of sublevels, a half-integer value has to be postulated for j.] Even though it had become possible then to describe the observed phenomena by means of mathematical formulas, they were far from understood. At the beginning of the twentieth century, it was generally believed that the doublet splitting had its origin in a nonzero angular momentum of the nucleus. Indeed, the natural isotope of the Na nucleus has a nonzero angular momentum (i.e., a nuclear spin $I = \frac{3}{2}$), which gives rise to a nuclear Zeeman effect. This splitting is much smaller, by a factor of about 2000, than the effect discussed here. It was Pauli who postulated in 1925 that the anomalous Zeeman effect in the spectra of alkali atoms was solely due to the valence electron and a strange, classically not describable ambiguity of its quantum theoretical properties:[9]

> "Die abgeschlossenen Elektronenkonfigurationen sollen nichts zum magnetischen Moment und zum Impulsmoment des Atoms beitragen. Insbesondere werden bei den Alkalien die Impulswerte des Atoms und seine Energieänderungen in einem äußeren Magnetfeld im wesentlichen als eine alleinige Wirkung des Leuchtelektrons angesehen, das auch als der Sitz der magneto-mechanischen Anomalie betrachtet wird. Die Dublettstruktur der Alkalispektren, sowie die Durchbrechung des Larmortheorems kommt gemäß diesem Standpunkt durch eine eigentümliche, klassisch nicht beschreibbare Art von Zweideutigkeit der quantentheoretischen Eigenschaften des Leuchtelektrons zustande."[a]

[a]English translation: The closed electron configurations shall not contribute to the magnetic moment and to the linear momentum of the atom. In particular, the (angular) momentum values of the alkali atoms and the changes of the atomic energies in an external magnetic field are regarded to be essentially solely due to the "Leuchtelekton" (valence electron), which is considered to be also the origin of the magnetomechanic anomaly. According to this view, the doublet structure of the alkali spectra as well as the violation of the Larmor theorem is caused by a strange, classically not describable kind of ambiguity in the quantum theoretical properties of the valence electron.

Carrying on the investigation of the anomalous Zeeman effect and the Paschen–Back effect on the spectra of the alkali atoms, Pauli postulated that an electron in an external magnetic field has to be described by four independent quantum numbers. Moreover, in order to justify the Bohr–Sommerfeld Aufbau (building-up) principle of the periodic system of the elements, he came up with his famous *exclusion principle* (Ausschließungsprinzip). In its original formulation it reads:[10]

> "Es kann niemals zwei oder mehrere äquivalente Elektronen im Atom geben, für welche in starken Feldern die Werte aller Quantenzahlen n, k_1, k_2, m_1 (oder, was dasselbe ist, n, k_1, m_1, m_2) übereinstimmen. Ist ein Elektron in Atom vorhanden, für das diese Quantenzahlen (im äußeren Felde) bestimmte Werte haben, so ist dieser Zustand 'besetzt' ".[b]

The four quantum numbers n, k_1, k_2, m_1 are related to the quantum numbers in our modern terminology, that is, n (principal quantum number), ℓ (orbital angular momentum), j (total angular momentum of an electron), and m_j (projection of j on the quantization axis), by

$$n = n \qquad k_1 = \ell + 1 \qquad k_2 = j + \frac{1}{2} \qquad m_1 = m_j \qquad [8]$$

The postulate of a fourth independent quantum number was difficult to understand, since electrons were assumed to have only three—the spatial—degrees of freedom.

Shortly after the formulation of the exclusion principle, Goudsmit and Uhlenbeck elegantly interpreted Pauli's fourth independent electronic quantum number by postulating a fourth (spin) electron degree of freedom, an intrinsic angular momentum that is quantized and can adopt two orientations in an external magnetic field.[3,4] Assuming that the spin quantum number s of an electron is $\frac{1}{2}$ and that the ratio of the spin magnetic moment and the spin angular momentum is twice as large as for the orbital angular momentum, the anomalous Zeeman effect of the alkali atoms is easily explained.

Spin Is a Quantum Effect

Although the electron spin is often referred to as an intrinsic angular momentum, it should be emphasized that spin has no classical analog. In particular, one should not imagine a spinning electron as a particle rotating about an inner axis (e.g., like the earth does). *Spin has to be regarded as a pure*

[b]English translation: There are never two or more electrons in an atom for which, in strong fields, the values of all quantum numbers n, k_1, k_2, m_1 (or equivalently n, k_1, m_1, m_2) agree. If in an atom an electron exists for which these quantum numbers (in an external field) have definite values, then this state is "occupied."

quantum effect. Sometimes we may try to simplify the description of its interactions by using the classical picture of a bar magnet, but we have to keep in mind that the rigorous treatment requires (relativistic) quantum theory. Contrary to what we find in many textbooks, electron spin does not automatically follow from relativistic invariance. Particles with different spin (0, $\frac{1}{2}$, 1, $\frac{3}{2}$, etc.) are described by different relativistic equations. A proper (and the most common) relativistic equation for a spin $\frac{1}{2}$ particles is the Dirac equation.[11] The most compelling definition, at least known to the author, for the spin of a particle with rest mass $m \neq 0$ is the following:[12]

Spin is the angular momentum of a particle at rest.

ANGULAR MOMENTA

Orbital Angular Momentum

Relation to an Infinitesimal Rotation

Imagine a particle with position vector $\vec{r}_a$ in a three-dimensional (3D) orthonormal coordinate system with axes $\vec{x}$, $\vec{y}$, $\vec{z}$ and a wave function $\Psi(\vec{r}_a) = \Psi(x_a, y_a, z_a)$. If the particle is rotated about one of the axes $\vec{x}$, $\vec{y}$, or $\vec{z}$ through an infinitesimal angle $\delta\varphi$, the new position coordinates $\vec{r}'_a$ are given in first order by

$$\begin{array}{lccc} x \text{ axis:} & x'_a = x_a & y'_a = y_a - \delta\varphi_x z_a & z'_a = z_a + \delta\varphi_x y_a \\ y \text{ axis:} & x'_a = x_a + \delta\varphi_y z_a & y'_a = y_a & z'_a = z_a - \delta\varphi_y x_a \\ z \text{ axis:} & x'_a = x_a - \delta\varphi_z y_a & y'_a = y_a + \delta\varphi_z x_a & z'_a = z_a \end{array} \quad [9]$$

To arrive at these equations, we have made use of the addition rules for the trigonometric functions and have set $\cos(\delta\varphi) \approx 1$ and $\sin(\delta\varphi) \approx \delta\varphi$. An example for a rotation about the z axis is shown in Figure 5. This result can be generalized, if we interpret $\delta\varphi_x$, $\delta\varphi_y$, and $\delta\varphi_z$ as the components of an infinitesimal rotation vector $\vec{\delta\varphi}$. As indicated in Figure 6, $\vec{\delta\varphi}$ points in the direction of the rotation axis; its length is equal to the angle $\delta\varphi$. Now, making use of the definition of the cross (vector) product, we can write the x component of the new position vector as

$$x'_a = x_a + \delta\varphi_y z_a - \delta\varphi_z y_a = x_a + \left(\vec{\delta\varphi} \times \vec{r}_a\right)_x \quad [10]$$

Similar relations are obtained for y'_a and z'_a. Thus, if the particle with position vector $\vec{r}_a$ is rotated about a general axis through $\delta\varphi$, its new position coordinates are given by

$$\vec{r}'_a = \vec{r}_a + \delta\vec{r}_a = \vec{r}_a + \vec{\delta\varphi} \times \vec{r}_a \quad [11]$$

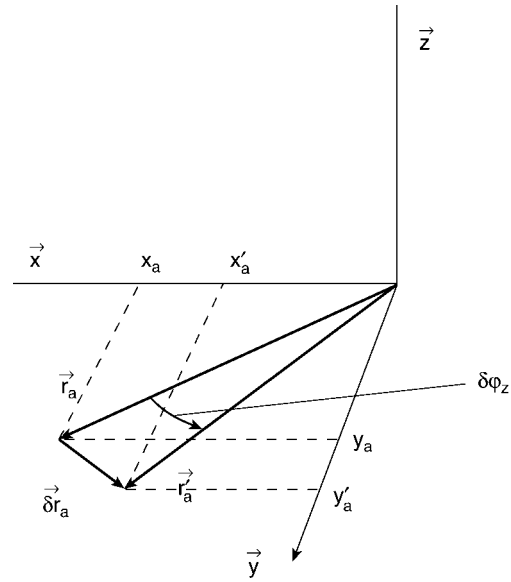

Figure 5 Rotation about $\vec{z}$ through $\delta\varphi_z$.

To first order, any function $\Psi(\vec{r}_a)$ is transformed under this coordinate rotation into

$$\Psi(\vec{r}_a + \vec{\delta r}_a) = \Psi(\vec{r}_a) + \vec{\delta r}_a \cdot \vec{\nabla}_a \Psi(\vec{r}_a) \quad [12]$$

$$= \Psi(\vec{r}_a) + (\vec{\delta\varphi} \times \vec{r}_a) \cdot \vec{\nabla}_a \Psi(\vec{r}_a) \quad [13]$$

$$= \left(1 + \vec{\delta\varphi} \cdot (\vec{r}_a \times \vec{\nabla}_a)\right) \Psi(\vec{r}_a) \quad [14]$$

The expression

$$1 + \vec{\delta\varphi} \cdot (\vec{r}_a \times \vec{\nabla}_a) \quad [15]$$

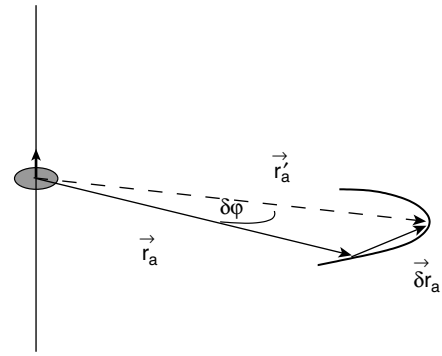

Figure 6 An infinitesimal rotation.

is the infinitesimal rotation operator. Since an infinitesimal rotation conserves the energy of a system, it must commute with the Hamiltonian. Because $\vec{\delta\varphi}$ is a constant vector, this condition can be formulated as

$$(\vec{r}_a \times \vec{\nabla}_a)\hat{\mathscr{H}} - \hat{\mathscr{H}}(\vec{r}_a \times \vec{\nabla}_a) = 0 \qquad [16]$$

On the other hand, the entity conserved in a closed system due to the isotropy of space is the *orbital angular momentum* of the system. Apart from a constant factor, the operator $\vec{r}_a \times \vec{\nabla}_a$ must therefore correspond to the orbital angular momentum. Further, the angular momentum is an observable (i.e., real valued). Thus the corresponding operator ought to be Hermitian. An operator $\hat{\mathscr{O}}$ is said to be Hermitian if it obeys the turn-over rule, that is,

$$\int \Psi^*(\hat{\mathscr{O}}\Phi)\,d\tau = \left(\int (\hat{\mathscr{O}}\Psi^*)\Phi\,d\tau\right)^* = \int \Phi^*(\hat{\mathscr{O}}^*\Psi)\,d\tau \qquad [17]$$

where the asterisk indicates complex conjugation. (Hermitian operators have only real eigenvalues.) Like $\vec{\nabla}$ itself, $\vec{r} \times \vec{\nabla}$ is an antisymmetric operator, that is, the exchange of the two wavefunctions Ψ and Φ in $\int \Psi^*(\vec{r}_a \times \vec{\nabla}_a)\Phi\,d\tau$ changes the sign of the integral. The requirement of antisymmetry and Hermiticity can be fulfilled, if the operator is purely imaginary. In vector notation, the angular momentum operator of a particle is given by

$$\hat{\vec{\ell}} = \vec{r} \times \hat{\vec{p}} = -i\hbar(\vec{r} \times \vec{\nabla}) \qquad [18]$$

or in form of its Cartesian components by

$$\hat{\ell}_x = \hat{y}\hat{p}_z - \hat{z}\hat{p}_y \qquad \hat{\ell}_y = \hat{z}\hat{p}_x - \hat{x}\hat{p}_z \qquad \hat{\ell}_z = \hat{x}\hat{p}_y - \hat{y}\hat{p}_x \qquad [19]$$

The corresponding many-particle operators are obtained by summing over the indices of the individual particles

$$\hat{\vec{\mathscr{L}}} = \sum_a \hat{\vec{\ell}}_a \qquad [20]$$

In the following, we shall denote the angular momentum operator of a single particle by a lower case letter and use a capital letter for the angular momentum operator of the total system.

In addition to the linear angular momenta, we introduce their square moduli

$$\hat{\ell}^2 = \hat{\vec{\ell}} \cdot \hat{\vec{\ell}} = \hat{\ell}_x^2 + \hat{\ell}_y^2 + \hat{\ell}_z^2 \qquad [21]$$

and

$$\hat{\mathscr{L}}^2 = \vec{\hat{\mathscr{L}}} \cdot \vec{\hat{\mathscr{L}}} = \hat{\mathscr{L}}_x^2 + \hat{\mathscr{L}}_y^2 + \hat{\mathscr{L}}_z^2$$
$$= \sum_a \hat{\ell}_{ax} \sum_b \hat{\ell}_{bx} + \sum_a \hat{\ell}_{ay} \sum_b \hat{\ell}_{by} + \sum_a \hat{\ell}_{az} \sum_b \hat{\ell}_{bz} \qquad [22]$$

For later convenience, it is useful to rewrite the infinitesimal rotation operator in terms of the angular momentum operator. If the rotation axis happens to coincide with one of the Cartesian axes, say the z axis, we can write

$$1 + \sum_a \delta\vec{\varphi}\,(\vec{r}_a \times \vec{\nabla}_a)_z = 1 + \frac{i}{\hbar}\delta\varphi\,\hat{\mathscr{L}}_z \qquad [23]$$

Analogously, for an infinitesimal rotation through $\delta\varphi$ about a general axis $\vec{n}$, one obtains

$$1 + \frac{i}{\hbar}\delta\varphi\,\vec{n} \cdot \vec{\hat{\mathscr{L}}} \qquad [24]$$

where $\vec{n} \cdot \vec{\hat{\mathscr{L}}}$ denotes the scalar product between $\vec{n}$ and $\vec{\hat{\mathscr{L}}}$ (i.e., the projection of $\vec{\hat{\mathscr{L}}}$ on the rotation axis $\vec{n}$).

Spherical Harmonics

In principle, knowledge of Eqs. [18]–[22] is sufficient to set up differential equations for the orbital angular momentum operators and to solve for eigenvalues and eigenfunctions. The solutions are most easily obtained employing spherical coordinates r, θ, ϕ (see Figure 7). The solutions, called spherical harmonics, can be found in any introductory textbook of quantum chemistry and shall be given here only for the sake of clarity.

$$Y_{\ell m}(\theta, \phi) = (-1)^{(m+|m|)/2} e^{i\ell\pi/2} \left[\frac{2\ell + 1}{4\pi} \frac{(l - |m|)!}{(l + |m|)!}\right]^{1/2} P_\ell^{|m|}(\cos\theta) e^{im\phi} \qquad [25]$$

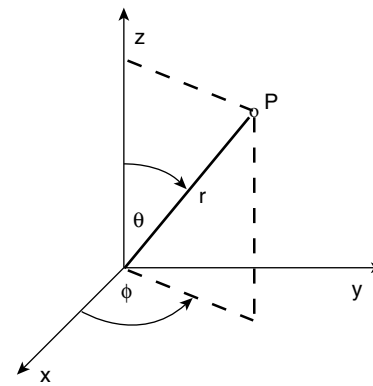

Figure 7 Polar coordinates r, θ, ϕ of a point P.

The spherical harmonics $Y_{\ell m}$ are simultaneously eigenfunctions of $\hat{\ell}^2$ with eigenvalue $\ell(\ell+1)$ and of $\hat{\ell}_z$ with eigenvalue m. They do not depend on the radial coordinate r, but are solely functions of the azimuthal angle θ and the polar angle ϕ. Although their functional form looks rather complicated, they are well known to chemists. In combination with specific radial functions, they constitute the solutions of the Schrödinger equation for the hydrogen atom, that is, Y_{00} is the angular part of an s orbital; p orbitals involve Y_{10} or $Y_{1\pm1}$; the θ and ϕ dependence of d orbitals is contained in Y_{20}, $Y_{2\pm1}$, or $Y_{2\pm2}$; the analytical expressions for f orbitals contain Y_{30}, $\cdots$, or $Y_{3\pm3}$, and so on.

Commutation Relations

The eigenvalues and eigenfunctions of the orbital angular momentum operators can also be derived solely on the basis their commutation relations. This derivability is particularly attractive because the spin operators and the total angular momentum obey the same commutation relations.

The commutation relations of the orbital angular momentum operators can be derived from those between the components of $\vec{\hat{r}}$ and $\vec{\hat{p}}$. If we denote the Cartesian components by the subindices i, k, and l, we can use the short-hand notation

$$[\hat{p}_i, \hat{r}_k] = -i\hbar\delta_{ik} \qquad [26]$$

where δ_{ik} is the Kronecker delta. This means that the diagonal elements are

$$[\hat{p}_x, \hat{x}] = -i\hbar \qquad [\hat{p}_y, \hat{y}] = -i\hbar \qquad [\hat{p}_z, \hat{z}] = -i\hbar \qquad [27]$$

and all off-diagonal elements are zero. By using these relations, we find

$$\begin{array}{lll} [\hat{\ell}_x, \hat{x}] = 0 & [\hat{\ell}_x, \hat{y}] = i\hbar\hat{z} & [\hat{\ell}_x, \hat{z}] = -i\hbar\hat{y} \\ [\hat{\ell}_y, \hat{y}] = 0 & [\hat{\ell}_y, \hat{z}] = i\hbar\hat{x} & [\hat{\ell}_y, \hat{x}] = -i\hbar\hat{z} \\ [\hat{\ell}_z, \hat{z}] = 0 & [\hat{\ell}_z, \hat{x}] = i\hbar\hat{y} & [\hat{\ell}_z, \hat{y}] = -i\hbar\hat{x} \end{array} \qquad [28]$$

Similar relations are obtained for the components of $[\hat{\ell}_i, \hat{p}_k]$. From these commutation relations, one directly derives

$$[\hat{\ell}_x, \hat{\ell}_y] = i\hbar\hat{\ell}_z \qquad [\hat{\ell}_y, \hat{\ell}_z] = i\hbar\hat{\ell}_x \qquad [\hat{\ell}_z, \hat{\ell}_x] = i\hbar\hat{\ell}_y \qquad [29]$$

The same results are obtained for the operators $\hat{\mathscr{L}}_x$, $\hat{\mathscr{L}}_y$, and $\hat{\mathscr{L}}_z$ since the angular momentum operators for different particles commute.

$$[\hat{\mathscr{L}}_x, \hat{\mathscr{L}}_y] = i\hbar\hat{\mathscr{L}}_z \qquad [\hat{\mathscr{L}}_y, \hat{\mathscr{L}}_z] = i\hbar\hat{\mathscr{L}}_x, \qquad [\hat{\mathscr{L}}_z, \hat{\mathscr{L}}_x] = i\hbar\hat{\mathscr{L}}_y \qquad [30]$$

From Eqs. [30], we conclude that the three components of the angular momentum cannot be determined simultaneously—with the trivial exception when all three components are zero.

By contrast, the square modulus of the orbital angular momentum (see Eq. [22]) commutes with all three components of $\hat{\vec{\mathscr{L}}}$, that is,

$$[\hat{\mathscr{L}}_x, \hat{\mathscr{L}}^2] = 0 \qquad [\hat{\mathscr{L}}_y, \hat{\mathscr{L}}^2] = 0 \qquad [\hat{\mathscr{L}}_z, \hat{\mathscr{L}}^2] = 0 \qquad [31]$$

General Angular Momenta

As mentioned above, the definition of an angular momentum is a direct consequence of the isotropy of space, and this property leads directly to the commutation relations. Instead of relying on the special properties of the orbital angular momentum, we shall solve the eigenvalue problem of the angular momentum solely based on the commutation relations that are common to all types of angular momenta. This approach has the advantage that cases with half-integer angular momentum quantum numbers are included; these are related to spin and do not occur for pure orbital angular momenta.

Step/Shift/Ladder and Tensor Operators

For the determination of matrix elements, it is often more convenient to use linear combinations of the Cartesian components of the angular momentum operator instead of the Cartesian components themselves. In the literature, two different kinds of operators are employed. The first type is defined by

$$\hat{\mathscr{J}}_+ = \hat{\mathscr{J}}_x + i\hat{\mathscr{J}}_y \qquad \hat{\mathscr{J}}_0 = \hat{\mathscr{J}}_z \qquad \hat{\mathscr{J}}_- = \hat{\mathscr{J}}_x - i\hat{\mathscr{J}}_y \qquad [32]$$

$\hat{\mathscr{J}}_+$ and $\hat{\mathscr{J}}_-$ are called *step-up and step-down operators*, respectively, or *shift operators* for reasons to become clear soon. These operators are also denominated *ladder operators*. $\hat{\mathscr{J}}_+$ and $\hat{\mathscr{J}}_-$ are not self-adjoint, that is, $(\hat{\mathscr{J}}_+)^* \neq \hat{\mathscr{J}}_+$ and so on; instead $\hat{\mathscr{J}}_+$ is the complex conjugate of $\hat{\mathscr{J}}_-$ and vice versa. Together with $\hat{\mathscr{J}}_z$ they form a linear independent set of components of $\hat{\vec{\mathscr{J}}}$. In terms of the shift operators, $\hat{\mathscr{J}}^2$ can be expressed as

$$\hat{\mathscr{J}}^2 = \hat{\mathscr{J}}_+\hat{\mathscr{J}}_- + \hat{\mathscr{J}}_z^2 - \hbar\hat{\mathscr{J}}z \qquad [33]$$

$$= \hat{\mathscr{J}}_-\hat{\mathscr{J}}_+ + \hat{\mathscr{J}}_z^2 + \hbar\hat{\mathscr{J}}_z \qquad [34]$$

$$= \frac{1}{2}(\hat{\mathscr{J}}_+\hat{\mathscr{J}}_- + \hat{\mathscr{J}}_-\hat{\mathscr{J}}_+) + \hat{\mathscr{J}}_z^2 \qquad [35]$$

Each of these forms will be used later.

A second set of operators, the so-called *tensor operators*, differ only slightly from the ladder operators. They are introduced here without further

explanation. Cartesian and spherical tensor components are related by

$$\hat{\mathscr{J}}_{+1} = -\frac{\hat{\mathscr{J}}_x + i\hat{\mathscr{J}}_y}{\sqrt{2}} \qquad \hat{\mathscr{J}}_0 = \hat{\mathscr{J}}_z \qquad \hat{\mathscr{J}}_{-1} = \frac{\hat{\mathscr{J}}_x - i\hat{\mathscr{J}}_y}{\sqrt{2}} \qquad [36]$$

or conversely

$$\mathscr{J}_x = \frac{\mathscr{J}_{-1} - \mathscr{J}_{+1}}{\sqrt{2}} \qquad \mathscr{J}_z = \mathscr{J}_0 \qquad \mathscr{J}_y = \frac{i(\mathscr{J}_{-1} + \mathscr{J}_{+1})}{\sqrt{2}} \qquad [37]$$

Phase conventions have been chosen to be consistent with those of Condon and Shortley.[13] In terms of tensor operators, the square modulus of $\vec{\hat{\mathscr{J}}}$ becomes

$$\hat{\mathscr{J}}^2 = -\hat{\mathscr{J}}_{+1}\hat{\mathscr{J}}_{-1} - \hat{\mathscr{J}}_{-1}\hat{\mathscr{J}}_{+1} + \hat{\mathscr{J}}_0^2 \qquad [38]$$

We shall come back to these operators after learning what a tensor is.

Commutation Relations

As for the orbital angular momentum, the commutation relations between the Cartesian components of a general angular momentum $\vec{\hat{\mathscr{J}}}$ and its square modulus $\hat{\mathscr{J}}^2$ read

$$[\hat{\mathscr{J}}_x, \hat{\mathscr{J}}_y] = i\hbar\hat{\mathscr{J}}_z \qquad [\hat{\mathscr{J}}_y, \hat{\mathscr{J}}_z] = i\hbar\hat{\mathscr{J}}_x \qquad [\hat{\mathscr{J}}_z, \hat{\mathscr{J}}_x] = i\hbar\hat{\mathscr{J}}_y \qquad [39]$$

and

$$[\hat{\mathscr{J}}_x, \hat{\mathscr{J}}^2] = 0 \qquad [\hat{\mathscr{J}}_y, \hat{\mathscr{J}}^2] = 0 \qquad [\hat{\mathscr{J}}_z, \hat{\mathscr{J}}^2] = 0 \qquad [40]$$

Using Eq. [39] and the definition of the step-up and step-down operators (Eq. [32]), one easily obtains their commutation relations

$$[\hat{\mathscr{J}}_+, \hat{\mathscr{J}}_-] = 2\hbar\hat{\mathscr{J}}_z \qquad [41]$$

and

$$[\hat{\mathscr{J}}_z, \hat{\mathscr{J}}_\pm] = \pm\hbar\hat{\mathscr{J}}_\pm \qquad [42]$$

From the latter, the useful relation

$$[\hat{\mathscr{J}}_z, \hat{\mathscr{J}}_\pm^n] = \pm n\hbar\hat{\mathscr{J}}_\pm \qquad [43]$$

can be derived.

Like the Cartesian components, the shift operators also commute with $\hat{\mathscr{J}}^2$

$$[\hat{\mathscr{J}}_{\pm}, \hat{\mathscr{J}}^2] = 0 \qquad [44]$$

The same is true for the tensor operators.

$$[\hat{\mathscr{J}}_{\pm 1}, \hat{\mathscr{J}}^2] = 0 \qquad [45]$$

The commutation relations among their components differ slightly from those of the shift operators. From Eqs. [36] and [39], it follows that

$$[\hat{\mathscr{J}}_{+1}, \hat{\mathscr{J}}_{-1}] = -\hbar \hat{\mathscr{J}}_0 \qquad [46]$$

and

$$[\hat{\mathscr{J}}_0, \hat{\mathscr{J}}_{\pm 1}] = \pm \hbar \hat{\mathscr{J}}_{\pm 1} \qquad [47]$$

The Eigenvalues of $\hat{\mathscr{J}}^2$ and $\hat{\mathscr{J}}_z$

Because $\hat{\mathscr{J}}^2$ and $\hat{\mathscr{J}}_z$ commute, they must have common eigenvectors that we shall denote by $|u\rangle$. The eigenvectors satisfy the equations

$$\hat{\mathscr{J}}^2 |u\rangle = a\hbar^2 |u\rangle \qquad [48]$$

$$\hat{\mathscr{J}}_z |u\rangle = M\hbar |u\rangle \qquad [49]$$

Applying Eqs. [42] to the eigenvectors $|u\rangle$

$$\hat{\mathscr{J}}_z \hat{\mathscr{J}}_{\pm} |u\rangle = \hat{\mathscr{J}}_{\pm} \hat{\mathscr{J}}_z |u\rangle \pm \hbar \hat{\mathscr{J}}_{\pm} |u\rangle = (M \pm 1)\hbar \hat{\mathscr{J}}_{\pm} |u\rangle \qquad [50]$$

we see that $\hat{\mathscr{J}}_{\pm}|u\rangle$ are also eigenvectors of $\hat{\mathscr{J}}_z$, but with eigenvalue $(M \pm 1)\hbar$. Moreover, $\hat{\mathscr{J}}_{\pm}|u\rangle$ is also an eigenvector of $\hat{\mathscr{J}}^2$ with eigenvalue $a\hbar^2$ because

$$\hat{\mathscr{J}}^2 \hat{\mathscr{J}}_{\pm} |u\rangle = \hat{\mathscr{J}}_{\pm} \hat{\mathscr{J}}^2 |u\rangle = a\hbar^2 \hat{\mathscr{J}}_{\pm} |u\rangle \qquad [51]$$

This means that the action of $\hat{\mathscr{J}}_+$ on an eigenvector of $\hat{\mathscr{J}}_z$ and $\hat{\mathscr{J}}^2$ is to step up the eigenvalue of $\hat{\mathscr{J}}_z$ by one unit while remaining within the subset of functions belonging to the eigenvalue a of $\hat{\mathscr{J}}^2$. $\hat{\mathscr{J}}_-$ analogously steps down the eigenvalue of $\hat{\mathscr{J}}_z$ by one unit.

To set up a connection between a and M, we express $\hat{\mathscr{J}}^2$ in the form Eq. [35] and note that $\hat{\mathscr{J}}_+$ and $\mathscr{J}_-$ are Hermitian conjugates. Applying the turnover rule (Eq. [17]) yields

$$\langle \hat{\mathscr{J}} \phi | \hat{\mathscr{J}} \phi \rangle = \frac{1}{2} \langle \hat{\mathscr{J}}_- \phi | \hat{\mathscr{J}}_- \phi \rangle + \frac{1}{2} \langle \hat{\mathscr{J}}_+ \phi | \hat{\mathscr{J}}_+ \phi \rangle + \langle \hat{\mathscr{J}}_z \phi | \hat{\mathscr{J}}_z \phi \rangle \qquad [52]$$

Each of these terms is always positive or zero—this follows from the fact that the length of a vector is always positive or zero. The fact that each of the terms in Eq. [52] is positive or zero leads us to the inequality

$$\langle\phi|\hat{\mathscr{J}}^2|\phi\rangle \geq \langle\phi|\hat{\mathscr{J}}_z^2|\phi\rangle \geq 0 \qquad [53]$$

By identifying $|\phi\rangle$ with an eigenvector $|u\rangle$, we obtain from Eqs. [48] and [49]

$$a \geq M^2 \geq 0 \qquad [54a]$$

or

$$-\sqrt{a} \leq M \leq \sqrt{a} \qquad [54b]$$

which means that the eigenvalues M are limited from above and below; there is a minimal value $M_{\min}$ and a maximal value $M_{\max}$.

In particular, if we apply Eq. [50] to an eigenvector $|u_{\max}\rangle$ belonging to $M_{\max}$

$$\hat{\mathscr{J}}_z\hat{\mathscr{J}}_+|u_{\max}\rangle = (M_{\max}+1)\hbar\hat{\mathscr{J}}_+|u_{\max}\rangle \qquad [56]$$

then this equation can only be satisfied, if

$$\hat{\mathscr{J}}_+|u_{\max}\rangle = 0 \qquad [57]$$

because $M_{\max}$ was assumed to be the maximal M value. Using expression [34] yields

$$\hat{\mathscr{J}}_-\hat{\mathscr{J}}_+|u_{\max}\rangle = (\hat{\mathscr{J}}^2 - \hat{\mathscr{J}}_z^2 - \hbar\hat{\mathscr{J}}_z)|u_{\max}\rangle = 0 \qquad [58]$$

leading to

$$a\hbar^2 - M_{\max}^2\hbar^2 - M_{\max}\hbar^2 = 0 \qquad [59]$$

From this, we obtain the relation between a and $M_{\max}$

$$a = M_{\max}(M_{\max}+1) \qquad [60]$$

Analogously, application of expression [33] to $|u_{\min}\rangle$ yields

$$a = M_{\min}(M_{\min}-1) \qquad [61]$$

Finally, stepping down $|u_{\max}\rangle$ repeatedly by applying Eq. [43]

$$\hat{\mathscr{J}}_z\hat{\mathscr{J}}_-^n|u_{\max}\rangle = (M_{\max} - n)\hbar\hat{\mathscr{J}}_-^n|u_{\max}\rangle \quad [62]$$

we can always find a positive integer n (the largest possible) such that

$$M_{\max} - n = M_{\min} \quad [63]$$

leading to

$$M_{\max}(M_{\max} + 1) = M_{\min}(M_{\min} - 1) = (M_{\max} - n)(M_{\max} - n - 1) \quad [64]$$

or

$$M_{\max} = \frac{n}{2} \quad [65]$$

In the following, we shall denote this maximal value $M_{\max}$ by J. This maximum J can only be an integer or a half-integer

$$J = 0, \frac{1}{2}, 1, \frac{3}{2}, \cdots \quad [66]$$

The eigenvalue a from Eq. [48] adopts the values $J(J + 1)$ and $M_{\min} = -J$. For M, we thus obtain the values

$$M = J, J - 1 \cdots - J \quad [67]$$

Finally, if we denominate the corresponding eigenvectors by $|u_J^M\rangle$, the eigenvalue equations of the angular momentum operators read

$$\hat{\mathscr{J}}^2|u_J^M\rangle = J(J + 1)\hbar^2|u_J^M\rangle \quad [68]$$

$$\hat{\mathscr{J}}_z|u_J^M\rangle = M\hbar|u_J^M\rangle \quad [69]$$

Because $\hat{\mathscr{J}}_z$ is Hermitian and each $|u_J^M\rangle$ belongs to a different eigenvalue, the eigenvectors $|u_J^M\rangle$ are orthogonal; after normalization we obtain

$$\langle u_J^M|u_{J'}^{M'}\rangle = \delta_{JJ'}\delta^{MM'} \quad [70]$$

The Action of $\hat{\mathscr{J}}_\pm$, $\hat{\mathscr{J}}_x$, and $\hat{\mathscr{J}}_y$ on Eigenvectors of $\hat{\mathscr{J}}^2$ and $\hat{\mathscr{J}}_z$

According to Eqs. [50] and [51], $\hat{\mathscr{J}}_\pm|u_J^M\rangle$ is both an eigenvector of $\hat{\mathscr{J}}_z$ with eigenvalue $(M \pm 1)\hbar$ and an eigenvector of $\hat{\mathscr{J}}^2$ with eigenvalue

$J(J+1)\hbar^2$. Thus $\hat{\mathscr{J}}_\pm|u_J^M\rangle$ and $|u_J^{M\pm1}\rangle$ are parallel, and we can determine the proportionality constants from the normalization constraint:

$$\langle\hat{\mathscr{J}}_\pm u_J^M|\hat{\mathscr{J}}_\pm u_J^M\rangle = \langle u_J^M|\hat{\mathscr{J}}_\mp\hat{\mathscr{J}}_\pm|u_J^M\rangle \quad [71]$$

$$= \langle u_J^M|(\hat{\mathscr{J}}^2 - \hat{\mathscr{J}}_z^2 \mp \hbar\hat{\mathscr{J}}_z)|u_J^M\rangle \quad [72]$$

$$= \big(J(J+1) - M^2 \mp M\big)\hbar^2 \quad [73]$$

Up to an arbitrary phase factor, this yields

$$\hat{\mathscr{J}}_\pm|u_J^M\rangle = \hbar\sqrt{J(J+1) - M(M\pm1)}\,|u_J^{M\pm1}\rangle \quad [74]$$

Accordingly, by inverting the definition of the ladder operators in terms of the Cartesian components, we can determine the actions of $\hat{\mathscr{J}}_x$ and $\hat{\mathscr{J}}_y$ on $|u_J^M\rangle$:

$$\hat{\mathscr{J}}_x|u_J^M\rangle = \frac{1}{2}\hbar\Big\{\sqrt{J(J+1) - M(M+1)}|u_J^{M+1}\rangle + \sqrt{J(J+1) - M(M-1)}|u_J^{M-1}\rangle\Big\} \quad [75]$$

$$\hat{\mathscr{J}}_y|u_J^M\rangle = -\frac{i}{2}\hbar\Big\{\sqrt{J(J+1) - M(M+1)}|u_J^{M+1}\rangle - \sqrt{J(J+1) - M(M-1)}|u_J^{M-1}\rangle\Big\} \quad [76]$$

Matrix Elements

By using the results of the last two subsections, the matrix elements of the angular momentum operators are easily determined. The $|u_J^M\rangle$ are eigenvectors of $\hat{\mathscr{J}}^2$ and $\hat{\mathscr{J}}_z$. Therefore, nonvanishing integrals of $\hat{\mathscr{J}}^2$ and $\hat{\mathscr{J}}_z$ are confined to the diagonal of the matrix.

$$\langle u_J^M|\hat{\mathscr{J}}^2|u_{J'}^{M'}\rangle = \hbar^2 J(J+1)\delta_{JJ'}\delta^{MM'} \quad [77]$$

and

$$\langle u_J^M|\hat{\mathscr{J}}_z|u_{J'}^{M'}\rangle = \hbar M\delta_{JJ'}\delta^{MM'} \quad [78]$$

In the $\hat{\mathscr{J}}^2$, $\hat{\mathscr{J}}_z$ representation, the shift operators $\hat{\mathscr{J}}_\pm$ have only off-diagonal matrix elements

$$\langle u_J^M|\hat{\mathscr{J}}_\pm|u_{J'}^{M'}\rangle = \hbar\sqrt{J(J+1) - M'(M'\pm1)}\delta_{JJ'}\delta^{MM'\mp1} \quad [79]$$

The same is true for the Cartesian $\hat{\mathscr{J}}_x$ and $\hat{\mathscr{J}}_y$

$$\langle u_J^M|\hat{\mathscr{J}}_x|u_{J'}^{M'}\rangle = \frac{1}{2}\hbar\sqrt{J(J+1)-M'(M'+1)}\delta_{JJ'}\delta^{MM'\mp 1} + \frac{1}{2}\hbar\sqrt{J(J+1)-M'(M'-1)}\delta_{JJ'}\delta^{MM'\pm 1} \quad [80]$$

$$\langle u_J^M|\hat{\mathscr{J}}_y|u_{J'}^{M'}\rangle = -\frac{i}{2}\hbar\sqrt{J(J+1)-M'(M'+1)}\delta_{JJ'}\delta^{MM'\mp 1} + \frac{i}{2}\hbar\sqrt{J(J+1)-M'(M'-1)}\delta_{JJ'}\delta^{MM'\pm 1} \quad [81]$$

In particular, for $J = \frac{1}{2}$ states we obtain

$\langle u_J^M|\hat{\mathscr{J}}^2|u_J^{M'}\rangle =$

M	M' $\frac{1}{2}$	$-\frac{1}{2}$
$\frac{1}{2}$	$\frac{3}{4}\hbar^2$	0
$-\frac{1}{2}$	0	$\frac{3}{4}\hbar^2$

[82]

$\langle u_J^M|\hat{\mathscr{J}}_z|u_J^{M'}\rangle =$

M	M' $\frac{1}{2}$	$-\frac{1}{2}$
$\frac{1}{2}$	$\frac{1}{2}\hbar$	0
$-\frac{1}{2}$	0	$-\frac{1}{2}\hbar$

[83]

$\langle u_J^M|\hat{\mathscr{J}}_+|u_J^{M'}\rangle =$

M	M' $\frac{1}{2}$	$-\frac{1}{2}$
$\frac{1}{2}$	0	$\hbar$
$-\frac{1}{2}$	0	0

[84]

$\langle u_J^M|\hat{\mathscr{J}}_-|u_J^{M'}\rangle =$

M	M' $\frac{1}{2}$	$-\frac{1}{2}$
$\frac{1}{2}$	0	0
$-\frac{1}{2}$	$\hbar$	0

[85]

$\langle u_J^M|\mathscr{J}_x|u_J^{M'}\rangle =$

M	M' $\frac{1}{2}$	$-\frac{1}{2}$
$\frac{1}{2}$	0	$\frac{1}{2}\hbar$
$-\frac{1}{2}$	$\frac{1}{2}\hbar$	0

[86]

$\langle u_J^M|\hat{\mathscr{J}}_y|u_J^{M'}\rangle =$

M	M' $\frac{1}{2}$	$-\frac{1}{2}$
$\frac{1}{2}$	0	$-\frac{i}{2}\hbar$
$-\frac{1}{2}$	$\frac{i}{2}\hbar$	0

[87]

A Pictorial Representation

In the previous sections, we learned that the modulus of the angular momentum $|\vec{\hat{\mathscr{J}}}|$ amounts to $\sqrt{J(J+1)}\hbar$ and that the eigenvalues of $\hat{\mathscr{J}}^2$ are $(2J+1)$-fold degenerate. Thus, if we measure the modulus of the angular momentum, we know that the state vector is located somewhere in the $2J+1$ dimensional vector space $\mathbf{u}_J$ of $\hat{\mathscr{J}}^2$. If we determine also the component of $\vec{\hat{\mathscr{J}}}$ along a direction (which we call the z axis), we will find values of $M\hbar$. Unlike the components of the linear momentum, not all components of the angular momentum can be determined simultaneously. Thus, it is not possible to represent the measured values of the quantum mechanical angular momentum as an arrow. Conveniently, the angular momentum is visualized as a cone (see Figure 8) with the axis oriented along the direction of the measured component (z axis), height $M\hbar$, and radius $\sqrt{J(J+1)-M^2}\,\hbar$. All that we can say is that the angular momentum vector lies somewhere on this cone; its x and y components remain undetermined.

We shall frequently encounter cases with angular momentum values of $J=\frac{1}{2}$ and $J=1$. For $J=\frac{1}{2}$, such as for electron spin, there are only two possible orientations of the angular momentum; they are depicted in Figure 9. Commemorative of the old Bohr–Sommerfeld theory, we say that the angular momentum is oriented *parallel* to the z axis in the case of $M=+\frac{1}{2}$ and *antiparallel* for $M=-\frac{1}{2}$.

In the case of $J=1$, three possibilities arise (Figure 10): an upward cone for $M=1$, a downward cone for $M=-1$, and a cone with height 0—thus reducing to a disk—for $M=0$.

Spin Angular Momentum

As mentioned earlier, we cannot make use of the correspondence principle to derive quantum mechanical spin operators, because spin has no classical analog. Instead, the spin eigenfunctions $|sm_s\rangle$ may be identified with $|u_{1/2}^{\pm 1/2}\rangle$

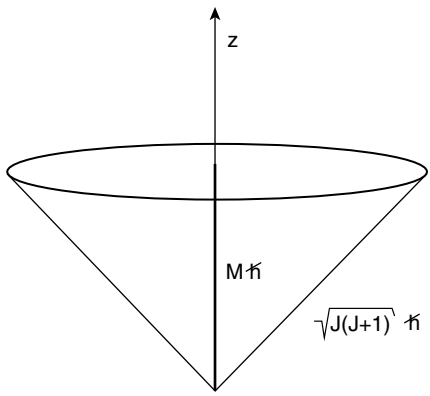

Figure 8 The cone of an angular momentum vector.

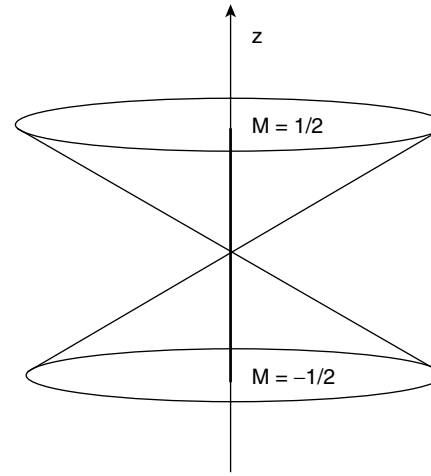

Figure 9 The two possible orientations of $J = \frac{1}{2}$.

and Eqs. [82]–[87] are then employed to define a matrix representation of the spin operators.

Spinors and Spin Operators

Obviously, the spin eigenfunction $|sm_s\rangle$ is not a function of the spatial coordinates; mathematically it is known as a *spinor*. Different notations are

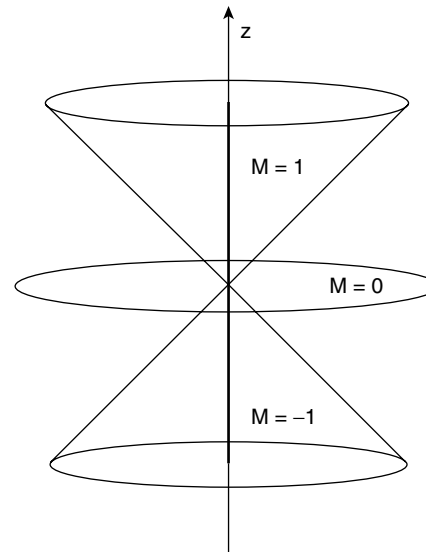

Figure 10 The three possible orientations of $J = 1$.

in common usage:

$$|sm_s\rangle = \begin{cases} |\tfrac{1}{2} \quad \tfrac{1}{2}\rangle \equiv \begin{pmatrix} 1 \\ 0 \end{pmatrix} \equiv \alpha \\ |\tfrac{1}{2} \quad -\tfrac{1}{2}\rangle \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix} \equiv \beta \end{cases} \qquad [88]$$

The symbols α and β are the ones most familiar to chemists. For the definition of spin operators, it is convenient to utilize the representation of the spin eigenfunctions as the orthonormal basis vectors of a two-dimensional (2D) vector space. In this representation, the spin operators may be written as matrices

$$\hat{\mathcal{S}}^2 = \hbar^2 \begin{pmatrix} \frac{3}{4} & 0 \\ 0 & \frac{3}{4} \end{pmatrix} \qquad [89]$$

$$\hat{\mathcal{S}}_0 = \hbar \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix} \qquad \hat{\mathcal{S}}_+ = \hbar \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \qquad \hat{\mathcal{S}}_- = \hbar \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \qquad [90]$$

$$\hat{\mathcal{S}}_x = \hbar \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix} \qquad \hat{\mathcal{S}}_y = \hbar \begin{pmatrix} 0 & -\frac{i}{2} \\ \frac{i}{2} & 0 \end{pmatrix} \qquad \hat{\mathcal{S}}_z = \hbar \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix} \qquad [91]$$

acting on the column vectors $\binom{1}{0}$ and $\binom{0}{1}$ by means of the usual matrix-vector product.

For later convenience, we also define the irreducible tensor operators

$$\hat{\mathcal{S}}_0 = \hbar \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix} \qquad \hat{\mathcal{S}}_{+1} = \hbar \begin{pmatrix} 0 & -\frac{1}{\sqrt{2}} \\ 0 & 0 \end{pmatrix} \qquad \hat{\mathcal{S}}_{-1} = \hbar \begin{pmatrix} 0 & 0 \\ \frac{1}{\sqrt{2}} & 0 \end{pmatrix} \qquad (92)$$

Pauli Spin Matrices

Pauli introduced slightly different spin operators known as the *Pauli spin matrices*. They are defined by

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad [93]$$

Apart from the numerical factor $2\hbar$, they are in fact identical with the Cartesian spin operators in Eq. [91]. The difference between $\vec{\sigma}$ and $\vec{\hat{S}}$ appears to be trivial; nevertheless, it is important to recognize that the operator $\vec{\sigma}$ does not qualify as an angular momentum operator because it does not fulfill the usual

commutation relations Eq. [39]. Instead of

$$\vec{\hat{\mathcal{S}}} \times \vec{\hat{\mathcal{S}}} = i\hbar\vec{\hat{\mathcal{S}}} \qquad [94]$$

the commutation relations for $\vec{\sigma}$ are given by

$$\vec{\sigma} \times \vec{\sigma} = 2i\vec{\sigma} \qquad [95]$$

A very useful equation employing the Pauli spin matrices is the so-called *Dirac relation*. For any pair of vector operators $\vec{u}$ and $\vec{v}$

$$(\vec{\sigma u})(\vec{\sigma v}) = \vec{u}\vec{v} + i\vec{\sigma}(\vec{u} \times \vec{v}) \qquad [96]$$

Further, it is easily seen that the square modulus of $\vec{\sigma}$ equals the unit matrix in two dimensions $\mathbf{1}_2$

$$\vec{\sigma} \cdot \vec{\sigma} = \mathbf{1}_2 \qquad [97]$$

Repeated application of the dot product yields

$$\vec{\sigma}^n = \begin{cases} \mathbf{1}_2 & \text{for even } n \\ \vec{\sigma} & \text{for odd } n \end{cases} \qquad [98]$$

SPIN–ORBIT HAMILTONIANS

Spin–orbit coupling arises naturally in Dirac theory, which is a fully relativistic one-particle theory for spin $\frac{1}{2}$ systems.[11] In one-electron atoms, spin s and orbital angular momentum ℓ of the electron are not separately conserved; they are coupled and only the resulting total electronic angular momentum j is a "good" quantum number.

The construction of relativistic two-body equations is an area of current research in relativistic field theory.[14] To the author's knowledge, a fully Lorentz invariant multielectron Hamiltonian (i.e., a multielectron Hamiltonian with correct relativistic transformation properties) has not yet been derived. In molecular applications approximate two-particle Hamiltonians are therefore utilized. The so-called Dirac–Coulomb–Breit operator contains a sum of Dirac terms representing the kinetic energy and the one-electron interactions. In nonrelativistic theories, a Coulomb term describes the instantaneous interaction of two electrons. One can imagine this interaction as being mediated by an exchange of "interaction particles", the so-called virtual photons. In relativistic theories, an instantaneous process cannot take place; even the exchange of virtual photons is limited by the speed of light. This

slowed exchange of virtual photons is taken into account by means of the so-called retardation term.[15] At the same level of approximation, a magnetic interaction term appears, first introduced by Gaunt in 1929.[16] Together, these two correction terms are named the *Breit interaction*. The resulting many-particle four-component Hamiltonian—the Dirac–Coulomb–Breit Hamiltonian—is frequently used as a starting point for deriving more approximate Hamiltonians that are suited for chemical applications.

Full One- and Two-Electron Spin–Orbit Operators

Spin–orbit interaction Hamiltonians are most elegantly derived by reducing the relativistic four-component Dirac–Coulomb–Breit operator to two components and separating spin-independent and spin-dependent terms. This reduction can be achieved in many different ways; for more details refer to the recent literature (e.g., Refs. 17–21).

The Breit–Pauli Spin–Orbit Hamiltonian

The most renowned one- and two-electron spin–orbit Hamiltonian is given by

$$\hat{\mathscr{H}}_{\mathrm{SO}}^{\mathrm{BP}} = \frac{e^2}{2m_e^2c^2}\left\{\sum_i\left(-\vec{\nabla}_i\left(\sum_I \frac{Z_I}{\hat{r}_{iI}}\right)\times\vec{\hat{p}}_i\right)\cdot\vec{\hat{s}}_i\right. \qquad [99]$$

$$+\sum_i\sum_{j\neq i}\left(\vec{\nabla}_i\left(\frac{1}{\hat{r}_{ij}}\right)\times\vec{\hat{p}}_i\right)\cdot\vec{\hat{s}}_i \qquad [100]$$

$$+\sum_i\sum_{j\neq i}\left(\vec{\nabla}_j\left(\frac{1}{\hat{r}_{ij}}\right)\times\vec{\hat{p}}_j\right)\cdot\vec{\hat{s}}_i \qquad [101]$$

$$\left.+\sum_j\sum_{i\neq j}\left(\vec{\nabla}_i\left(\frac{1}{\hat{r}_{ji}}\right)\times\vec{\hat{p}}_i\right)\cdot\vec{\hat{s}}_j\right\} \qquad [102]$$

where I labels nuclei and i and j electrons, and where Z_I is the charge of nucleus I, e and m_e are the charge and the mass of an electron, respectively, and c is the speed of light. The so-called Breit–Pauli spin–orbit operator was derived originally by Pauli in 1927,[22] one year before the advent of four-component Dirac theory.[11] Pauli started from the Schrödinger equation of a molecule in an external electric and magnetic field. The Breit-Pauli operator is then obtained by assuming that the scalar potential is purely Coulombic and the magnetic field arises from the electronic spin associated with a magnetic moment $\vec{\mu}_e = 2(e\hbar/2m_ec)\vec{S}$. Terms [99] and [100] constitute the spin–same-orbit part. Herein, the one-electron operator [99] describes the interaction of the spin magnetic moment of an electron i with the magnetic moment that arises from its orbiting in the field of the nucleus I. Term [100] is the two-electron analog relating to the motion of electron i in the field of electron

j. Operators [101] and [102] are usually called spin–other-orbit terms. They describe the coupling between the spin magnetic moment of electron *i* and the orbital magnetic moment of electron *j* and vice versa.

From the four-component Dirac–Coulomb–Breit equation, the terms [99]–[102] can be deduced without assuming external fields. A Foldy–Wouthuysen transformation[23] of the electron-nuclear Coulomb attraction and collecting terms to order v^2/c^2 yields the one-electron part [99]. Similarly, the two-electron part [100] of the spin–same–orbit operator stems from the transformation of the two-electron Coulomb interaction. The spin–other-orbit terms [101] and [102] have a different origin. They result, among other terms, from the reduction of the Gaunt interaction.

The Breit-Pauli spin–orbit Hamiltonian is found in many different forms in the literature. In expressions [101] and [102], we have chosen a form in which the connection to the Coulomb potential and the symmetry in the particle indices is apparent. Mostly $\hat{\mathscr{H}}_{\mathrm{SO}}^{\mathrm{BP}}$ is written in a short form where spin–same- and spin–other-orbit parts of the two-electron Hamiltonian have been contracted to a single term, either as

$$\hat{\mathscr{H}}_{\mathrm{SO}}^{\mathrm{BP}} = \frac{e^2}{2m_e^2c^2}\sum_i\left\{\sum_I Z_I\left(\frac{\vec{\hat{r}}_{iI}}{\hat{r}_{iI}^3}\times\vec{\hat{p}}_i\right)\vec{\hat{s}}_i - \sum_{j\neq i}\left(\frac{\vec{\hat{r}}_{ij}}{\hat{r}_{ij}^3}\times\vec{\hat{p}}_i\right)\cdot(\vec{\hat{s}}_i + 2\vec{\hat{s}}_j)\right\} \qquad [103]$$

or

$$\hat{\mathscr{H}}_{\mathrm{SO}}^{\mathrm{BP}} = \frac{e^2}{2m_e^2c^2}\sum_i\left\{\sum_I Z_I\left(\frac{\vec{\hat{r}}_{iI}}{\hat{r}_{iI}^3}\times\vec{\hat{p}}_i\right) - \sum_{j\neq j}\frac{\vec{\hat{r}}_{ij}}{\hat{r}_{ij}^3}\times(\vec{\hat{p}}_i - 2\vec{\hat{p}}_j)\right\}\cdot\vec{\hat{s}}_i \qquad [104]$$

The Breit-Pauli spin–orbit operator has one major drawback. It implicitly contains terms coupling electronic states (with positive energy) and positronic states (in the negative energy continuum) and is thus unbounded from below. It can be employed safely only in first-order perturbation theory.

The No-Pair Spin–Orbit Hamiltonian

A more appropriate spin–orbit coupling Hamiltonian can be derived if electron–positron pair creation processes are excluded right from the beginning (no-pair approximation). After projection on the positive energy states, a variationally stable Hamiltonian is obtained if one avoids expansion in reciprocal powers of *c*. Instead the Hamiltonian is transformed by properly chosen

unitary operators. If closed expressions cannot be obtained, one rather expands in the coupling strength Ze^2.[24–26]

$$\hat{\mathscr{H}}^{+}_{\mathrm{SO}} = e^2\hbar c \sum_i \left\{ \sum_I Z_I \, \frac{\hat{A}_i}{\hat{E}_i + m_e c^2} \left(\frac{\vec{\hat{r}}_{iI}}{\hat{r}^3_{iI}} \times \vec{\hat{p}}_i \right) \vec{\hat{s}}_i \, \frac{\hat{A}_i}{\hat{E}_i + m_e c^2} - \sum_{j \neq i} \frac{\hat{A}_i \hat{A}_j}{\hat{E}_i + m_e c^2} \left[\left(\frac{\vec{\hat{r}}_{ij}}{\hat{r}^3_{ij}} \times \vec{\hat{p}}_i \right) (\vec{\hat{s}}_i + 2\vec{\hat{s}}_j) \right] \frac{\hat{A}_i \hat{A}_j}{\hat{E}_i + m_e c^2} \right\} \qquad [105]$$

Herein, $\hat{E}_i$ denotes the relativistic kinetic energy operator

$$\hat{E}_i = \sqrt{\hat{p}_i^2 c^2 + m_e^2 c^4} \qquad [106]$$

and the factors $\hat{A}_i$ or $\hat{A}_j$ are defined as

$$\hat{A}_i = \sqrt{\frac{\hat{E}_i + m_e c^2}{2\hat{E}_i}} \qquad [107]$$

As may be seen by comparing Eqs. [103] and [105], the no-pair spin–orbit Hamiltonian has exactly the same structure as the Breit-Pauli spin–orbit Hamiltonian. It differs from the Breit-Pauli operator only by kinematical factors that damp the $1/\hat{r}^3_{iI}$ and $1/\hat{r}^3_{ij}$ singularities.

Valence-Only Spin–Orbit Hamiltonians

The computational effort of a molecular calculation can be reduced significantly, if only a few electrons are taken into account explicitly and the interaction with the rest is approximated by means of effective Hamiltonians. The first step in the hierarchy of approximations is the so-called frozen-core approximation.

The Spin-Free Frozen-Core Approximation

In electron correlation treatments, it is a common procedure to divide the orbital space into various subspaces: orbitals with large binding energy (core), occupied orbitals with low-binding energy (valence), and unoccupied orbitals (virtual). One of the reasons for this subdivision is the possibility to "freeze" the core (i.e., to restrict excitations to the valence and virtual spaces). Consequently, all determinants in a configuration interaction (CI) expansion share a set of frozen-core orbitals. For this approximation to be valid, one has to assume that excitation energies are not affected by correlation contributions of the inner shells. It is then sufficient to describe the interaction between core and valence electrons by some kind of mean-field expression.

Let us first become familiar with the spin-free (sf) case: we have a pair of Slater determinants interacting via $\hat{\mathscr{H}}_{+}^{\mathrm{sf}}$ (i.e., a sum of spin-independent one- and two-electron operators, $\hat{h}_1$ and $\hat{g}_{12}$). Their matrix element is given by (Slater–Condon rules):

1. Diagonal case:

$$\langle\Phi|\hat{\mathscr{H}}_{+}^{\mathrm{sf}}|\Phi\rangle = \sum_i \langle i(1)|\hat{h}_1|i(1)\rangle + \sum_i \sum_{j>i} \{\langle i(1)j(2)|\hat{g}_{12}|i(1)j(2)\rangle - \langle i(1)j(2)|\hat{g}_{12}|j(1)i(2)\rangle\} \quad [108]$$

   The summation indices i and j run over all occupied spin orbitals in the determinant Φ.
2. Single excitations: Φ and Φ_i^j differ by one pair of spin orbitals i, j.

$$\langle\Phi|\hat{\mathscr{H}}_{+}^{\mathrm{sf}}|\Phi_i^j\rangle = \langle i(1)|\hat{h}_1|j(1)\rangle + \sum_k \{\langle i(1)k(2)|\hat{g}_{12}|j(1)k(2)\rangle - \langle i(1)k(2)|\hat{g}_{12}|k(1)j(2)\rangle\} \quad [109]$$

   Here the summation index k runs over all common spin orbitals.
3. Double excitations: Φ and Φ_{ik}^{jl} differ by two pairs of spin orbitals ij, kl.

$$\langle\Phi|\hat{\mathscr{H}}_{+}^{\mathrm{sf}}|\Phi_{ik}^{jl}\rangle = \langle i(1)k(2)|\hat{g}_{12}|j(1)l(2)\rangle - \langle i(1)k(2)|\hat{g}_{12}|l(1)j(2)\rangle \quad [110]$$

For convenience, let us separate core and valence (val) contributions in these expressions and use the short-hand bra–ket notation $\langle i|j\rangle$ for the one-electron integral $\langle i(1)|\hat{h}_1|j(1)\rangle$, and $\langle ik|jl\rangle$ is defined as $\langle i(1)k(2)|\hat{g}_{12}|j(1)l(2)\rangle$. If core excitations are not allowed, one obtains:

1. Diagonal case:

$$\langle\Phi|\hat{\mathscr{H}}_{+}^{\mathrm{sf}}|\Phi\rangle = \sum_i^{\mathrm{val}} \left\{ \langle i|i\rangle + \sum_{j>i}^{\mathrm{val}} \{\langle ij|ij\rangle - \langle ij|ji\rangle\} \right. \quad [111]$$

$$\left. + \sum_j^{\mathrm{core}} \{\langle ij|ij\rangle - \langle ij|ji\rangle\} \right\} \quad [112]$$

$$+ \sum_i^{\mathrm{core}} \left\{ \langle i|i\rangle + \sum_{j>i}^{\mathrm{core}} \{\langle ij|ij\rangle - \langle ij|ji\rangle\} \right\} \quad [113]$$

2. Single excitations: Φ and Φ_i^j differ by one pair of valence spin orbitals i, j.

$$\langle\Phi|\hat{\mathscr{H}}_+^{\mathrm{sf}}|\Phi_i^j\rangle = \langle i|j\rangle + \sum_k^{\mathrm{val}}\{\langle ik|jk\rangle - \langle ik|kj\rangle\} \quad [114]$$

$$+ \sum_k^{\mathrm{core}}\{\langle ik|jk\rangle - \langle ik|kj\rangle\} \quad [115]$$

3. Double excitations: Φ and Φ_{ik}^{jl} differ by two pairs of valence spin orbitals ij, kl.

$$\langle\Phi|\hat{\mathscr{H}}_+^{\mathrm{sf}}|\Phi_{ik}^{jl}\rangle = \langle ik|jl\rangle - \langle ik|lj\rangle \quad [116]$$

We see that integrals with core indices contribute solely to the diagonal and the single excitation cases. Because frozen-core orbitals are common to all determinants by definition, we may interchange the sums over determinants and core orbitals. The pure core terms [113] are constant and are conveniently added to the nuclear repulsion. After summing over the core indices, the mixed core–valence integrals, [112] and [115], can be added to the corresponding one-electron integrals $\langle i|i\rangle$ and $\langle i|j\rangle$, respectively. One has to be aware, however, that this summation includes an integration over the spin of the core electrons. For a Coulomb integral $\langle ik|jk\rangle$, the same rule applies as for the corresponding one-electron integral $\langle i|j\rangle$; it survives if the spin of orbitals i and j are equal, whatever the spin of k may be. Because each spatial core orbital is occupied by two electrons, it contributes two Coulomb integrals. Exchange integrals $\langle ik|kj\rangle$, on the other hand, integrate to zero if the spin of k differs from the spin of i and j. Therefore only one of the exchange integrals will survive. Since the Hamiltonian is spin-independent, the spin part integrates to one, irrespective whether it amounts to $\langle\alpha\alpha|\alpha\alpha\rangle$ or $\langle\beta\beta|\beta\beta\rangle$.

The Frozen-Core Spin–Orbit Hamiltonian

The frozen-core (fc) approach is not restricted to spin-independent electronic interactions; the spin–orbit (SO) interaction between core and valence electrons can be expressed by a sum of Coulomb- and exchange-type operators. The matrix element formulas can be derived in a similar way as the Slater–Condon rules.[27] Here, it is not important whether the Breit-Pauli spin–orbit operators or their no-pair analogs are employed as these are structurally equivalent. Differences with respect to the Slater–Condon rules occur due to the symmetry properties of the angular momentum operators and because of the presence of the spin–other-orbit interaction. It is easily shown by partial integration that the linear momentum operator $\hat{\vec{p}}$ is antisymmetric with respect to orbital exchange, and the same applies to $\vec{\hat{\ell}} = \vec{r} \times \vec{\hat{p}}$. Therefore, spin–orbit

Hamiltonians do not have diagonal matrix elements in a basis of real (Cartesian) orbitals. Further, the use of two-electron operators of the forms [103] or [105] requires a symmetrization in the particle indices

$$\hat{g}_{12} \rightarrow \frac{1}{2}(\hat{g}_{12} + \hat{g}_{21}) \qquad [117]$$

thus doubling the number of two-electron contributions. Combined with the above mentioned integral matrix properties, this leads to the following expressions for a spin–orbit matrix element between a pair of Slater determinants built from real spin orbitals:

1. Diagonal case:

$$\langle \Phi | \hat{\mathscr{H}}_{\mathrm{SO}} | \Phi \rangle = 0 \qquad [118]$$

2. Single excitations: Φ and Φ_i^j differ by one pair of spin orbitals i, j. Contributions come from the one-electron spin–orbit integral and three-index two-electron integrals

$$\begin{aligned}\langle \Phi | \hat{\mathscr{H}}_{\mathrm{SO}} | \Phi_i^j \rangle = \langle i(1) | \hat{h}_{\mathrm{SO}}(1) | j(1) \rangle + \frac{1}{2} \sum_k \{ & \langle i(1)k(2) | \hat{\mathscr{H}}_{\mathrm{SO}}(1,2) | j(1)k(2) \rangle \\ & - \langle k(1)i(2) | \hat{\mathscr{H}}_{\mathrm{SO}}(1,2) | j(1)k(2) \rangle \\ & - \langle i(1)k(2) | \hat{\mathscr{H}}_{\mathrm{SO}}(1,2) | k(1)j(2) \rangle \}\end{aligned} \qquad [119]$$

   The summation index k runs over common spin orbitals. The second Coulomb-type integral $\langle k(1)i(2) | \hat{\mathscr{H}}_{\mathrm{SO}}(1,2) | k(1)j(2) \rangle$ vanishes because $\hat{\mathscr{H}}$ is linear in $\vec{\hat{p}}_1$.
3. Double excitations: Φ and Φ_{ik}^{jl} differ by two pairs of spin orbitals ij, kl.

Only four-index two-electron integrals contribute to the spin–orbit coupling matrix element:

$$\begin{aligned}\langle \Phi | \hat{\mathscr{H}}_{\mathrm{SO}} | \Phi_{ik}^{jl} \rangle = \frac{1}{2} \{ & \langle i(1)k(2) | \hat{\mathscr{H}}_{\mathrm{SO}}(1,2) | j(1)l(2) \rangle + \langle k(1)i(2) | \hat{\mathscr{H}}_{\mathrm{SO}}(1,2) | l(1)j(2) \rangle \\ & - \langle k(1)i(2) | \hat{\mathscr{H}}_{\mathrm{SO}}(1,2) | j(1)l(2) \rangle - \langle i(1)k(2) | \hat{\mathscr{H}}_{\mathrm{SO}}(1,2) | l(1)j(2) \rangle \}\end{aligned} \qquad [120]$$

In these formulas, a spin integration has not yet been performed.

Core excitation processes, of course, rule out the possibility of applying a frozen-core approximation. Otherwise the indices i and j represent valence orbitals. The same applies to k and l in the double excitation case. The only

Table 1 One-Electron Spin Integrals [$\hbar$] for the $\vec{\hat{s}}_1$ Tensor Operator

	$\|\alpha(1)\rangle$	$\|\beta(1)\rangle$
$\langle\alpha(1)\|$	$1/2$	$-1/\sqrt{2}$
$\langle\beta(1)\|$	$1/\sqrt{2}$	$-1/2$

core contributions occur for singly excited determinants—and they are by far the largest terms in the sum over k in Eq. [119]. Summing up all core contributions to a given index pair i, j, one is left with an effective one-electron integral that may be contracted with the proper one-electron term $\langle i|\hat{h}_{\mathrm{so}}|j\rangle$.

Spin integration is more elaborate than for a spin-free Hamiltonian.[28] Assume for the moment that i and j represent α orbitals. According to Table 1 the one-electron spin integral amounts to $\frac{1}{2}$. The frozen-core orbitals are doubly occupied and contribute with both α and β spin parts. From Table 2, we see that the spin parts of the Coulomb-type integral $\langle i(1)k(2)|\hat{\mathscr{H}}_{\mathrm{SO}}(1,2)|j(1)k(2)\rangle$, $\langle\alpha\alpha|\alpha\alpha\rangle$ and $\langle\alpha\beta|\alpha\beta\rangle$, integrate to $\frac{3}{2}$ and $-\frac{1}{2}$, respectively, yielding $\frac{2}{2}$ in total. Because the effective one-electron integral will later be multiplied by $\frac{1}{2}$, the prefactor of the Coulomb-type integral has to be 2. The first exchange integral $\langle i(1)k(2)|\hat{\mathscr{H}}_{\mathrm{SO}}|k(1)j(2)\rangle$ is connected to the spin integrals $\langle\alpha\alpha|\alpha\alpha\rangle$ and $\langle\alpha\beta|\beta\alpha\rangle$. The former amounts to $\frac{3}{2}$, whereas the latter one vanishes, thus yielding a prefactor of -3 for this exchange integral. The same is true for the second exchange integral. Contrary to the matrix elements of a spin-independent Hamiltonian, spin–orbit coupling matrix elements do not automatically vanish if i and j represent orbitals with different spins, say α and β. The spatial one-electron integral is then to be multiplied by a spin factor of $-1/\sqrt{2}$. The spin parts of the Coulomb-type integral, $\langle\alpha\alpha|\beta\alpha\rangle$ and $\langle\alpha\beta|\beta\beta\rangle$, are equal in this case and add up to a total of $-2/\sqrt{2}$. This value is again twice that of the one-electron spin integral. The spin parts of the exchange integrals yield $-3/\sqrt{2}$; thus, like in the previous case, -3 times the spatial exchange integral must be added to the effective one-electron integral. By continuing with the case studies, we find that the prefactor of the spatial Coulomb-type integral always amounts to 2, whereas it assumes a value of -3 for the exchange integrals. A matrix element of the effective frozen-core spin–orbit Hamiltonian is therefore

Table 2 Two-Electron Spin Integrals [$\hbar$] for the $\vec{\hat{s}}_1 + 2\vec{\hat{s}}_2$ Operator

	$\|\alpha(1)\alpha(2)\rangle$	$\|\alpha(1)\beta(2)\rangle$	$\|\beta(1)\alpha(2)\rangle$	$\|\beta(1)\beta(2)\rangle$
$\langle\alpha(1)\alpha(2)\|$	$3/2$	$-2/\sqrt{2}$	$-1/\sqrt{2}$	0
$\langle\alpha(1)\beta(2)\|$	$2/\sqrt{2}$	$-1/2$	0	$-1/\sqrt{2}$
$\langle\beta(1)\alpha(2)\|$	$1/\sqrt{2}$	0	$1/2$	$-2/\sqrt{2}$
$\langle\beta(1)\beta(2)\|$	0	$1/\sqrt{2}$	$2/\sqrt{2}$	$-3/2$

given by

$$\langle i(1)|\hat{\mathscr{H}}_{\mathrm{SO}}^{\mathrm{fc}}|j(1)\rangle = \langle i(1)|\hat{h}_{\mathrm{SO}}(1)|j(1)\rangle + \frac{1}{2}\sum_{k}^{\mathrm{fc}}\{2\langle i(1)k(2)|\hat{\mathscr{H}}_{\mathrm{SO}}(1,2)|j(1)k(2)\rangle - 3\langle k(1)i(2)|\hat{\mathscr{H}}_{\mathrm{SO}}(1,2)|j(1)k(2)\rangle - 3\langle i(1)k(2)|\hat{\mathscr{H}}_{\mathrm{SO}}(1,2)|k(1)j(2)\rangle\} \quad [121]$$

Frozen-core orbitals are doubly occupied and a spin integration has been performed for the core electrons. The summation index k therefore runs over spatial orbitals only. Employing the frozen-core approximation considerably shortens the summation procedure in single excitation cases. Double excitation cases are left unaltered at this level of approximation. The computational effort can be substantially reduced further if one manages to get rid of all explicit two-electron terms.

Effective One-Electron Spin–Orbit Hamiltonians

Since spin–orbit coupling is very important in heavy element compounds and the structure of the full microscopic Hamiltonians is rather complicated, several attempts have been made to develop approximate one-electron spin–orbit Hamiltonians. The application of an (effective) one-electron spin–orbit Hamiltonian has several computational advantages in spin–orbit CI or perturbation calculations: (1) all integrals may be kept in central memory, (2) there is no need for a summation over common indices in singly excited Slater determinants, and (3) matrix elements coupling doubly excited configurations do not occur. In many approximate schemes, even the tedious four-index transformation of two-electron integrals ceases to apply. The central question that comes up in this context deals with the accuracy of such an approximation, of course.

Empirical One-Electron Operators

One of the simplest and least demanding approaches is to take the electron–electron interactions into account through screening of the nuclear potential.

$$\hat{\mathscr{H}}_{\mathrm{SO}}^{\mathrm{eff}} = \frac{e^2}{2m_e^2c^2}\sum_{I}\sum_{i}{}' \frac{Z_{I,\ell}^{\mathrm{eff}}}{r_{I_{i_I}}^3}\,\vec{\hat{\ell}}_{I_{i_I}}\vec{\hat{s}}_{i_I} \quad [122]$$

In this one-electron one-center spin–orbit operator, I denotes an atom and i_I an electron occupying an orbital located at center I. Likewise, $\vec{\hat{\ell}}_{I_{i_I}}$ labels the angular momentum of electron i_I with respect to the orbital origin at atom I. The

prime on the second summation (over electrons) indicates that it includes only the open shell of an atom I with azimuthal quantum number ℓ.

The effective nuclear charge $Z_{I,\ell}^{\text{eff}}$ is parameterized to fit experimental fine-structure splittings of one or more electronic states on atom I. Frequently, this is achieved by estimating a spin–orbit coupling parameter on the basis of the Landé interval rule (see the later subsection on this rule). This procedure has serious drawbacks, however. Open-shell electronic states of heavy atoms, in particular, are often not well separated energetically and may interact via spin–orbit coupling. In these cases, the multiplet splitting does not follow the Landé pattern, according to which neighboring fine-structure levels ought to be separated by $E(J) - E(J-1) = A_{\text{SO}}J$ in first order. Furthermore, in light atoms spin–spin coupling contributes to the deviation of the atomic splitting from the Landé interval rule.

Higher-order spin–orbit effects can be determined by quasi–degenerate perturbation theory (see the section on Computational Aspects). In this approach, an interaction matrix is set up including all electrostatic and spin–orbit coupling matrix elements between the states under consideration. Diagonalization of this matrix directly yields the fine-structure splitting. The reverse procedure may be used for fitting the effective nuclear charge in expression [122]. In principle, the interaction matrix can be diagonalized for various coupling strengths, and the spin–orbit parameter is then fitted to reproduce the splitting in a least-squares sense. However, the deperturbation of an experimental spectrum is complicated by the fact that higher-order spin–orbit shifts depend both on the location of fictitious spin-free states and the spin–orbit coupling parameter. If the spin–orbit splitting is small compared to the term energy separation by Coulombic repulsion, a weighted average of the term energies may serve as zero-order energies.[29] For heavy atoms, the latter approximation is not applicable; instead excitation energies are determined from correlated ab initio calculations employing a spin-free Hamiltonian. Examples for the latter procedure may be found in Refs. 30–32.

Blume and Watson[33] showed in 1962 that only parts of the two-electron spin–orbit interaction may be expressed in the form [122]. According to these authors, the neglect of the remaining terms leads to a dependence of the effective nuclear charge from a particular state. For d^n configurations, for example, they found that $Z_{I,\ell}^{\text{eff}}$ can vary by as much as 20%.[34]

Pseudo-Potential Spin–Orbit Operators

Most common among the approximate spin–orbit Hamiltonians are those derived from relativistic effective core potentials (RECPs).[35–38] Spin–orbit coupling operators for pseudo-potentials were developed in the 1970s.[39,40] In the meantime, different schools have devised different procedures for tailoring such operators. All these procedures to parameterize the spin–orbit interaction for pseudo-potentials have one thing in common: The predominant action of the spin–orbit operator has to be transferred from

the region close to the nucleus to the valence region. The reason for this relocation is simply the fact that amplitudes of pseudo-orbitals are small in the proximity of the nucleus. If the original $1/r^3$ dependence is left untouched as in the semiempirical expression [122], huge effective charges may result.

Teichteil et al.[41] fit a spin–orbit pseudo-operator such that its action on a pseudo-orbital optimally reproduces the effect of the true spin–orbit operator on the corresponding all-electron orbital. Ermler, Ross, Christiansen, and co-workers[42–49] and Titov and Mosyagin[50] define a spin–orbit operator as the difference between the ℓ and j dependent relativistic effective pseudopotentials (REPs)[51]

$$\hat{\mathscr{H}}_{\mathrm{SO}}(r) = \sum_{\ell=1}^{\ell_{\max}} \Delta V_\ell(r) \left(\frac{\ell}{2\ell+1} \sum_{m_j=-j}^{j} |\ell, j, m_j\rangle\langle \ell, j, m_j| - \frac{\ell+1}{2\ell+1} \sum_{m_j'=-j'}^{j'} |\ell, j', m_j'\rangle\langle \ell, j', m_j'| \right) \quad [123]$$

with

$$j = \ell + (1/2), \qquad j' = \ell - (1/2) \quad [124]$$

and

$$\Delta V_\ell(r) = V_{\ell,\ell+1/2}(r) - V_{\ell,\ell-1/2}(r) \quad [125]$$

This latter difference is generally fitted to a linear combination of a few basis functions. According to Pitzer and Winter,[52] the operator [123] can be written equivalently as

$$\hat{\mathscr{H}}_{\mathrm{SO}}(r) = \sum_{\ell=1}^{\ell_{\max}} \frac{2\Delta V_\ell(r)}{2\ell+1} \vec{\ell} \cdot \vec{s} \sum_{m_\ell=-\ell}^{+\ell} |\ell, m_\ell\rangle\langle \ell, m_\ell| \quad [126]$$

$$= \sum_{\ell=1}^{\ell_{\max}} \xi_\ell(r) \vec{\ell} \cdot \vec{s} \sum_{m_\ell=-\ell}^{+\ell} |\ell, m_\ell\rangle\langle \ell, m_\ell| \quad [127]$$

Stoll, Dolg, and co-workers[53–58] also make use of this form to define an effective spin–orbit Hamiltonian. However, they fit the parameters in expression [127] directly to atomic fine-structure splittings. Seijo,[59] on the other hand, determines the parameters through least-squares fitting to the radial parts of the numerical Wood–Boring spin–orbit operators and introduces a correction through a change of AO basis set contraction coefficients.

Spin–Orbit Mean-Field Operator

One of the most rigorous approaches toward the definition of an effective one-electron spin–orbit operator appears to be the recently proposed

spin–orbit mean-field Hamiltonian.[60] It constitutes an extension of the frozen-core approximation.

In our effort to get rid of all explicit two-electron terms, we may in most cases neglect four-index spin–orbit integrals (case 3, Eq. [120]), since they make a minor contribution to the total spin–orbit coupling matrix element between two CI wave functions. However, unacceptable errors result—in particular for molecules built from light constituents—if also all three-index two-electron valence integrals (case 2, Eq. [119]) are set to zero. A much better effective one-electron spin–orbit Hamiltonian is obtained when the frozen-core summation over k in ansatz [121] is extended to the valence space and an average valence orbital occupation—corresponding, for instance, to the Hartree–Fock configuration—is assumed. Partially filled orbitals are taken into account through occupation numbers n_k, ranging from 0 to 2. Further, the two-electron spin–orbit interaction of partially filled orbitals k is averaged over α and β spin orientations before spin-integration. A matrix element of this effective one-electron spin–orbit mean-field (mf) Hamiltonian is given by

$$\begin{aligned}\langle i(1)|\hat{\mathscr{H}}_{\mathrm{SO}}^{mf}|j(1)\rangle = \left\langle i(1)|\hat{h}_{SO}(1)|j(1)\right\rangle + \frac{1}{2}\sum_k n_k \{ & 2\langle i(1)k(2)|\hat{\mathscr{H}}_{\mathrm{SO}}(1,2)|j(1)k(2)\rangle \\ & - 3\langle k(1)i(2)|\hat{\mathscr{H}}_{\mathrm{SO}}(1,2)|j(1)k(2)\rangle \\ & - 3\langle i(1)k(2)|\hat{\mathscr{H}}_{\mathrm{SO}}(1,2)|k(1)j(2)\rangle\}\end{aligned} \qquad [128]$$

The summation index k now runs over all spatial molecular orbitals.

The molecular spin–orbit mean-field Hamiltonian appears to approximate very well the full one- and two-particle operator. For the platinum atom, a deviation from the full operator result of less than 4 cm^{-1} out of a total of 4000 cm^{-1} was observed (i.e., the error amounts to less than 0.1%).[60] Even light molecules, where the two-electron contributions to the total matrix element are in the order of 50%, may be treated in this approximation if the leading configurations of the considered electronic states are singly excited with respect to each other. The singlet–triplet transition matrix element of ethene and the ground-state splitting of HC_6H^+, NC_5H^+, and NC_4N^+ fall into this class;[61,62] here errors did not exceed 1%.

Like all effective one-electron approaches, the mean-field approximation considerably quickens the calculation of spin–orbit coupling matrix elements. Nevertheless, the fact that the construction of the molecular mean-field necessitates the evaluation of two-electron spin–orbit integrals in the complete AO basis represents a serious bottleneck in large applications. An enormous speed-up can be achieved if a further approximation is introduced and the molecular mean field is replaced by a sum of atomic mean fields. In this case, only two-electron integrals for basis functions located at the same center have to be evaluated. This idea is based on two observations: first, the spin–orbit Hamiltonian exhibits a $1/r^3$ radial dependence and falls off much faster

than the Coulomb interaction, and second, the largest two-electron contributions stem from the molecular core orbitals that are fairly localized and closely resemble their atomic counterparts. The two-electron one-center approximation works very well for heavy element compounds.[60] Larger percentage errors result for molecules built up from light elements. Thorough investigations on a variety of light molecules show that multicenter one- and two-electron contributions partially compensate in a systematic manner.[62,63] In these cases, it appears to be advantageous either to neglect all multicenter contributions, if demanded by the problem size, or to take into account all of them, if high accuracy is required.

The one-center approximation allows for an extremely rapid evaluation of spin–orbit mean-field integrals if the atomic symmetry is fully exploited.[64] Even more efficiency may be gained, if also the spin-independent core-valence interactions are replaced by atom-centered effective core potentials (ECPs). In this case, the inner shells do not even emerge in the molecular orbital optimization step, and the size of the atomic orbital basis set can be kept small. A prerequisite for the use of the all-electron atomic mean-field Hamiltonian in ECP calculations is to find a prescription for setting up a correspondence between the valence orbitals of the all-electron and ECP treatments.[65–67]

SYMMETRY

All the terms in the molecular Hamiltonian transform as the totally symmetric irreducible representation (irrep) of the molecular point group. Symbolically,

$$\Gamma(\hat{\mathscr{H}}_{\mathrm{SO}}) = \Gamma_1 \qquad [129]$$

For a matrix element of the spin–orbit operator to be different from zero, it is in principle sufficient that the direct products of the irreps of space and spin functions on both sides of the matrix element be equal. Thus,

$$\langle \Psi_a | \hat{\mathscr{H}}_{\mathrm{SO}} | \Psi_b \rangle \neq 0 \quad \text{if} \quad \Gamma(\mathrm{Space}_a) \otimes \Gamma(\mathrm{Spin}_a) = \Gamma(\mathrm{Space}_b) \otimes \Gamma(\mathrm{Spin}_b) \qquad [130]$$

Representation theory of molecular point groups tells us how a rotation or a reflection of a molecule can be represented as an orthogonal transformation in 3D coordinate space. We can therefore easily determine the irreducible representation for the spatial part of the wave function. By contrast, a spin eigenfunction is not a function of the spatial coordinates. If we want to study the transformation properties of the spinors

$$\alpha = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \beta = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad [131]$$

under symmetry operations of some molecular point group, we will have to find a mapping between the transformation matrices **R** in 3D coordinate space and corresponding matrices **u** in a properly chosen 2D vector space.

The fact that the magnetic interaction Hamiltonians are compound tensor operators can be exploited to derive more specific selection rules than the one given above. Furthermore, as we shall see later, the number of matrix elements between multiplet components that actually have to be computed can be considerably reduced by use of the Wigner–Eckart theorem.

Transformation Properties of the Wave Function

We are not going to review here the transformation properties of spatial wave functions under the symmetry operations of molecular point groups. To prepare the discussion of the transformation properties of spinors, we shall put some effort, however, in discussing the symmetry operations of $O(3)^+$, the group of proper rotations in 3D coordinate space (i.e., orthogonal transformations with determinant $+1$). Reflections and improper rotations (orthogonal transformations with determinant -1) will be dealt with later.

Rotations in Three-Dimensional Coordinate Space

Any rotation in 3D space can be composed of successive rotations through the three Euler angles (see Figure 11):

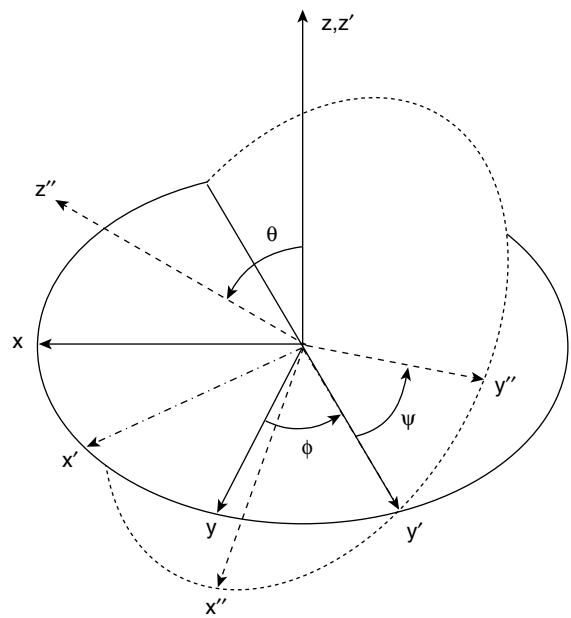

Figure 11 Definition of the three Euler angles.

1. Rotation about the z axis through an angle ϕ ($0 \leq \phi \leq 2\pi$).
2. Rotation about the new position of the y axis (y' in Figure 11) through an angle θ ($0 \leq \theta \leq \pi$).
3. Rotation about the new position of the z axis (z'' in Figure 11) through an angle ψ ($0 \leq \psi \leq 2\pi$).

Alternatively, we can view the angles θ and ϕ as defining the orientation of a general rotation axis z'', about which a rotation through ψ is to be performed, relative to a space-fixed coordinate system x, y, z.

The first and the third of the Euler angles, ϕ and ψ, refer to rotations about z axes. The corresponding transformation matrices in 3D coordinate space read

$$\mathbf{R}_z(\phi) = \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \mathbf{R}_{z''}(\psi) = \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad [132]$$

A rotation through the Euler angle θ refers to the y' axis and is represented by

$$\mathbf{R}_{y'}(\theta) = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \qquad [133]$$

These three rotations do in general not commute, that is, the product depends on the order in which the matrices are multiplied. The representation matrix of the complete transformation can be composed from these individual rotation matrices as

$$\mathbf{R}(\phi, \theta, \psi) = \mathbf{R}_{z''}(\psi)\mathbf{R}_{y'}(\theta)\mathbf{R}_z(\phi) \qquad [134]$$

In the earlier section relating to an infinitesimal rotation, we introduced a relation between the operator for an infinitesimal rotation through $\delta\varphi$ about a general axis $\vec{n}$ and the angular momentum operator $\hat{\vec{\mathscr{L}}}$. We can carry this connection further and find a corresponding expression for a finite rotation through a finite angle γ

$$\Psi(\vec{r} + \vec{\delta r}) = \hat{U}_{\mathbf{R}_{\vec{n}}(\gamma)}\Psi(\vec{r}) \qquad [135]$$

To this end, we develop $\Psi(\vec{r} + \vec{\delta r})$ into a Taylor series.

$$\begin{aligned} \Psi(\vec{r} + \vec{\delta r}) &= \Psi(\vec{r}) + \vec{\delta r} \cdot \vec{\nabla}\Psi(\vec{r}) + \frac{1}{2}(\vec{\delta r} \cdot \vec{\nabla})^2\Psi(\vec{r}) + \cdots \\ &= \left(1 + \vec{\gamma} \cdot (\vec{r} \times \vec{\nabla}) + \frac{1}{2}(\vec{\gamma} \cdot (\vec{r} \times \vec{\nabla}))^2 + \cdots\right)\Psi(\vec{r}) \end{aligned} \qquad [136]$$

or after introducing the projection of the angular momentum operator on the rotation axis

$$\Psi(\vec{r}+\vec{\delta r}) = \left(1 + \frac{i}{\hbar}\gamma\,\vec{n}\cdot\vec{\hat{\mathscr{L}}} + \frac{1}{2}\left(\frac{i}{\hbar}\gamma\,\vec{n}\cdot\vec{\hat{\mathscr{L}}}\right)^2 + \cdots\right)\Psi(\vec{r}) \qquad [137]$$

The term in parentheses is the desired operator for finite rotations. It is often written in a short form as

$$\hat{\mathscr{U}}_{\mathbf{R}_{\vec{n}}(\gamma)} = \exp\left(\frac{i}{\hbar}\gamma\,\vec{n}\cdot\vec{\hat{\mathscr{L}}}\right) \qquad [138]$$

where the exponential of an operator or a matrix is defined via its Taylor expansion. The operator for the three successive finite rotations through the Euler angles can therefore be expressed as

$$\hat{\mathscr{U}}_{\mathbf{R}(\phi,\theta,\psi)} = \exp\left(\frac{i}{\hbar}\psi\hat{\mathscr{L}}_{z''}\right)\exp\left(\frac{i}{\hbar}\theta\hat{\mathscr{L}}_{y'}\right)\exp\left(\frac{i}{\hbar}\phi\hat{\mathscr{L}}_{z}\right) \qquad [139]$$

Spinor Transformations

We can employ an analog of the spatial transformation operator [138] for analyzing the transformation properties of a spinor under coordinate rotations. The evaluation of the corresponding 2D transformation matrices is simplified if we rewrite

$$\hat{\mathscr{U}}_{\mathbf{u}_n(\gamma)} = \exp\left(\frac{i}{\hbar}\gamma\,\vec{n}\cdot\vec{\hat{s}}\right) \qquad [140]$$

by making use of the relation between the spin operators and the Pauli matrices $\vec{\hat{s}} = \hbar\,\vec{\sigma}/2$

$$\hat{\mathscr{U}}_{\mathbf{u}_n(\gamma)} = \exp(i\gamma\,\vec{n}\cdot\vec{\sigma}/2) \qquad [141]$$

In expanding the Taylor series, we exploit Eq. [98]:

$$\begin{aligned}\hat{\mathscr{U}}_{\mathbf{u}_n(\gamma)} &= \mathbf{1}_2 + \frac{i}{1!}\vec{n}\cdot\vec{\sigma}\left(\frac{\gamma}{2}\right) - \frac{1}{2!}\mathbf{1}_2\left(\frac{\gamma}{2}\right)^2 - \frac{i}{3!}\vec{n}\cdot\vec{\sigma}\left(\frac{\gamma}{2}\right)^3 + \frac{1}{4!}\mathbf{1}_2\left(\frac{\gamma}{2}\right)^4 + \cdots \\ &= \mathbf{1}_2\cos\left(\frac{\gamma}{2}\right) + i\vec{n}\cdot\vec{\sigma}\sin\left(\frac{\gamma}{2}\right)\end{aligned} \qquad [142]$$

For a rotation about the Cartesian axes through an angle γ, we obtain

x axis:

$$\hat{U}_{\mathbf{u}_x(\gamma)} = \mathbf{1}_2 \cos\left(\frac{\gamma}{2}\right) + i\sigma_x \sin\left(\frac{\gamma}{2}\right) = \begin{pmatrix} \cos\left(\frac{\gamma}{2}\right) & i\sin\left(\frac{\gamma}{2}\right) \\ i\sin\left(\frac{\gamma}{2}\right) & \cos\left(\frac{\gamma}{2}\right) \end{pmatrix} \quad [143]$$

y axis:

$$\hat{U}_{\mathbf{u}_y(\gamma)} = \mathbf{1}_2 \cos\left(\frac{\gamma}{2}\right) + i\sigma_y \sin\left(\frac{\gamma}{2}\right) = \begin{pmatrix} \cos\left(\frac{\gamma}{2}\right) & \sin\left(\frac{\gamma}{2}\right) \\ -\sin\left(\frac{\gamma}{2}\right) & \cos\left(\frac{\gamma}{2}\right) \end{pmatrix} \quad [144]$$

z axis:

$$\hat{\mathscr{U}}_{\mathbf{u}_z(\gamma)} = \mathbf{1}_2 \cos\left(\frac{\gamma}{2}\right) + i\sigma_z \sin\left(\frac{\gamma}{2}\right) = \begin{pmatrix} \exp\left(\frac{i\gamma}{2}\right) & 0 \\ 0 & \exp\left(\frac{-i\gamma}{2}\right) \end{pmatrix} \quad [145]$$

The matrix $\mathbf{u}(\phi, \theta, \psi)$ corresponding to the Euler rotation matrix $\mathbf{R}(\phi, \theta, \psi)$ is thus found to be

$$\mathbf{u}(\phi, \theta, \psi) = \begin{pmatrix} \exp\left(\frac{i\psi}{2}\right)\cos\left(\frac{\theta}{2}\right)\exp\left(\frac{i\phi}{2}\right) & \exp\left(\frac{i\psi}{2}\right)\sin\left(\frac{\theta}{2}\right)\exp\left(\frac{-i\phi}{2}\right) \\ -\exp\left(\frac{-i\psi}{2}\right)\sin\left(\frac{\theta}{2}\right)\exp\left(\frac{i\phi}{2}\right) & \exp\left(\frac{-i\psi}{2}\right)\cos\left(\frac{\theta}{2}\right)\exp\left(\frac{-i\phi}{2}\right) \end{pmatrix} \quad [146]$$

It has to be noted that the relation between the elements of $O(3)^+$ (also called SO(3), the group representing proper rotations in 3D coordinate space) and SU(2) (the special unitary group in two dimensions) is not a one-to-one correspondence. Rather, each $\mathbf{R}$ matches two matrices $\pm\mathbf{u}$. Molecular point groups including symmetry operations for spinors therefore exhibit two times as many elements as ordinary point groups and are dubbed double groups.

Double Groups

In ordinary space, a rotation through 2π is identical to unity. The matrix $\mathbf{u}$ representing this operation in spinor space is not equal to the unit matrix in two dimensions, but rather $-\mathbf{1}_2$ is obtained. The identity corresponds to a rotation through 4π instead. The distinction between a rotation through 2π, denoted by the symbol $\hat{\bar{E}}$, and the identity operation $\hat{E}$ leads to a reduplication of the number of molecular point group operations. Irreducible representations (irreps) of double groups can be subdivided into two types: *boson* irreps and *fermion* irreps. The boson irreps, corresponding to the normal molecular point group irreps, exhibit real positive characters for the $\hat{\bar{E}}$ operation. All *spatial* wave functions as well as *spin* functions for an *even* number of electrons transform according to one of the boson irreps. *Spin* functions for an *odd* number of electrons belong to one of the fermion irreps.

We exemplify the procedure of determining the spinor transformation properties under molecular point group operations for the $\bar{C}_{2v}$ double group. Other double groups can be treated analogously. The character tables of the 32 molecular double groups may be found, e.g., in Ref. 68.

The C_{2v} molecular point group is of order four with symmetry operations:

$\hat{E}$ the identity, also corresponding to a rotation through 2π.
$\hat{C}_2(z)$ rotation about the z axis through π.
$\hat{\sigma}_{xz}$ reflection in the xz plane.
$\hat{\sigma}_{yz}$ reflection in the yz plane.

Please, note that in this context $\hat{\sigma}$ denotes a reflection and not one of the Pauli matrices. The group has four one-dimensional (1D) irreducible representations. For convenience, its character table is displayed in Table 3.

The corresponding double group $\bar{C}_{2v}$ comprises eight symmetry operations:

$\hat{E}$ the identity, also corresponding to a rotation through 4π.
$\hat{\bar{E}}$ a rotation through 2π.
$\hat{C}_2(z)$ rotation about the z axis through π.
$\hat{\bar{C}}_2(z)$ $= \hat{\bar{E}} \cdot \hat{C}_2(z)$.
$\hat{\sigma}_{xz}$ reflection in the xz plane.
$\hat{\bar{\sigma}}_{xz}$ $= \hat{\bar{E}} \cdot \hat{\sigma}_{xz}$.
$\hat{\sigma}_{yz}$ reflection in the yz plane.
$\hat{\bar{\sigma}}_{yz}$ $= \hat{\bar{E}} \cdot \hat{\sigma}_{yz}$.

Let us set up a 2D unitary matrix representation for the transformation of the spin functions α and β in $\bar{C}_{2v}$. So far, we have established only a relation between $O(3)^+$ and SU(2). The matrix representations of reflections or improper rotations do not belong to $O(3)^+$ because their determinants have a value of -1. To find out how α and β behave under reflections, we notice that any reflection in a plane can be thought of as a rotation through π about an axis perpendicular to that plane followed by the inversion operation. For instance, $\hat{\sigma}_{xz}$ may be constructed as $\hat{\sigma}_{xz} = \hat{C}_2(y) \cdot \hat{i}$. Herein, it is not necessarily required

Table 3 Character Table of the C_{2v} Point Group

C_{2v}	$\hat{E}$	$\hat{C}_2(z)$	$\hat{\sigma}_{xz}$	$\hat{\sigma}_{yz}$
A_1	1	1	1	1
A_2	1	1	-1	-1
B_1	1	-1	1	-1
B_2	1	-1	-1	1

Table 4 The $E_{1/2}$ Representation of $\bar{C}_{2v}$

$\hat{O}$	$E_{1/2}$ Representation	$\hat{O}$	$E_{1/2}$ Representation
$\hat{E}$	$\mathbf{u}(4\pi) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\hat{\bar{E}}$	$\mathbf{u}(2\pi) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$
$\hat{C}_2$	$\mathbf{u}_z(\pi) = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$	$\hat{\bar{C}}_2$	$\mathbf{u}_z(3\pi) = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}$
$\hat{\sigma}_{xz}$	$\mathbf{u}_y(\pi) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$	$\hat{\bar{\sigma}}_{xz}$	$\mathbf{u}_y(3\pi) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$
$\hat{\sigma}_{yz}$	$\mathbf{u}_x(\pi) = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$	$\hat{\bar{\sigma}}_{yz}$	$\mathbf{u}_x(3\pi) = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}$

that this particular rotation and the inversion are contained separately in the considered point group. Further, the fact that angular momenta are invariant with respect to inversion can be exploited. By utilizing this prescription, we obtain a 2D representation of $\bar{C}_{2v}$ (Table 4), which is equivalent to the $E_{1/2}$ irreducible representation.

The character table of $\bar{C}_{2v}$ (Table 5) is easily constructed considering that in this group the barred and unbarred symmetry operations belong to the same class, with the exception of $\hat{E}$ and $\hat{\bar{E}}$, which always represent a class of their own.

The spin functions α and β (i.e., the components of a spin doublet) belong to the $E_{1/2}$ irrep of $\bar{C}_{2v}$. But what about singlet and triplet spin functions? For this purpose, we look at the action of the symmetry operators on typical two-electron spin functions such as $\alpha\alpha$ and $\alpha\beta$. The results, displayed in Table 6, are easily verified.

A look at Table 6 tells us that the $M_S = +1$ ($\alpha\alpha$) and $M_S = -1$ ($\beta\beta$) components of the triplet spin function do not separately transform according to one of the irreps of $\bar{C}_{2v}$. Their positive linear combination, $\alpha\alpha + \beta\beta$, exhibits B_1 character, the negative one, $\alpha\alpha - \beta\beta$, belongs to the B_2 irrep. These linear combinations closely resemble expressions [37] for Cartesian tensor components. In accordance with the tensor normalization and phase conventions, we choose $\mathrm{T}_y = \frac{i}{\sqrt{2}}(\alpha\alpha + \beta\beta)$ and $\mathrm{T}_x = \frac{1}{\sqrt{2}}(\beta\beta - \alpha\alpha)$. Among the $M_S = 0$ two-electron functions, the singlet function $\mathrm{S}_0 = \frac{1}{\sqrt{2}}(\alpha\beta - \beta\alpha)$ behaves totally symmetrically (A_1), whereas the $M_S = 0$ component of the triplet, $\mathrm{T}_0 = \mathrm{T}_z = \frac{1}{\sqrt{2}}(\alpha\beta + \beta\alpha)$, transforms according to A_2.

Table 5 Character Table of the $\bar{C}_{2v}$ Double Group

$\bar{C}_{2v}$	$\hat{E}$	$\hat{\bar{E}}$	$(\hat{C}_2(z), \hat{\bar{C}}_2(z))$	$(\hat{\sigma}_{xz}, \hat{\bar{\sigma}}_{xz})$	$(\hat{\sigma}_{yz}, \hat{\bar{\sigma}}_{yz})$
A_1	1	1	1	1	1
A_2	1	1	1	−1	−1
B_1	1	1	−1	1	−1
B_2	1	1	−1	−1	1
$E_{1/2}$	2	−2	0	0	0

Table 6 Action of the Symmetry Operators of $\bar{C}_{2v}$ on Two-Electron Spin Functions

E	$\alpha\alpha =$	$\alpha\alpha$	$\hat{E}$	$\alpha\beta =$	$\alpha\beta$
$\hat{\bar{E}}$	$\alpha\alpha =$	$\alpha\alpha$	$\hat{\bar{E}}$	$\alpha\beta =$	$\alpha\beta$
$\hat{C}_2(z)$	$\alpha\alpha =$	$-\alpha\alpha$	$\hat{C}_2(z)$	$\alpha\beta =$	$\alpha\beta$
$\hat{\sigma}_{xz}$	$\alpha\alpha =$	$\beta\beta$	$\hat{\sigma}_{xz}$	$\alpha\beta =$	$-\beta\alpha$
$\hat{\sigma}_{yz}$	$\alpha\alpha =$	$-\beta\beta$	$\hat{\sigma}_{yz}$	$\alpha\beta =$	$-\beta\alpha$
$\hat{E}$	$\beta\beta =$	$\beta\beta$	$\hat{E}$	$\beta\alpha =$	$\beta\alpha$
$\hat{\bar{E}}$	$\beta\beta =$	$\beta\beta$	$\hat{\bar{E}}$	$\beta\alpha =$	$\beta\alpha$
$\hat{C}_2(z)$	$\beta\beta =$	$-\beta\beta$	$\hat{C}_2(z)$	$\beta\alpha =$	$\beta\alpha$
$\hat{\sigma}_{xz}$	$\beta\beta =$	$\alpha\alpha$	$\hat{\sigma}_{xz}$	$\beta\alpha =$	$-\alpha\beta$
$\hat{\sigma}_{yz}$	$\beta\beta =$	$-\alpha\alpha$	$\hat{\sigma}_{yz}$	$\beta\alpha =$	$-\alpha\beta$

As already mentioned, every spin function corresponding to an even number of electrons transforms according to a representation of the usual molecular point group—the boson representations—whereas the spin functions with an odd number of electrons belong to the representations added in the double group—the fermion representations. Further, for all molecules with a center of inversion, the spin functions transform according to one of the *gerade* (g) (even) representations. The transformation properties of frequently occurring spin functions are shown in Table 7 for some of the most common molecular double groups. A more complete list is found, e.g., in Ref. 69 (p. 569).

Transformation Properties of the Hamiltonian

Tensor Operators

An irreducible tensor operator $\hat{\mathcal{T}}_q^{(k)}$ of rank k is an operator with the property that it transforms under rotations according to

$$\hat{\mathcal{U}}(R)\hat{\mathcal{T}}_q^{(k)}\hat{\mathcal{U}}(R)^{-1} = \sum_{q'} D_{qq'}^{(k)}(R)\hat{\mathcal{T}}_{q'}^{(k)} \qquad [147]$$

Table 7 Transformation Properties of Spin Functions in Frequently Occurring Molecular Double Groups[a]

$S =$	0	$\frac{1}{2}$	1	$\frac{3}{2}$
$\bar{C}_s$	A'	$E_{\frac{1}{2}}$	$A' \oplus 2A''$	$2E_{\frac{1}{2}}$
$\bar{C}_{2v}$	A_1	$E_{\frac{1}{2}}$	$A_2 \oplus B_1 \oplus B_2$	$2E_{\frac{1}{2}}$
$\bar{D}_2(\bar{D}_{2h})$	$A_{(g)}$	$E_{\frac{1}{2}(g)}$	$B_{1(g)} \oplus B_{2(g)} \oplus B_{3(g)}$	$2E_{\frac{1}{2}(g)}$
$\bar{C}_{3v}/\bar{D}_3(\bar{D}_{3d})$	$A_{(g)}$	$E_{\frac{1}{2}(g)}$	$A_{2(g)} \oplus E_{(g)}$	$E_{\frac{1}{2}(g)} \oplus E_{\frac{3}{2}(g)}$
$\bar{C}_{\infty v}(\bar{D}_{\infty h})$	$\Sigma^+_{(g)}$	$E_{\frac{1}{2}(g)}$	$\Sigma^-_{(g)} \oplus \Pi_{(g)}$	$E_{\frac{1}{2}(g)} \oplus E_{\frac{3}{2}(g)}$

[a]The symbol $\oplus$ indicates a direct sum in group theory. The parenthetical gerade subscripts refer to the point groups in parentheses in the first column.

that is, the components of $\hat{\mathscr{T}}^{(k)}$ are transformed among themselves upon a coordinate rotation. The coefficients $D^{(k)}_{qq'}(R)$ are elements of a $2k+1$ dimensional irreducible matrix representation $D^{(k)}$ of the full rotation group in three dimensions, $O^+(3)$ or $SO(3)$. Equivalently, it can be stated that the $2k+1$ components $\hat{\mathscr{T}}^{(k)}_q$ with q ranging from $-k$ to k form a basis for the irreducible representation $D^{(k)}$ of $O^+(3)$. Actually, the coefficients $D^{(k)}_{qq'}(R)$ are the same as those obtained from a rotation operation transferring the spherical harmonic $Y_{kq}(\theta,\phi)$ (see Eq. [25]) to $Y_{kq'}(\theta,\phi)$.

The simplest example is a *scalar* operator $\hat{\mathscr{T}}^{(0)}_0$ with the transformation property

$$\hat{\mathscr{U}}(R)\hat{\mathscr{T}}^{(0)}_0\hat{\mathscr{U}}(R)^{-1} = \hat{T}^{(0)}_0 \qquad [148]$$

This means that scalar operators are invariant with respect to rotations in coordinate or spin space. An example for a scalar operator is the Hamiltonian, i.e., the operator of the energy.

A first-rank tensor operator $\hat{\mathscr{T}}^{(1)}$ is also called a *vector* operator. It has three components, $\hat{\mathscr{T}}^{(1)}_0$ and $\hat{\mathscr{T}}^{(1)}_{\pm1}$. Operators of this type are the angular momentum operators, for instance. Relations between spherical and Cartesian components of first-rank tensor operators are given in Eqs. [36] and [37].

Operating with the components of an arbitrary vector operator $\hat{\mathscr{P}}^{(1)}$ on an eigenfunction $|u^{M_P}_P\rangle$ of the corresponding operators $(\hat{\mathscr{P}}^{(1)})^2$ and $\hat{\mathscr{P}}^{(1)}_0$ yields the following results:

$$\hat{\mathscr{P}}^{(1)}_0|u^{M_P}_P\rangle = \hbar M_P|u^{M_P}_P\rangle \qquad [149]$$

$$\hat{\mathscr{P}}^{(1)}_{\pm1}|u^{M_P}_P\rangle = \mp\hbar\sqrt{\tfrac{1}{2}[P(P+1)-(M_P\pm1)M_P]}\,|u^{M_P\pm1}_P\rangle \qquad [150]$$

$\hat{\mathscr{P}}^{(1)}_0$ conserves the projection M_P of P on the z axis, $\hat{\mathscr{P}}^{(1)}_{+1}$ increases M_P by one unit, and $\hat{\mathscr{P}}^{(1)}_{-1}$ decreases M_P by one unit. The prefactors in Eq. [150] differ slightly from those obtained in Eq. [74] by operating with step operators on $|u^M_J\rangle$ (see the earlier section on step/shift/ladder and tensor operators).

A general Cartesian second-rank tensor operator is represented by a 3×3 matrix.

$$\hat{T}^{(2)}(x,y,z) = \begin{pmatrix} \hat{\mathscr{T}}^{(2)}_{xx} & \hat{\mathscr{T}}^{(2)}_{xy} & \hat{\mathscr{T}}^{(2)}_{xz} \\ \hat{\mathscr{T}}^{(2)}_{yx} & \hat{\mathscr{T}}^{(2)}_{yy} & \hat{\mathscr{T}}^{(2)}_{yz} \\ \hat{\mathscr{T}}^{(2)}_{zx} & \hat{\mathscr{T}}^{(2)}_{zy} & \hat{\mathscr{T}}^{(2)}_{zz} \end{pmatrix} \qquad [151]$$

If the Cartesian representation corresponds to an irreducible second-rank tensor, its nine entries are not independent: The four conditions $\hat{\mathscr{T}}^{(2)}_{ij} = \hat{\mathscr{T}}^{(2)}_{ji}$ and $\sum_i \hat{\mathscr{T}}^{(2)}_{ii} = 0$ with $i,j \in \{x,y,z\}$ reduce the number of independent

components to five. Spherical and Cartesian components of a second-rank irreducible tensor are related by

$$\hat{\mathscr{T}}^{(2)}_{\pm 2} = (\hat{\mathscr{T}}^{(2)}_{xx} - \hat{\mathscr{T}}^{(2)}_{yy} \pm 2i\hat{\mathscr{T}}^{(2)}_{xy})/2 \quad [152]$$

$$\hat{\mathscr{T}}^{(2)}_{\pm 1} = \mp\left(\hat{\mathscr{T}}^{(2)}_{xz} \pm i\hat{\mathscr{T}}^{(2)}_{yz}\right) \quad [153]$$

$$\hat{\mathscr{T}}^{(2)}_{0} = \left(2\hat{\mathscr{T}}^{(2)}_{zz} - \hat{\mathscr{T}}^{(2)}_{xx} - \hat{\mathscr{T}}^{(2)}_{yy}\right)/\sqrt{6} \quad [154]$$

Symmetric, but reducible, second-rank tensor operators ($\hat{\mathscr{T}}^{(2)}_{ij} = \hat{\mathscr{T}}^{(2)}_{ji}$ and $\sum_i \hat{\mathscr{T}}^{(2)}_{ii} \neq 0$) with six independent components are widespread: the polarizability and the hyperfine coupling tensors are well-known representatives.

Compound Tensor Operators

In a tensor product $\{\hat{\mathscr{P}}^{(k)} \otimes \hat{\mathscr{Q}}^{(j)}\}$, each component of $\hat{\mathscr{P}}^{(k)}$ is to be multiplied with each component of $\hat{\mathscr{Q}}^{(j)}$. The resulting $2(k+1)\cdot 2(j+1)$-dimensional tensor is in general not irreducible. Reduction yields irreducible tensor operators with ranks ranging from $k+j$, $|k+j-1|$, $\cdots$, $|k-j|$.

As an example, consider the product of two arbitrary first-rank tensor operators $\hat{\mathscr{P}}^{(1)}$ and $\hat{\mathscr{Q}}^{(1)}$. It is nine-dimensional and can be reduced to a sum of compound irreducible tensor operators of ranks 2, 1, and 0, respectively. Operators of this type play a role in spin–spin coupling Hamiltonians. In terms of spherical and Cartesian components of $\hat{\mathscr{P}}$ and $\hat{\mathscr{Q}}$, the resulting irreducible tensors are given in Tables 8 and 9, respectively.[70]

We are mainly interested in compound tensor operators of rank zero (i.e., scalar operators such as the Hamiltonian). To form a scalar from two tensor operators $\hat{\mathscr{P}}^{(k)}$ and $\hat{\mathscr{Q}}^{(j)}$, their ranks k and j have to be equal. Further, the $+q$ component of $\hat{\mathscr{P}}^{(k)}$ has to be combined with the $-q$ component of $\hat{\mathscr{Q}}^{(k)}$ and

Table 8 Irreducible Spherical Compound Tensor Operators Resulting from a Product of Two First-Rank Tensor Operators $\hat{\mathscr{P}}^{(1)}$ and $\hat{\mathscr{Q}}^{(1)}$

Compound Tensor	m	Spherical Component
$\{\hat{\mathscr{P}}^{(1)} \otimes \hat{\mathscr{Q}}^{(1)}\}^{(2)}_m$	+2	$\hat{\mathscr{P}}^{(1)}_{+1}\hat{\mathscr{Q}}^{(1)}_{+1}$
	+1	$(\hat{\mathscr{P}}^{(1)}_{+1}\hat{\mathscr{Q}}^{(1)}_{0} + \hat{\mathscr{P}}^{(1)}_{0}\hat{\mathscr{Q}}^{(1)}_{+1})/\sqrt{2}$
	0	$(\hat{\mathscr{P}}^{(1)}_{-1}\hat{\mathscr{Q}}^{(1)}_{+1} + 2\hat{\mathscr{P}}^{(1)}_{0}\hat{\mathscr{Q}}^{(1)}_{0} + \hat{\mathscr{P}}^{(1)}_{+1}\hat{\mathscr{Q}}^{(1)}_{-1})/\sqrt{6}$
	−1	$(\hat{\mathscr{P}}^{(1)}_{-1}\hat{\mathscr{Q}}^{(1)}_{0} + \hat{\mathscr{P}}^{(1)}_{0}\hat{\mathscr{Q}}^{(1)}_{-1})/\sqrt{2}$
	−2	$\hat{\mathscr{P}}^{(1)}_{-1}\hat{\mathscr{Q}}^{(1)}_{-1}$
$\{\hat{\mathscr{P}}^{(1)} \otimes \hat{\mathscr{Q}}^{(1)}\}^{(1)}_m$	+1	$(\hat{\mathscr{P}}^{(1)}_{+1}\hat{\mathscr{Q}}^{(1)}_{0} - \hat{\mathscr{P}}^{(1)}_{0}\hat{\mathscr{Q}}^{(1)}_{+1})/\sqrt{2}$
	0	$(\hat{\mathscr{P}}^{(1)}_{+1}\hat{\mathscr{Q}}^{(1)}_{-1} - \hat{\mathscr{P}}^{(1)}_{-1}\hat{\mathscr{Q}}^{(1)}_{+1})/\sqrt{2}$
	−1	$(\hat{\mathscr{P}}^{(1)}_{0}\hat{\mathscr{Q}}^{(1)}_{-1} - \hat{\mathscr{P}}^{(1)}_{-1}\hat{\mathscr{Q}}^{(1)}_{0})/\sqrt{2}$
$\{\hat{\mathscr{P}}^{(1)} \otimes \hat{\mathscr{Q}}^{(1)}\}^{(0)}_m$	0	$(\hat{\mathscr{P}}^{(1)}_{-1}\hat{\mathscr{Q}}^{(1)}_{+1} - \hat{\mathscr{P}}^{(1)}_{0}\hat{\mathscr{Q}}^{(1)}_{0} + \hat{\mathscr{P}}^{(1)}_{+1}\hat{\mathscr{Q}}^{(1)}_{-1})/\sqrt{3}$

Table 9 Irreducible Cartesian Compound Tensor Operators Resulting from a Product of Two First-Rank Tensor Operators $\hat{\mathscr{P}}^{(1)}$ and $\hat{\mathscr{Q}}^{(1)}$

Compound Tensor	m	Cartesian Component
$\{\hat{\mathscr{P}}^{(1)} \otimes \hat{\mathscr{Q}}^{(1)}\}^{(2)}_m$	+2	$(\hat{\mathscr{P}}^{(1)}_x \hat{\mathscr{Q}}^{(1)}_x - \hat{\mathscr{P}}^{(1)}_y \hat{\mathscr{Q}}^{(1)}_y + i\hat{\mathscr{P}}^{(1)}_x \hat{\mathscr{Q}}^{(1)}_y + i\hat{\mathscr{P}}^{(1)}_y \hat{\mathscr{Q}}^{(1)}_x)/2$
	+1	$-(\hat{\mathscr{P}}^{(1)}_x \hat{\mathscr{Q}}^{(1)}_z + \hat{\mathscr{P}}^{(1)}_z \hat{\mathscr{Q}}^{(1)}_x + i\hat{\mathscr{P}}^{(1)}_y \hat{\mathscr{Q}}^{(1)}_z + i\hat{\mathscr{P}}^{(1)}_z \hat{\mathscr{Q}}^{(1)}_y)/2$
	0	$(2\hat{\mathscr{P}}^{(1)}_z \hat{\mathscr{Q}}^{(1)}_z - \hat{\mathscr{P}}^{(1)}_x \hat{\mathscr{Q}}^{(1)}_x - \hat{\mathscr{P}}^{(1)}_y \hat{\mathscr{Q}}^{(1)}_y)/\sqrt{6}$
	−1	$(\hat{\mathscr{P}}^{(1)}_x \hat{\mathscr{Q}}^{(1)}_z + \hat{\mathscr{P}}^{(1)}_z \hat{\mathscr{Q}}^{(1)}_x - i\hat{\mathscr{P}}^{(1)}_y \hat{\mathscr{Q}}^{(1)}_z - i\hat{\mathscr{P}}^{(1)}_z \hat{\mathscr{Q}}^{(1)}_y)/2$
	−2	$(\hat{\mathscr{P}}^{(1)}_x \hat{\mathscr{Q}}^{(1)}_x - \hat{\mathscr{P}}^{(1)}_y \hat{\mathscr{Q}}^{(1)}_y - i\hat{\mathscr{P}}^{(1)}_x \hat{\mathscr{Q}}^{(1)}_y - i\hat{\mathscr{P}}^{(1)}_y \hat{\mathscr{Q}}^{(1)}_x)/2$
$\{\hat{\mathscr{P}}^{(1)} \otimes \hat{\mathscr{Q}}^{(1)}\}^{(1)}_m$	+1	$-(\hat{\mathscr{P}}^{(1)}_x \hat{\mathscr{Q}}^{(1)}_z - \hat{\mathscr{P}}^{(1)}_z \hat{\mathscr{Q}}^{(1)}_x + i\hat{\mathscr{P}}^{(1)}_y \hat{\mathscr{Q}}^{(1)}_z - i\hat{\mathscr{P}}^{(1)}_z \hat{\mathscr{Q}}^{(1)}_y)/2$
	0	$i(\hat{\mathscr{P}}^{(1)}_x \hat{\mathscr{Q}}^{(1)}_y - \hat{\mathscr{P}}^{(1)}_y \hat{\mathscr{Q}}^{(1)}_x)/\sqrt{2}$
	−1	$(\hat{\mathscr{P}}^{(1)}_z \hat{\mathscr{Q}}^{(1)}_x - \hat{\mathscr{P}}^{(1)}_x \hat{\mathscr{Q}}^{(1)}_z - i\hat{\mathscr{P}}^{(1)}_z \hat{\mathscr{Q}}^{(1)}_y + i\hat{\mathscr{P}}^{(1)}_y \hat{\mathscr{Q}}^{(1)}_z)/2$
$\{\hat{\mathscr{P}}^{(1)} \otimes \hat{\mathscr{Q}}^{(1)}\}^{(0)}_m$	0	$-(\hat{\mathscr{P}}^{(1)}_x \hat{\mathscr{Q}}^{(1)}_x + \hat{\mathscr{P}}^{(1)}_y \hat{\mathscr{Q}}^{(1)}_y + \hat{\mathscr{P}}^{(1)}_z \hat{\mathscr{Q}}^{(1)}_z)/\sqrt{3}$

vice versa. Two notations are found in the literature for these tensor products differing only in normalization constants and phase factors. The properly normalized product of two tensor operators forming a scalar operator reads[71]

$$\{\hat{\mathscr{P}}^{(k)} \otimes \hat{\mathscr{Q}}^{(k)}\}^{(0)}_0 = (2k+1)^{-1/2} \sum_{q=-k}^{+k} (-1)^{k-q} \hat{\mathscr{P}}^{(k)}_{-q} \hat{\mathscr{Q}}^{(k)}_q \qquad [155]$$

Instead, we utilize the simplified expression that differs from Eq. [155] by a factor of $\sqrt{2k+1}$:

$$(\hat{\mathscr{P}}^{(k)} \cdot \hat{\mathscr{Q}}^{(k)}) = \sum_{q=-1}^{+k} (-1)^{k-q} \hat{\mathscr{P}}^{(k)}_{-q} \hat{\mathscr{Q}}^{(k)}_q \qquad [156]$$

Tensor Properties of Magnetic Interaction Terms

The operators $\hat{\mathscr{H}}_{\mathrm{SO}}$ and $\hat{\mathscr{H}}_{\mathrm{SS}}$ are compound tensor operators of rank zero (scalars) composed of vector (first-rank tensor) operators and matrix (second-rank tensor) operators. We will make use of this tensorial structure when it comes to selection rules for the magnetic interaction Hamiltonians and symmetry relations between their matrix elements. Similar considerations apply to the molecular rotation $\hat{\mathscr{H}}_{\mathrm{rot}}$ and hyperfine splitting interaction Hamiltonians $\hat{\mathscr{H}}_{\mathrm{hfs}}$.

Spin–Orbit Coupling For the derivation of selection rules, it is sufficient to employ a simplified Hamiltonian. To this end, we rewrite each term in the microscopic spin–orbit Hamiltonians in form of a scalar product between an appropriately chosen spatial angular momentum $\vec{\hat{\mathscr{L}}}$ and a spin angular momentum $\vec{\hat{\mathcal{S}}}$

$$\hat{\mathscr{H}}_{\mathrm{SO}} = A_{\mathrm{SO}}(r) \vec{\hat{\mathscr{L}}} \cdot \vec{\hat{\mathcal{S}}} \qquad [157]$$

The operator [157] is a phenomenological spin–orbit operator. In addition to being useful for symmetry considerations, Eq. [157] can be utilized for setting up a connection between theoretically and experimentally determined fine-structure splittings via the so-called spin–orbit parameter A_{SO} (see the later section on first-order spin–orbit splitting). In terms of its tensor components, the phenomenological spin–orbit Hamiltonian reads

$$\hat{\mathscr{H}}_{\mathrm{SO}} = A_{\mathrm{SO}}(r)\vec{\hat{\mathscr{L}}} \cdot \vec{\hat{\mathcal{S}}} \qquad [158]$$

$$= A_{\mathrm{SO}}(r)(\hat{\mathscr{L}}_0\hat{\mathcal{S}}_0 - \hat{\mathscr{L}}_{+1}\hat{\mathcal{S}}_{-1} - \hat{\mathscr{L}}_{-1}\hat{\mathcal{S}}_{+1}) \qquad [159]$$

$$= A_{\mathrm{SO}}(r)(\hat{\mathscr{L}}_x\hat{\mathcal{S}}_x + \hat{\mathscr{L}}_y\hat{\mathcal{S}}_y + \hat{\mathscr{L}}_z\hat{\mathcal{S}}_z) \qquad [160]$$

The phenomenological spin–orbit Hamiltonian ought not to be used for computing spin–orbit matrix elements, though. An example for a failure of such a procedure will be discussed in detail in the later subsection on a word of caution.

Spin–Spin Coupling Although we focus on spin–orbit coupling (SOC), we need to consider the tensorial structure of electronic spin–spin coupling: Second-order SOC mimics perfectly first-order spin–spin coupling and vice versa, so that they cannot be told apart (see the later section on second-order spin–orbit splitting).

The dipolar spin–spin coupling operators are scalar operators of the form $\hat{\mathscr{P}}^{(1)} \cdot \hat{\mathscr{T}}^{(2)} \cdot \hat{\mathscr{Q}}^{(1)}$. The tensorial structure of $\hat{\mathscr{H}}_{\mathrm{SS}}$ becomes apparent if we write the Breit–Pauli spin–spin coupling operator as

$$\hat{\mathscr{H}}_{\mathrm{SS}}^{\mathrm{BP}} = \frac{e^2\hbar^2}{2m_e^2c^2}\sum_{i\neq j}\left\{-\frac{8\pi}{3}\delta(\vec{r}_{ij})\vec{\hat{s}}_i\vec{\hat{s}}_j + \frac{\vec{\hat{s}}_i\vec{\hat{s}}_j}{\hat{r}_{ij}^3} - \frac{\vec{\hat{s}}_i(3\vec{r}_{ij}\otimes\vec{r}_{ij})^{(2)}\vec{\hat{s}}_j)}{\hat{r}_{ij}^5}\right\} \qquad [161]$$

$$= \frac{e^2\hbar^2}{2m_e^2c^2}\sum_{i\neq j}\left\{-\frac{8\pi}{3}\delta(\vec{r}_{ij})\vec{\hat{s}}_i\vec{\hat{s}}_j + \left(\vec{\hat{s}}_i\hat{\mathscr{D}}_{ij}^{(2)}\vec{\hat{s}}_j\right)\right\} \qquad [162]$$

The first term is a tensor of rank zero involving only spin variables. It does not contribute to the multiplet splitting of an electronic state but yields only a (small) overall shift of the energy and is, henceforth, neglected. The operator $\hat{\mathscr{D}}_{ij}^{(2)}$ is a traceless (irreducible) second-rank tensor operator, the form of which in Cartesian components is

$$\hat{\mathscr{D}}_{ij}^{(2)} = \frac{1}{\hat{r}_{ij}^5}\begin{pmatrix} \hat{r}_{ij}^2 - 3\hat{x}_{ij}^2 & -3\hat{x}_{ij}\hat{y}_{ij} & -3\hat{x}_{ij}\hat{z}_{ij} \\ -3\hat{x}_{ij}\hat{y}_{ij} & \hat{r}_{ij}^2 - 3\hat{y}_{ij}^2 & -3\hat{y}_{ij}\hat{z}_{ij} \\ -3\hat{x}_{ij}\hat{z}_{ij} & -3\hat{y}_{ij}\hat{z}_{ij} & \hat{r}_{ij}^2 - 3\hat{z}_{ij}^2 \end{pmatrix} \qquad [163]$$

For the definition of a phenomenological electronic spin–spin operator, one makes use of this tensorial structure

$$\hat{\mathscr{H}}_{\mathrm{SS}} = \left\{\hat{\mathcal{S}}^{(1)}\hat{\mathscr{D}}^{(2)}_{\mathrm{SS}}\hat{\mathcal{S}}^{(1)}\right\}^{(0)} \qquad [164]$$

An overall coupling to a scalar poses again certain conditions on the combination of the tensor components. All products $\hat{\mathscr{P}}^{(1)}_p\hat{\mathscr{T}}^{(2)}_t\hat{\mathscr{Q}}^{(1)}_q$, which fulfill the condition that $p + t + q = 0$, may contribute to the compound scalar operator. Actual prefactors of the terms depend on the order in which the three tensor operators are coupled. In general there is no unique way to combine three angular momenta j_1, j_2, and j_3 to a total angular momentum J. The $6j$ symbols allow a conversion between two possible coupling schemes involving direct products, for example, between $\{\{j_1 \otimes j_2\} \otimes j_3\} = J$ and $\{j_1 \otimes \{j_2 \otimes j_3\}\} = J$.[72–74] In the electronic spin–spin Hamiltonian, the two spin operators are coupled first to form a second-rank tensor operator, which is then combined with $\hat{\mathscr{D}}^{(2)}_{ij}$ to form a scalar, that is,

$$\{\{\vec{\hat{s}}_i \otimes \vec{\hat{s}}_j\}^{(2)} \otimes \hat{\mathscr{D}}^{(2)}_{ij}\}^{(0)} \qquad [165]$$

Matrix Elements

The Wigner–Eckart Theorem

The main reason for working with irreducible tensor operators stems from an important theorem, known as the Wigner–Eckart Theorem (WET)[75,76] for matrix elements of tensor operators:

$$\langle \alpha' j' m' | \hat{\mathscr{T}}^{(k)}_q | \alpha j m \rangle = \underbrace{(-1)^{j'-m'}\begin{pmatrix} j' & k & j \\ -m' & q & m \end{pmatrix}}_{\text{geometrical part: } 3j \text{ symbol}} \underbrace{\langle \alpha' j' \| \hat{\mathscr{T}}^{(k)} \| \alpha j \rangle}_{\text{physical part: reduced matrix element}} \qquad [166]$$

Here, j and j' represent angular momentum quantum numbers of the initial and final wave functions related to the tensor $\hat{\mathscr{T}}^{(k)}$, and m and m' denote the corresponding magnetic quantum numbers. Note that α and α' do not mean spin states in this context but stand for all other quantum numbers.

From the WET, Eq. [166], it is obvious that the reduced matrix element (RME) depends on the specific wave functions and the operator, whereas it is independent of magnetic quantum numbers m. The $3j$ symbol depends only on rotational symmetry properties. It is related to the corresponding vector

addition coefficient or *Clebsch–Gordan* (CG) coefficient by[77]

$$\begin{pmatrix} j' & k & j \\ -m' & q & m \end{pmatrix} = \frac{(-1)^{j'-k-m}}{\sqrt{2j+1}} \langle j'km'q|j'kjm\rangle \qquad [167]$$

The general selection rules emerging from the properties of the $3j$ symbols are

$$m = m' + q \qquad \triangle(jkj') \qquad [168]$$

where the triangle condition $\triangle(jkj')$ is fulfilled if $j + j' \geq k \geq |j - j'|$.

Specifically, for tensor operators of rank 0, 1, or 2, one obtains:

$$\langle \alpha' j' m'|\hat{\mathscr{T}}^{(0)}_0|\alpha jm\rangle = 0 \quad \text{unless} \quad \begin{cases} \Delta j = 0 \\ \Delta m = 0 \end{cases} \qquad [169]$$

$$\langle \alpha' j' m'|\hat{\mathscr{T}}^{(1)}_q|\alpha jm\rangle = 0 \quad \text{unless} \quad \begin{cases} \Delta j = 0, \pm 1 \\ \Delta m = 0, \pm 1 \\ j + j' \geq 1 \end{cases} \qquad [170]$$

$$\langle \alpha' j' m'|\hat{\mathscr{T}}^{(2)}_q|\alpha jm\rangle = 0 \quad \text{unless} \quad \begin{cases} \Delta j = 0, \pm 1, \pm 2 \\ \Delta m = 0, \pm 1, \pm 2 \\ j + j' \geq 2 \end{cases} \qquad [171]$$

The condition $j + j' \geq 1$ for a matrix element of a first rank tensor operator implies, e.g., that there is no first-order SOC of singlet wave functions. Two doublet spin wave functions may interact via SOC, but the selection rule $j + j' \geq 2$ for $\hat{\mathscr{T}}^{(2)}_q$ (Eq. [171]) tells us that electronic spin–spin interaction does not contribute to their fine-structure splitting in first order.

Besides providing us with selection rules, the WET can be employed to considerably reduce the computational effort: if a single matrix element $\langle \alpha' j' m'|\hat{\mathscr{T}}^{(k)}_q|\alpha jm\rangle$ is known, all possible $(2j' + 1)(2k + 1)(2j + 1)$ matrix elements $\langle \alpha' j' m'''|\hat{\mathscr{T}}^{(k)}_q|\alpha jm''\rangle$ can be calculated with the aid of $3j$ symbols:

$$\langle \alpha' j' m'''|\hat{\mathscr{T}}^{(k)}_{q'}|\alpha jm''\rangle = (-1)^{m'''-m'} \frac{\begin{pmatrix} j' & k & j \\ -m''' & q' & m'' \end{pmatrix}}{\begin{pmatrix} j' & k & j \\ -m' & q & m \end{pmatrix}} \cdot \langle \alpha' j' m'|\hat{\mathscr{T}}^{(k)}_q|\alpha jm\rangle \qquad [172]$$

Equation [172] or related expressions (Table 10) are applied extensively when evaluating the spin part of spin–orbit matrix elements, for configuration interaction (CI) wave functions. The latter are usually provided for a single M_S component only.

Table 10 Spin-Coupling Coefficients

ΔS	M'_S	M_S	$(-1)^{S'-M'_S}\begin{pmatrix} S' & 1 & S \\ -M'_S & q' & M_S \end{pmatrix}\begin{pmatrix} S' & 1 & S \\ -S' & q & S \end{pmatrix}^{-1}$
$S' = S > 0$	M	M	$M \cdot S^{-1}$
	$M \pm 1$	M	$\mp((S \pm M + 1)(S \mp M)/2)^{1/2} \cdot S^{-1}$
$S' = S + 1$	M	M	$((S - M + 1)(S + M + 1)/2)^{1/2} \cdot ((2S + 1)(S + 1)/2)^{-1/2}$
	$M \pm 1$	M	$((S \pm M + 1)(S \pm M + 2)/4)^{1/2} \cdot ((2S + 1)(S + 1)/2)^{-1/2}$

Numerical values for the $3j$ symbols can be looked up in tables.[72,77] Alternatively, they can be computed using analytic formulas revised by Roothaan.[78,79] Here, only some of their symmetry properties shall be mentioned.

$$\text{even permutations:} \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = \begin{pmatrix} j_3 & j_1 & j_2 \\ m_3 & m_1 & m_2 \end{pmatrix} \quad [173]$$

$$\text{odd permutations:} = (-1)^{j_1+j_2+j_3}\begin{pmatrix} j_2 & j_1 & j_3 \\ m_2 & m_1 & m_3 \end{pmatrix} \quad [174]$$

$$\text{interchange of rows:} = \begin{pmatrix} m_1 & m_2 & m_3 \\ j_1 & j_2 & j_3 \end{pmatrix} \quad [175]$$

$$m \text{ changed to } -m\text{:} = (-1)^{j_1+j_2+j_3}\begin{pmatrix} j_1 & j_2 & j_3 \\ -m_1 & -m_2 & -m_3 \end{pmatrix} \quad [176]$$

For the special case of a first-rank tensor operator and $m' = j'$ and $m = j$, analytical formulas for the symmetry related factors in Eq. [172] have been worked out by McWeeny[70] and by Cooper and Musher.[80] Note, however, that the formulas in both publications contain typos concerning a sign or a square root.

The WET can in principle be applied to both the spatial and spin parts of a spin–orbit coupling matrix element. In molecular applications, however, the question arises how to use the WET, since L is not a good quantum number.[c] There is a way out: We can work with Cartesian spatial functions and spherical spin functions and apply relations [36] and [37] for transforming back and forth!

[c] In quantum theory and spectroscopy, a "good" quantum number is one that is independent of the theoretical model.

Tips and Tricks

So far we know the selection rules for spin–orbit coupling. Further, given a reduced matrix element (RME), we are able to calculate the matrix elements (MEs) of all multiplet components by means of the WET. What remains to be done is thus to compute RMEs. Technical procedures how this can be achieved for CI wave functions are presented in the later section on Computational Aspects. Regarding symmetry, often a complication arises in this step: CI wave functions are usually determined only for a single spin component, mostly $M_S = S$. The M_S quantum numbers determine the component of the spin tensor operator for which the spin matrix element $\langle S'|\hat{\mathcal{S}}_q|S\rangle$ does not vanish. The component q does not always match, however, the selection rules dictated by the spatial part of the ME.

Given a molecule that possesses C_{2v} symmetry, let us try to figure out how to calculate $\langle {}^3A_2|\hat{\mathscr{H}}_{\mathrm{SO}}|{}^3B_1\rangle$ from wave functions with $M_S = 1$. The coupling of an A_2 and a B_1 state requires a spatial angular momentum operator of B_2 symmetry. From Table 11, we read that this is just the x component of $\hat{\vec{\mathscr{L}}}$. A direct computation of $\langle {}^3A_2, M_S = 1|\hat{\mathscr{L}}_x\hat{\mathcal{S}}_x|{}^3B_1, M_S = 1\rangle$ yields zero because the integral of $\hat{\mathcal{S}}_x$ with two $M_S = 1$ functions is zero. Instead, one calculates $\mathrm{ME} = \langle {}^3A_2, M_S = 1|\hat{\mathscr{L}}_x\hat{\mathcal{S}}_0|{}^3B_1, M_S = 1\rangle$, where $\hat{\mathscr{L}}_x$ stands symbolically for the spatial part of the microscopic spin–orbit Hamiltonian with x symmetry and $\hat{S}_0$ correspondingly for the zero-component of the spin tensor. This is the only nonzero matrix element for the given wave functions.

The RME (only spin) is then given by

$$\mathrm{RME} = \frac{\langle {}^3A_2, M_S = 1|\hat{\mathscr{L}}_x\hat{\mathcal{S}}_0|{}^3B_1, M_S = 1\rangle}{\begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix}} \qquad [177]$$

Table 11 Irreps of Singlet (S) and Triplet (T) Spin Functions, the Angular Momentum Operators ($\hat{\vec{\mathscr{L}}}$ and $\hat{\vec{\mathcal{S}}}$), an Irreducible Second-Rank Tensor Operator $\hat{\mathscr{D}}$, and the Position Operators $\hat{\mathscr{X}}, \hat{\mathscr{Y}}, \hat{\mathscr{Z}}$ in C_{2v} Symmetry

Function/Operator	Irreducible Representation
$\mathrm{Singlet}_{M_S=0}(\mathrm{S}_0)$	A_1
$\mathrm{Triplet}_{M_S=0}(\mathrm{T}_0)$	A_2
$\mathrm{Triplet}_{M_S=\pm1}(\mathrm{T}_{\pm1})$	$\{B_1, B_2\}$
$\hat{\mathscr{L}}_x, \hat{\mathcal{S}}_x, \mathrm{Triplet\ T}_x$	B_2
$\hat{\mathscr{L}}_y, \hat{\mathcal{S}}_y, \mathrm{Triplet\ T}_y$	B_1
$\hat{\mathscr{L}}_z, \hat{\mathcal{S}}_z, \mathrm{Triplet\ T}_z$	A_2
$\hat{\mathscr{D}}_{2z^2-x^2-y^2}$	A_1
$\hat{\mathscr{D}}_{x^2-y^2}$	A_1
$\hat{\mathscr{D}}_{xy}$	A_2
$\hat{\mathscr{D}}_{xz}$	B_2
$\hat{\mathscr{D}}_{yz}$	B_1
$\hat{\mathscr{X}}$	B_1
$\hat{\mathscr{Y}}$	B_2
$\hat{\mathscr{Z}}$	A_1

with

$$\begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix} = \frac{1}{\sqrt{6}}$$

Since the spatial wave functions are coupled by $\hat{\mathscr{L}}_x$, a spin operator of x symmetry has to be used also. The operator $\hat{\mathcal{S}}_x$ contributes to both $\hat{\mathcal{S}}_{-1}$ and $\hat{\mathcal{S}}_{+1}$. Thus, nonzero matrix elements are expected for $\langle M_S = 1|\hat{\mathcal{S}}_x|M_S = 0\rangle$, $\langle M_S = 0|\hat{\mathcal{S}}_x|M_S = 1\rangle$, $\langle M_S = 0|\hat{\mathcal{S}}_x|M_S = -1\rangle$, and $\langle M_S = -1|\hat{\mathcal{S}}_x|M_S = 0\rangle$. The $3j$ symbols for the coupling of these spin functions by $\hat{\mathcal{S}}_{-1}$ and $\hat{\mathcal{S}}_{+1}$ are easily evaluated as

$$\begin{aligned} -\begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & 0 \end{pmatrix} &= \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \end{pmatrix} \\ = -\begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 1 \end{pmatrix} &= \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{pmatrix} = \frac{1}{\sqrt{6}} \end{aligned} \qquad [178]$$

Applying relations [37] and [172], the following interaction matrix for $\langle {}^3A_2|\hat{\mathscr{L}}_x\hat{\mathcal{S}}_x|{}^3B_1\rangle$ is obtained:

$M_S =$	1	0	−1
$M_S = 1$	0	$+\text{RME}/\sqrt{2}$	0
0	$+\text{RME}/\sqrt{2}$	0	$+\text{RME}/\sqrt{2}$
−1	0	$+\text{RME}/\sqrt{2}$	0

Note, that it is not possible to compute the nonvanishing MEs of a triplet–triplet coupling by using $M_S = 0$ wave functions, because the spin part of $\langle \Psi, M_S = 0|\vec{\hat{\mathscr{L}}} \cdot \vec{\hat{\mathcal{S}}}|\Psi', M_S = 0\rangle$ is zero and cannot serve for determining the RME.

A Word of Caution

Phenomenological operators should be utilized with great care. Do not oversimplify!

As an example, let us look at the spin–orbit coupling between a ${}^3\Sigma^-$ and a ${}^1\Sigma^+$ state of a linear molecule. This example is of some importance for the understanding of spin-forbidden transitions in the O_2 molecule which exhibits a ${}^3\Sigma_g^-$ ground state and a low-lying excited ${}^1\Sigma_g^+$ state. The coupling is symmetry allowed, that is, the matrix element $\langle {}^3\Sigma_0^-|\mathscr{H}_{\text{SO}z}|{}^1\Sigma_0^+\rangle$ is different from zero. To see this, we may employ the phenomenological operator $A_{\text{SO}}\vec{\hat{\mathscr{L}}}\vec{\hat{\mathcal{S}}}$ (Eq. [157]) in the symmetry analysis (the spin–orbit parameter A_{SO} was introduced in the earlier section on tensor properties of magnetic interaction terms):

1. The $\hat{\mathscr{L}}_z$ transforms like $\Sigma^-_{(g)}$ in the symmetry group of a linear molecule and thus allows an interaction between Σ^- and Σ^+ states.
2. The $M_S = 0$ component of a triplet has Σ^- symmetry, $\hat{\mathcal{S}}_z$ transforms like $\Sigma^-_{(g)}$, and singlets are totally symmetric, resulting again in a symmetry-allowed interaction.

However, if $\langle {}^3\Sigma_0^- | \mathscr{H}_{\mathrm{SO}z} | {}^1\Sigma_0^+ \rangle$ is evaluated using the phenomenological operator ([Eq. 157]) explicitly, we obtain $A_{\mathrm{SO}} \langle \Sigma^- | \hat{\mathscr{L}}_z | \Sigma^+ \rangle \langle \mathrm{T}_0 | \hat{\mathcal{S}}_z | \mathrm{S}_0 \rangle = 0$:

1. $\mathscr{L}_z$ operating on a Σ state gives 0.
2. The same holds true for the action of $\mathcal{S}_0$ on a state with $M_S = 0$.

The resolution of this discrepancy is closely related to another question: How is an operator such as $\vec{\hat{\mathcal{S}}}$, when combined with $\vec{\hat{\mathscr{L}}}$, capable of coupling electronic states of different multiplicities while, according to Eqs. [149] and [150], $\vec{\hat{\mathcal{S}}}$ as a first rank tensor operator is only able to change the M_S quantum number of a state, but not its S value.

To understand these seemingly opposite facts, we have to leave the global $\vec{\hat{\mathscr{L}}} \cdot \vec{\hat{\mathcal{S}}}$ expression and rather write the spin–orbit Hamiltonian as a sum of one-particle operators

$$\vec{\hat{\mathscr{L}}} \cdot \vec{\hat{\mathcal{S}}} = \vec{\hat{\ell}}(1) \vec{\hat{s}}(1) + \vec{\hat{\ell}}(2) \vec{\hat{s}}(2) \qquad [179]$$

As we shall see, it is the $M_S = 0$ level of the triplet state that couples to the singlet state in this case, and the coupling is brought about by the z component of the spin–orbit coupling operator. The wave functions of the ${}^3\Sigma_g^-$ and ${}^1\Sigma_g^+$ states of O_2 can be written as

$$\langle {}^1\Sigma^+ | = \frac{1}{2} \langle [\pi_x(1)\pi_x(2) + \pi_y(1)\pi_y(2))(\alpha(1)\beta(2) - \beta(1)\alpha(2)] | \qquad [180]$$

$$| {}^3\Sigma^- \rangle = \frac{1}{2} | [\pi_x(1)\pi_y(2) - \pi_y(1)\pi_x(2))(\alpha(1)\beta(2) + \beta(1)\alpha(2)] \rangle \qquad [181]$$

or in a short-hand notation

$$\langle {}^1\Sigma^+ | = \frac{1}{2} \langle \pi_x \bar{\pi}_x - \bar{\pi}_x \pi_x + \pi_y \bar{\pi}_y - \bar{\pi}_y \pi_y | \qquad [182]$$

$$| {}^3\Sigma^- \rangle = \frac{1}{2} | \pi_x \bar{\pi}_y + \bar{\pi}_x \pi_y - \pi_y \bar{\pi}_x - \bar{\pi}_y \pi_x \rangle \qquad [183]$$

where unbarred and barred entities denote α and β electrons, respectively, and the particle label has been suppressed. The (one-electron part of the) spin–orbit matrix element $\langle {}^3\Sigma_0^- | \mathscr{H}_{\mathrm{SO}z} | {}^1\Sigma_0^+ \rangle$ is given by

$$\frac{1}{4} \langle \pi_x \bar{\pi}_x - \bar{\pi}_x \pi_x + \pi_y \bar{\pi}_y - \bar{\pi}_y \pi_y | [\ell_0(1) s_0(1) + \ell_0(2) s_0(2)] | \pi_x \bar{\pi}_y + \bar{\pi}_x \pi_y - \pi_y \bar{\pi}_x - \bar{\pi}_y \pi_x \rangle \qquad [184]$$

At this stage, all the way down at the integral level, we can finally exploit the fact that $\vec{\hat{\ell}}$ and $\vec{\hat{s}}$ act on different coordinates: we can split up the integrals into space- and spin-dependent parts. In the above matrix element (expression [184]), only eight terms are different from zero because $\hat{s}_z$ does not change α into β spin or vice versa. The remaining integrals are

$$\begin{aligned}
&\langle \pi_x \bar{\pi}_x | \hat{\ell}_z(2)\hat{s}_z(2) | \pi_x \bar{\pi}_y \rangle - \langle \bar{\pi}_x \pi_x | \hat{\ell}_z(2)\hat{s}_z(2) | \bar{\pi}_x \pi_y \rangle - \langle \pi_y \bar{\pi}_y | \hat{\ell}_z(2)\hat{s}_z(2) | \pi_y \bar{\pi}_x \rangle \\
&+ \langle \bar{\pi}_y \pi_y | \hat{\ell}_z(2)\hat{s}_z(2) | \bar{\pi}_y \pi_x \rangle - \langle \pi_x \bar{\pi}_x | \hat{\ell}_z(1)\hat{s}_z(1) | \pi_y \bar{\pi}_x \rangle + \langle \bar{\pi}_x \pi_x | \hat{\ell}_z(1)\hat{s}_z(1) | \bar{\pi}_y \pi_x \rangle \\
&+ \langle \pi_y \bar{\pi}_y | \hat{\ell}_z(1)\hat{s}_z(1) | \pi_x \bar{\pi}_y \rangle - \langle \bar{\pi}_y \pi_y | \hat{\ell}_z(1)\hat{s}_z(1) | \bar{\pi}_x \pi_y \rangle
\end{aligned} \quad [185]$$

This expression can be reduced further by integrating over the spin and exploiting the fact that electrons 1 and 2 are indistinguishable. Remembering that $\langle \alpha | \hat{s}_z | \alpha \rangle = \frac{\hbar}{2}$ and $\langle \beta | \hat{s}_z | \beta \rangle = -\frac{\hbar}{2}$, one obtains

$$\begin{aligned}
&-\frac{\hbar}{2}\langle \pi_x | \hat{\ell}_z | \pi_y \rangle - \frac{\hbar}{2}\langle \pi_x | \hat{\ell}_z | \pi_y \rangle + \frac{\hbar}{2}\langle \pi_y | \hat{\ell}_z | \pi_x \rangle + \frac{\hbar}{2}\langle \pi_y | \hat{\ell}_z | \pi_x \rangle \\
&-\frac{\hbar}{2}\langle \pi_x | \hat{\ell}_z | \pi_y \rangle - \frac{\hbar}{2}\langle \pi_x | \hat{\ell}_z | \pi_y \rangle + \frac{\hbar}{2}\langle \pi_y | \hat{\ell}_z | \pi_x \rangle + \frac{\hbar}{2}\langle \pi_y | \hat{\ell}_z | \pi_x \rangle
\end{aligned} \quad [186]$$

$\hat{\ell}_z$ is a pure imaginary operator (see the earlier section on Angular Momenta). Therefore,

$$\frac{\hbar}{2}\langle \pi_x | \hat{\ell}_z | \pi_y \rangle = -\frac{\hbar}{2}\langle \pi_y | \hat{\ell}_z | \pi_x \rangle \quad [187]$$

which finally yields (at this level of approximation)

$$\langle {}^1\Sigma^+ | \hat{\mathscr{H}}_{\mathrm{SO}} | {}^3\Sigma^- \rangle = -\hbar \langle \pi_x | \hat{\ell}_z | \pi_y \rangle \quad [188]$$

It is the change of sign of individual integrals due to spin integration that lets the matrix element survive. In other cases, where $\hat{s}_{+1}$ or $\hat{s}_{-1}$ are involved, spin flips of individual orbitals occur in addition. Because of these individual spin flips, the spin–orbit operator can alter not only the M_S quantum number by $\Delta M_S = 0, \pm 1$, but also may change the spin quantum number S by at most one unit (i.e., $\Delta S = 0, \pm 1$).

Examples

Nonlinear Molecules

For nonlinear molecules, no component of the orbital angular momentum is conserved. Although formulas for use in highly symmetric molecules such as octahedral and tetrahedral complexes have been worked out by

Silver,[71] we generally do not employ the WET for the spatial part of a spin–orbit matrix element. In the Russell–Saunders (*LS*) coupling scheme, we still assume, however, that S is a (fairly) good quantum number[d] and that we can expand the true electronic wavefunction in terms of $\vec{S}^2$ eigenfunctions. Given two electronic states, we should like to know whether electronic spin–orbit or spin–spin coupling is symmetry allowed and which of the spin sublevels interact.

Methylene (CH_2, Figure 12) is one of the rare molecules that exhibit a triplet electronic ground state. Assume we are interested in its fine-structure splitting. Apart from the ground state[e] X $^3B_1$, three low-lying excited singlet states exist: a $^1A_1$, b $^1B_1$, and c $^1A_1$.

We should like to answer the following questions:

1. Is there any first-order fine-structure splitting in the electronic ground state, due to either electronic spin–orbit or spin–spin coupling?
2. Which of the excited states, a $^1A_1$ and c $^1A_1$, or b $^1B_1$ is by symmetry allowed to contribute to the fine-structure splitting of X $^3B_1$ in second order?
3. If there is a second-order spin–orbit splitting, can we predict which of the triplet sublevels is lowered in energy due to spin–orbit coupling?

For a matrix element to be nonzero, the direct products of the irreps of space and spin functions on each side have to be equal. Let us check this out first. Spatially, A_1 and B_1 states are involved. Table 11 presents the irreps according to which singlet and triplet spin functions transform under symmetry operations of the C_{2v} molecular point group.

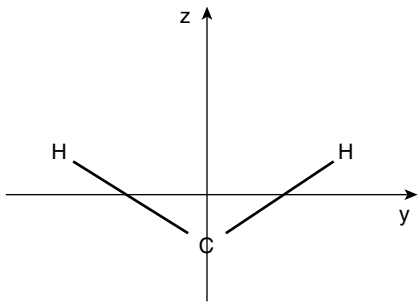

Figure 12 Ground-state equilibrium structure of methylene. The molecule is chosen to lie in the yz plane with z as symmetry axis. With this convention, the triplet ground state is of $^3B_1$ symmetry. If $\vec{x}$ and $\vec{z}$ span the molecular plane, B_1 and B_2 have to be interchanged.

[d] A fairly good quantum number means that coupling between states of different multiplicity is small, and the spin quantum number is close to zero (singlet), one (triplet), and so on.

[e] Following convention in spectroscopy, X labels the electronic ground state, A is the first known excited state of the same spin symmetry as X, B the second, and so on; "a" labels the first excited state of different spin symmetry, b the second, and so on. Symmetry labels for each state are in italics.

Table 12 Direct Product Representations for the Spatial and Spin Wave Functions of the Low-Lying Electronic States of Methylene

State	Spatial Symmetry	Spin Symmetry	$\otimes$
X $^3B_1$	B_1	A_2	B_2
		B_1	A_1
		B_2	A_2
b $^1B_1$	B_1	A_1	B_1
$\{a, c\}$ $^1A_1$	A_1	A_1	A_1

The direct product representations of the space and spin functions are in this particular case given in Table 12. The overall symmetries of the electronic states involved tell us that there is no coupling between b $^1B_1$ and X $^3B_1$, whereas the $^1A_1$ states may interact with X $^3B_1$. However, we can give much more specific answers:

1. Concerning spin–orbit coupling, no component of the angular momentum operator is found in A_1 symmetry. Therefore, there is no first-order contribution of $\hat{\mathscr{H}}_{SO}$ to the fine-structure splitting of X $^3B_1$. This statement is true for all spatially nondegenerate electronic states.
2. On the contrary, both $\mathscr{D}_{2z^2-x^2-y^2}$ and $\mathscr{D}_{x^2-y^2}$ are totally symmetric. Hence, the triplet is split by spin–spin coupling (SSC) into three distinct multiplet levels.
3. The spatial parts of the $\{a, b\}$ $^1A_1$ states can couple to X $^3B_1$ via the y component of the spin–orbit operator. The operator $\hat{\mathcal{S}}_y$ couples the singlet spin function S_0 (A_1) to the B_1 triplet function.

Summarizing, we find that spin–spin interaction is capable of splitting the ground-state triplet spin multiplet into three distinct levels. Which of the multiplet levels is lowered in energy and which is raised due to spin–spin coupling cannot be predicted by means of group theory. This raising or lowering in energy depends on the specific electronic configuration in the considered state. Among the low-lying states, only the $^1A_1$ states are allowed to perturb the X $^3B_1$ ground state via spin–orbit coupling. One of the triplet components (T_y) is lowered in energy due to this perturbation, the two others remain unaltered.

Linear Molecules

In systems with orbitally degenerate states, we can also exploit the Wigner–Eckart theorem for the spatial part of the wave function. Use of the WET further reduces the number of matrix elements that have to be computed explicitly.

In linear molecules or atoms where the separate projection of the total spatial and spin angular momenta on the z axis is meaningful (Russell–

Table 13 Irreps of Singlet (S) and Triplet (T) Spin Functions, the Angular Momentum Operators ($\hat{\vec{\mathscr{L}}}$ and $\hat{\vec{\mathscr{S}}}$), an Irreducible Second-Rank Tensor Operator $\hat{\mathscr{D}}$, and the Position Operators $\hat{\mathscr{X}}\hat{\mathscr{Y}}\hat{\mathscr{Z}}$ in $C_{\infty v}$ ($D_{\infty h}$) Symmetry

Function/Operator		Irrep.
$\mathrm{Triplet}_{M_S=0}$	T_0	$\Sigma^-_{(g)}$
$\mathrm{Triplet}_{M_S=\pm1}$	$\mathrm{T}_{\pm1}$	$\Pi_{(g)}$
$\mathrm{Singlet}_{M_S=0}$	S_0	$\Sigma^+_{(g)}$
$\hat{\mathscr{L}}_z, \hat{\mathscr{S}}_z$	$\hat{\mathscr{L}}_0, \hat{\mathscr{S}}_0$	$\Sigma^-_{(g)}$
$\hat{\mathscr{L}}_{\mp(x\pm iy)/\sqrt{2}}, \hat{\mathscr{S}}_{\mp(x\pm iy)/\sqrt{2}}$	$\hat{\mathscr{L}}_{\pm1}, \hat{\mathscr{S}}_{\pm1}$	$\Pi_{(g)}$
$\hat{\mathscr{D}}_{(2z^2-x^2-y^2)/\sqrt{6}}$	$\mathscr{D}_0$	$\Sigma^+_{(g)}$
$\hat{\mathscr{D}}_{\mp(xz+zx\pm i(yz+zy))/2}$	$\mathscr{D}_{\pm1}$	$\Pi_{(g)}$
$\hat{\mathscr{D}}_{(x^2-y^2\pm i(xy+yx))/2}$	$\mathscr{D}_{\pm2}$	$\Delta_{(g)}$
$\hat{\mathscr{Z}}$	$\hat{\mathscr{T}}_0$	$\Sigma^+_{(u)}$
$\mp(\hat{\mathscr{X}} \pm i\hat{\mathscr{Y}})/\sqrt{2}$	$\hat{\mathscr{T}}_{\pm1}$	$\Pi_{(u)}$

Saunders coupling), M_L and M_S quantum numbers can be changed by the operator $\hat{\mathscr{H}}_{\mathrm{SO}}$ by at most one unit each, while their sum M_J has to remain constant. Employing the nomenclature specific to linear molecules (i.e., $M_J = \Omega$, $M_L = \Lambda$, and $M_S = \Sigma$), the selection rule for spin–orbit coupling reads:

$$\Delta\Omega = 0 \qquad \Delta\Lambda = 0, \pm1 \qquad \Delta\Sigma = 0, \mp1 \qquad [189]$$

To determine the first-order spin–orbit splitting pattern of an orbitally degenerate electronic state, we shall make use of the energy expression obtained from the phenomenological operator, which in this case reduces to $A_{\mathrm{SO}} \cdot \Lambda \cdot \Sigma$ because only the z component of the spin–orbit operator is involved.

Inspection of Table 13 shows that the $M_S = 0$ component of a triplet state transforms like Σ^- in $C_{\infty v}$, and the $M_S = \pm1$ components like the Π irrep. The direct product of space and spin functions is irreducible in the case of the $M_S = 0$ component. For a $^3\Pi$ state, the direct product results in an overall Π symmetry. Reduction of the direct product for $M_S = \pm1$ components is readily carried out and gives $\Pi \otimes \Pi = \Sigma^+ \oplus \Sigma^- \oplus \Delta$. More specifically, one obtains the splitting pattern shown in Table 14. From the entries in Table 14, we read the following results:

- First-order SOC lifts certain degeneracies but not all of them. In the particular example, the six components of a $^3\Pi$ state are split into

Table 14 First-Order Spin–Orbit Splitting of a $^3\Pi$ State

	Product State	$\lvert\Omega\rvert = \lvert\Lambda + \Sigma\rvert$	$\Lambda \cdot \Sigma$	E
Δ	$\Pi_{+1}\mathrm{T}_{+1};\ \Pi_{-1}\mathrm{T}_{-1}$	2	$+1$	$+A_{\mathrm{SO}}$
Π	$\Pi_{+1}\mathrm{T}_0;\ \Pi_{-1}\mathrm{T}_0$	1	0	0
$\Sigma^\pm$	$(1/\sqrt{2})[\Pi_{+1}\mathrm{T}_{-1} \mp \Pi_{-1}\mathrm{T}_{+1}]$	$0^\pm$	-1	$-A_{\mathrm{SO}}$

three doubly degenerate levels. (We should point out that a degeneracy of the $0^{\pm}$ levels is not required by group theory. The $\mathcal{S}_{-1}\mathcal{D}_{+2}\mathcal{S}_{-1}$ and $\mathcal{S}_{+1}\mathcal{D}_{-2}\mathcal{S}_{+1}$ components of the electronic spin–spin coupling operator and higher-order SOC lift their degeneracy. It requires a different type of interaction, that is, either the coupling to an external magnetic field or the rotation of the nuclear framework to split also the $\Omega = 2$ and $\Omega = 1$ sublevels.)

- There is a first-order splitting pattern common to all $^3\Pi$ states, independent of the physical content. All the molecule-dependent physical information is contained in the parameter A_{SO}. These facts are, of course a consequence of the tensor properties expressed in the Wigner–Eckart theorem.

Summary

1. Only in spatially degenerate states, $\hat{\mathscr{H}}_{SO}$ may cause a multiplet splitting in first order.
2. The operator $\hat{\mathscr{H}}_{SO}$ couples states of different spin and space symmetries in second order, independent of spatial degeneracies.
3. It is possible to use the Wigner–Eckart theorem for reducing the number of matrix elements that have to be calculated. Symmetry rules can be obtained from the tensorial structure of the interaction Hamiltonians.
4. As a consequence of the WET, there is a first-order splitting pattern common to all states of a specific space + spin symmetry, independent of the physical content.
5. Double groups: For describing the transformation properties of the α and β spin functions, it is necessary to augment the ordinary molecular point group by symmetry operations that result from multiplying the original symmetry operations by a rotation through 2π. The resulting enlarged point groups are called double groups.
6. Many-particle spin functions:
   Odd number of electrons $\rightarrow$ Fermion irreps of the double group
   Even number of electrons $\rightarrow$ Boson irreps of the double group
7. Selection rules for $\hat{\mathscr{H}}_{SO}$ ($\longleftrightarrow$: coupling allowed; $\leftarrow\backslash\rightarrow$: no interaction; *u* signifies ungerade states)
   a. Systems with an inversion center strictly: $g \longleftrightarrow g$; $u \longleftrightarrow u$; $g \leftarrow\backslash\rightarrow u$
   b. Atoms strictly: $\delta_{JJ'}$, $\delta_{M_J M'_J}$
      LS coupling: $L + L' \geq 1 \geq |L - L'|$; $S' + S \geq 1 \geq |S' - S|$
   c. Linear molecules strictly: $\Delta\Omega = 0$
      LS coupling: $\Delta\Lambda = 0, \pm 1$; $\Delta\Sigma = 0, \mp 1$; $S' + S \geq 1 \geq |S' - S|$
   d. Nonlinear molecules strictly: refer to direct product representations of the appropriate double group.
      LS coupling: S, S' as above

COMPUTATIONAL ASPECTS

General Considerations

Various approaches can be pursued to compute spin–orbit effects. Four-component ab initio methods automatically include scalar and magnetic relativistic corrections, but they put high demands on computer resources. (For reviews on this subject, see, e.g., Refs. 18,19,81,82.) The following discussion focuses on two-component methods treating SOC either perturbationally or variationally. Most of these procedures start off with orbitals optimized for a spin-free Hamiltonian. Spin–orbit coupling is added then at a later stage. The latter approaches can be divided again into so-called one-step or two-step procedures as explained below.

In light molecules, SOC predominantly affects spectral properties such as fine-structure and transition probabilities. Fine-structure splittings originate both from first- and higher-order spin–orbit contributions. In the language of magnetic resonance, these are also dubbed diamagnetic and paramagnetic contributions, respectively. The latter depend both on spin–orbit coupling matrix elements and on spin-free energy differences. Independent of the spin–orbit interaction scheme, it is therefore indispensable to employ methods that take electron correlation into account.

In heavy element compounds, spin–orbit interaction is of concern also for binding energies because the mutual spin–orbit interaction between molecular states will in general be smaller than in the dissociation limit. (Sometimes this is also addressed as quenching of SOC, although the interaction does not disappear completely.) Those molecular states that correlate with the lower spin–orbit component of a heavy element atomic state will therefore be more loosely bound. In contrast, the states that dissociate to the upper atomic spin–orbit level are stabilized by SOC.

Approximately from the 3*d* elements onward, it is necessary to add spin-independent (scalar) relativistic corrections to the Hamiltonian.[18] In early works, the differential scalar relativistic effect to molecular excitation energies was often estimated by adding mass–velocity and Darwin corrections[14] perturbationally. However, even for first- and second-row transition metals, it turned out that scalar relativistic effects are preferably taken into account during the orbital optimization,[83,84] whereas this is mandatory for heavier element compounds.[85] In two-component methods, integrals including kinematic relativistic corrections can be generated either by utilizing relativistic effective core potentials (see, e.g., Refs. 35,37,38) or by employing a one-component relativistic all-electron Hamiltonian that is bounded from below. An overview over state-of-the-art one-component relativistic electronic structure methods has recently been given by Hess and Marian.[19]

Different computational strategies ought to be pursued for heavy main group elements and transition metals:

- In compounds containing heavy main group elements, electron correlation depends on the particular spin–orbit component. The *jj* coupled $6p_{1/2}$ and $6p_{3/2}$ orbitals of thallium, for example, exhibit very different radial amplitudes (Figure 13). As a consequence, electron correlation in the *p* shell, which has been computed at the spin-free level, is not transferable to the spin–orbit coupled case. This feature is named *spin-polarization*. It is best recovered in spin–orbit CI procedures where electron correlation and spin–orbit interaction can be treated on the same footing—in principle at least. As illustrated below, complications arise when configuration selection is necessary to reduce the size of the CI space. The relativistic contraction of the thallium 6*s* orbital, on the other hand, is mainly covered by scalar relativistic effects.
- The most critical part in electronic structure calculations on transition metal compounds is the determination of electron correlation at the spin-free level. Spin–orbit coupling in open *d* shells can be computed very accurately by perturbational expansions.[86] One reason for this behavior lies in the fact that densities and radial expectation values of $d_{3/2}$ and $d_{5/2}$ orbitals do not differ dramatically. Spin-polarization effects are therefore small. Another cause is related to the length of the perturbation expansion. In transition metal atoms, roughly speaking, only terms with equal *d* occupation interact via SOC. At the orbital level, an $s \rightarrow d$ excitation or vice versa yields spin–orbit integrals of the type $\langle s|\hat{\mathscr{H}}_{\mathrm{SO}}|d\rangle$, which are zero. As long as configurations of different *d* occupations do not mix extensively at the spin-free level, perturbation sums can be confined to a manifold of states with a particular *d* occupation.

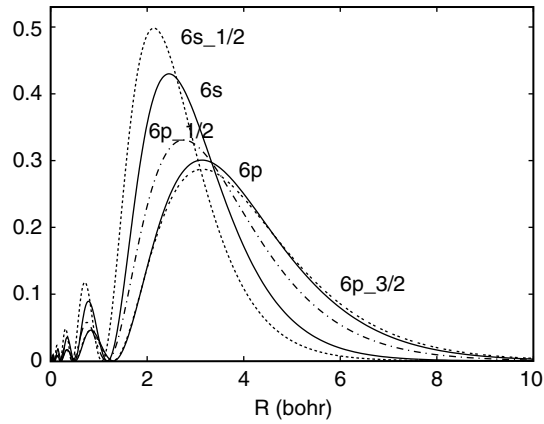

Figure 13 Nonrelativistic 6*s* and 6*p* radial wave functions (solid) versus relativistic $6s_{1/2}$ (dotted), $6p_{1/2}$ (dashed–dotted), $6p_{3/2}$ (dashed) radial wave functions of the thallium atom calculated at the Hartree–Fock and Dirac–Fock levels, respectively.

- Spin-polarization should also play a minor role in lanthanides and actinides. Differential correlation between states of different f occupation and spin–orbit coupling within a given $s^l d^m f^n$ manifold are huge, however. Further, the number of electronic states that can be derived from an $s^l d^m f^n$ occupation is in general too large for an explicit expansion of the spin–orbit coupled states in unperturbed ones. In spite of these difficulties, good progress has been made in recent years.[56]

We confine the discussion in the remainder of this section to the treatment of electronic wave functions. This confinement to electronic wave functions is justified as long as no (sharply avoided) intersystem crossings are present or other non-Born–Oppenheimer effects such as rovibronic (rotational/vibrational/electronic) coupling are involved. Intersystem crossings will be discussed in connection to nonradiative transitions.

Evaluation of Spin–Orbit Integrals

Early implementations of Breit-Pauli spin–orbit integrals were based on Slater-type orbitals (STOs).[63,87–90] All these programs involve numerical integration schemes that become prohibitively expensive in polyatomic molecules. Kern and Karplus[91] proposed a Gaussian transform for STOs, but explicit formulas were only given for s functions. For a while, Gaussian lobe functions were popular because they are composed only of s functions.[92–94] Besides s orbitals, p orbitals are easily described by linear combinations of lobe functions, but the extension to higher angular momentum basis functions leads to numerical problems. King and Furlani[95] derived formulas for evaluating Breit-Pauli (BP) integrals in the basis of Cartesian Gaussians numerically by means of Rys quadrature techniques.

Actual computer codes for polyatomic molecules employ either Cartesian or Hermite Gaussian functions. The calculation of all-electron spin–orbit integrals makes use of second derivatives of Coulomb integrals with respect to nuclear coordinates.[93,96–104] The relation between Coulomb and BP spin–obit integrals is immediately apparent, if the BP operator is written as in expressions [101] and [102]. Instead of acting with ∇ on $1/\hat{r}$ in integrals of $\nabla(1/\hat{r}) \times \nabla$, the ∇ on the left can be applied to the bra vector and the other one to the ket vector. The derivative of an s function is a p function; taking the derivative of a p function yields an s and a d function, and so on. In this way, the spatial parts of a BP integral can be written as a linear combination of Coulomb integrals.

The no-pair spin–orbit Hamiltonian [105] differs from the corresponding BP terms [103] by momentum dependent factors of the type $\hat{A}_i/(\hat{E}_i + mc^2)$ or $(\hat{A}_i\hat{A}_j)/(\hat{E}_i + m_e c^2)$, where $\hat{E}_i$ and $\hat{A}_i$ or $\hat{A}_j$ have been defined in [106] and [107], respectively. There are essentially two ways of taking these factors into account.

- The first was pioneered by Samzow et al.[103] and makes use of a method proposed by Hess for spin-independent one-electron no-pair operators.[25] This approach resembles a resolution-of-the-identity ansatz. As the kinematical factors are functions of the momentum, they are most easily evaluated in momentum space. The auxiliary basis set should span the momentum space as completely as possible. If the set of uncontracted functions is not sufficient for this purpose, it has to be augmented by further primitives. First, the spatial auxiliary basis functions are orthogonalized. The matrix of the nonrelativistic kinetic energy operator $\hat{p}_i^2/2m_e$ is diagonalized in a second step. Its eigenvectors form a discrete representation of the (continuous) spectrum of the momentum operator. If the basis were complete, then every other operator that can be written as a polynomial of the momentum would be diagonal. In a finite basis, this property is only approximately fulfilled. The eigenvectors of the kinetic energy matrix are utilized to evaluate the momentum dependent factors in the no-pair Hamiltonian. The last step is a transformation back to the original auxiliary basis.
- The second way avoids transformations back and forth, which are particularly time consuming for two-electron integrals. Almlöf and co-workers[105] noticed that kinematical factors such as $\hat{A}_i/(\hat{E}_i + m_e c^2)$ or $(\hat{A}_i\hat{A}_j)/(\hat{E}_i + m_e c^2)$ can be turned over to the basis functions instead of applying them to the BP operator. Under certain circumstances, the modified basis functions can be reexpanded in the original basis so that only the contraction coefficients change. In this way, any program that evaluates BP integrals can be utilized to compute approximate no-pair integrals.

Effective core-potential [127] and atomic mean-field spin–orbit operators [128] are in essence one-center operators. Only the projectors contain multicenter terms, but these yield merely overlap integrals. One-center spin–orbit integrals are therefore most easily evaluated in the basis of spherical Gaussians.[64] The computation reduces then to a 1D radial integration and multiplication by analytically determined factors from the angular part. Exploiting the spherical symmetry of the one-center terms thus appreciably speeds up the integral evaluation time and appears to be the only tractable way to perform all-electron spin–orbit calculations in large molecules. For further usage in a molecular code, a basis set transformation to a Cartesian or so-called real spherical Gaussian basis is performed.[f] Atomic mean-field

[f]Real spherical Gaussian basis functions are not proper $\hat{\ell}_z$ eigenfunctions. They are linear combinations of spherical Gaussians $\exp(-\alpha r^2) \cdot (a_m Y_m^\ell + b_{-m} Y_{-m}^\ell)$ with coefficients a_m and b_{-m} chosen such that real basis functions result.

integrals for ab initio model potentials (AIMPs) are evaluated in an all-electron basis first and are transferred then to the AIMP basis.[65] A prerequisite for this procedure to work is an approximate matching of the respective valence orbital exponents and contraction coefficients. Therefore, this approach should only be applied in connection with generalized contracted basis sets of the Raffenetti or atomic natural orbital types.[106,107]

Perturbational Approaches to Spin–Orbit Coupling

As in all perturbational approaches, the Hamiltonian is divided into an unperturbed part $\mathscr{H}^{(0)}$ and a perturbation V. The operator $\mathscr{H}^{(0)}$ is a spin-free, one-component Hamiltonian and the spin–orbit coupling operator takes the role of the perturbation. There is no natural perturbation parameter λ in this particular case. Instead, $\hat{\mathscr{H}}_{\mathrm{SO}}$ is assumed to represent a first-order perturbation $\mathscr{H}^{(1)}$. The perturbational treatment of fine structure is an inherent two-step approach. It starts with the computation of correlated wave functions and energies for pure spin states—mostly at the CI level. In a second step, spin–orbit perturbed energies and wavefunctions are determined.

Rayleigh–Schrödinger Expansion

In Rayleigh–Schrödinger perturbation theory, perturbed wave functions are expanded in the infinite set of eigenfunctions $\Psi^{(0)}$ of the unperturbed Hamiltonian. In practical applications, the sum over states is truncated to a finite number, of course. In principle, the eigenstates $\Psi^{(0)}$ could be constructed from determinants using separate sets of molecular orbitals. Mutually nonorthogonal orbital bases add a complication to the evaluation of matrix elements, in particular for two-electron operators, as the Slater–Condon rules [119] and [120] are not applicable right away. Formulas for self-consistent field wave functions have been worked out by Bearpark et al.[104] These involve bi-orthogonalization of the orbital sets and cofactors resulting from nonunity overlaps. Employing different orbitals for different states (DODS) is desirable in many cases and certainly yields more accurate excitation energies than the use of a common set of molecular orbitals (MOs), but the evaluation of spin–orbit matrix elements for general CI expansions is prohibitively expensive. In most cases, a common set of MOs is used as a one-particle basis for the construction of the determinants. Further, the order of the perturbation expansion is often confined to second order in the energy and to first order in the wave function.

A first-order contribution to the energy is obtained only for spatially degenerate states. Let us assume that the unperturbed state $|\Psi_k^{(0)}\rangle$ has a d-fold degenerate eigenvalue, including both spin and space degeneracies. According to the rules of degenerate perturbation theory, the first-order

energies $E^{(1)}_{k;a}$ are obtained by diagonalizing the interaction matrix in a basis spanned by the d components $|\Psi^{(0)}_{k;b}\rangle$ of $|\Psi^{(0)}_{k}\rangle$, where b is an index from 1 to d.

$$\begin{pmatrix} \langle\Psi^{(0)}_{k;1}|\hat{\mathscr{H}}_{\mathrm{SO}}|\Psi^{(0)}_{k;1}\rangle & \langle\Psi^{(0)}_{k;2}|\hat{\mathscr{H}}_{\mathrm{SO}}|\Psi^{(0)}_{k;1}\rangle & \dots & \langle\Psi^{(0)}_{k;d}|\hat{\mathscr{H}}_{\mathrm{SO}}|\Psi^{(0)}_{k;1}\rangle \\ \langle\Psi^{(0)}_{k;1}|\hat{\mathscr{H}}_{\mathrm{SO}}|\Psi^{(0)}_{k;2}\rangle & \langle\Psi^{(0)}_{k;2}|\hat{\mathscr{H}}_{\mathrm{SO}}|\Psi^{(0)}_{k;2}\rangle & \dots & \langle\Psi^{(0)}_{k;d}|\hat{\mathscr{H}}_{\mathrm{SO}}|\Psi^{(0)}_{k;2}\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle\Psi^{(0)}_{k;1}|\hat{\mathscr{H}}_{\mathrm{SO}}|\Psi^{(0)}_{k;d}\rangle & \langle\Psi^{(0)}_{k;2}|\hat{\mathscr{H}}_{\mathrm{SO}}|\Psi^{(0)}_{k;d}\rangle & \dots & \langle\Psi^{(0)}_{k;d}|\hat{\mathscr{H}}_{\mathrm{SO}}|\Psi^{(0)}_{k;d}\rangle \end{pmatrix} \quad [190]$$

Only one of the matrix elements needs to be evaluated explicitly. All others can be obtained from the reduced matrix element by means of the Wigner–Eckart theorem, Eq. [166]. The eigenvectors

$$|\Psi^{(0)}_{k;a}\rangle = \sum_{b=1}^{d} |c^{(0)}_{k;a,b}|\Psi^{(0)}_{k;b}\rangle \quad [191]$$

are zeroth-order wave functions adapted to this particular perturbation.

If the degeneracy is lifted completely in first-order or if at least second-order effects do not introduce an additional splitting of degenerate levels, the second-order energy $E^{(2)}_{k,l}$ can be expressed as

$$E^{(2)}_{k;a} = \sum_{i\neq k}^{\infty} \sum_{b'=1}^{d'} \frac{\langle\Psi^{(0)}_{k;a}|\hat{\mathscr{H}}_{\mathrm{SO}}|\Psi^{(0)}_{i;b'}\rangle\langle\Psi^{(0)}_{i;b'}|\hat{\mathscr{H}}_{\mathrm{SO}}|\Psi^{(0)}_{k,a}\rangle}{E^{(0)}_{i} - E^{(0)}_{k}} \quad [192]$$

The corresponding first-order perturbed wave function reads

$$|\Psi^{(1)}_{k;a}\rangle = \sum_{i\neq k}^{\infty} \sum_{b'=1}^{d'} \frac{\langle\Psi^{(0)}_{i;b'}|\hat{\mathscr{H}}_{\mathrm{SO}}|\Psi^{(0)}_{k,a}\rangle}{E^{(0)}_{i} - E^{(0)}_{k}} |\Psi^{(0)}_{i;b'}\rangle \quad [193]$$

If these conditions are not fulfilled, an auxiliary condition is necessary to determine the proper linear combinations of the $|\Psi^{(0)}_{k,a}\rangle$: although each set of orthogonal eigenfunctions qualifies as a proper wave function as long as the degeneracy pertains, the wave function coefficients are required to change smoothly when the additional perturbation is switched on adiabatically. Symmetry properties of the involved states and operators may aid in predetermining these linear combinations. In general, this procedure is too complicated, however, and one resorts to quasi-degenerate perturbation theory.

Moreover, first and second orders of perturbation theory are not defined in a stringent way in a molecule. Consider, for instance, all potential energy curves of a diatomic molecule that correspond to a specific dissociation

channel. Near the equilibrium geometry, the low-lying electronic states are usually well separated. Their mutual interaction is then described properly by second-order perturbation theory. In the separated atom limit, first-order degenerate perturbation theory applies. In between, but still close to the dissociation limit, the states are nearly degenerate—or quasi-degenerate as we might say. Neither of the procedures is then strictly applicable and orders of perturbation theory are not well defined.

Quasi-Degenerate Perturbation Theory

An often chosen way out of this dilemma is to set up a so-called perturbation matrix in the basis of eigenvectors of a spin-free secular equation, multiplied by an appropriate spin function. In this basis, matrix elements of the spin-free Hamiltonian occur only in the diagonal, of course. Due to symmetry, diagonal spin–orbit matrix elements come along only with complex wave functions. In a Cartesian basis, the integrals of the electrostatic Hamiltonian and the y component of $\hat{\mathscr{H}}_{\mathrm{SO}}$ are real, whereas x and z integrals exhibit an imaginary phase. (The spatial parts of the integrals are imaginary for all Cartesian components. It is the choice of an imaginary phase for $\hat{\mathcal{S}}_y$ that makes the matrix elements of the y component of $\hat{\mathscr{H}}_{\mathrm{SO}}$ real.) As before, matrix elements are computed only for one representative of a spin multiplet; all other matrix elements are generated by use of the WET.

Similar to first-order degenerate perturbation theory, perturbed energies and wave functions are obtained by matrix diagonalization. This approach has therefore been named quasi-degenerate perturbation theory. In spite of this designation, the procedure is also applicable to states with large energy separations. Other authors prefer to call it *LS contracted* spin–orbit configuration interaction (SOCI) in order to stress its relation to configuration interaction procedures in the limit of a complete set of eigenvectors.[108] The perturbation matrix is in most cases small enough for the application of standard complex Hermitian eigenvalue and eigenvector solvers. Otherwise one can resort to the same methods as in SOCI.

One of the great advantages of quasi-degenerate and Rayleigh–Schrödinger perturbation theories over SOCI procedures is that different levels of sophistication can easily be mixed. For reasons of efficiency or technical limitations, it is in general not possible to a perform a full CI calculation that yields the exact solution within a given basis set. Instead, configurations are selected according to some criterion such as excitation class or energy. Unfortunately, electron correlation contributions are slowly convergent. Furthermore, truncated CI expansions are not size extensive, that is, the correlation energy does not scale properly with the number of electrons. For spin-free states, long-standing experience exists on how to estimate correlation contributions from discarded configurations or excitation classes.[109–115] These extrapolated energies or the eigenvalues of correspondingly dressed Hamiltonians can be taken as diagonal elements combined with spin–orbit matrix

elements of single excitation CI or smaller single and double excitation CI expansions.[41,108,116]

A weakness of these methods lies in the limited number of zeroth-order states that are used for an expansion of the first-order perturbed wave function. In particular, it has been demonstrated that probabilities of spin-forbidden radiative transitions converge slowly with the length of the perturbation expansion.[92]

Variational Perturbation and Response Theory

As an alternative to sum-over-states methods, the perturbation equations can be solved directly. In the context of spin–orbit coupling, reviews on this subject have recently been given by Yarkony[117] and by Ågren et al.[118]

Hess et al.[119] utilized a Hamiltonian matrix approach to determine the spin–orbit coupling between a spin-free correlated wave function and the configuration state functions (CSFs) of the perturbing symmetries. Havriliak and Yarkony[120] proposed to solve the matrix equation

$$(\mathbf{H}^{(0)} - E_i^{(0)}\mathbf{1})\Psi_i^{(1)} = \mathbf{H}_{\mathrm{SO}}\Psi_i^{(0)} \qquad [194]$$

directly for $\Psi_i^{(1)}$ in the basis of CSFs. The direct solution of such a perturbation equation is generally known as Hylleraas (or variational) perturbation theory.[121,122] Yarkony[123] developed this approach further by introducing a symbolic matrix element method, thereby extending the limits relating to the dimensionality of the Hamiltonian matrix.

Also in response theory the summation over excited states is effectively replaced by solving a system of linear equations. Spin–orbit matrix elements are obtained from linear response functions, whereas quadratic response functions can most elegantly be utilized to compute spin-forbidden radiative transition probabilities. We refrain from going into details here, because an excellent review on this subject has been published by Ågren et al.[118] While these authors focus on response theory and its application in the framework of CI and multiconfiguration self-consistent field (MCSCF) procedures, an analogous scheme using coupled-cluster electronic structure methods was presented lately by Christiansen et al.[124]

Variational Procedures

For compounds containing heavy atoms, spin–orbit and electron correlation energies are approximately of the same size, and one cannot expect these effects to be independent of each other. A variational approach that treats both interactions at the same level is then preferable to a perturbation expansion. Special care is advisable in the choice of the spin–orbit operator in this case. The variational determination of spin–orbit coupling requires a spin–orbit

operator that is bounded from below such as the no-pair Hamiltonian,[103] the zeroth-order regular approximation Hamiltonian (ZORA),[125–127] or the ECP spin–orbit operators.[39,40,52] Otherwise one has to constrain the basis set (e.g., by an energy selection) to prevent a variational collapse.

jj *Coupling Methods*

Two-component approaches employing one-particle spinors as basis functions are referred to as *jj* coupling methods. They are typical one-step procedures that take electron correlation and spin–orbit coupling into account simultaneously. Kramers-restricted two-component calculations at the Hartree–Fock, second-order Møller-Plesset, singles and doubles CI, and coupled cluster levels have been reported for atoms and diatomic molecules.[128,129] Technically Lee and co-workers utilized modified versions of four-component programs.[130] Two-component wave function approaches that include SOC right away at the orbital optimization stage are rarely applied to molecules with more than two atoms. Possible reasons are the complications that arise from a formulation of subsequent correlation treatments due to complex-valued orbital coefficients. Among the most widely used two-component methods in the field of chemistry are spin density functional methods.[131–133] In the Kohn–Sham approximation, complex orbital coefficients emerge, but this is not problematic due to the restriction to a single determinant.

Intermediate Coupling Procedures

Most of the variational treatments of spin–orbit interaction utilize one-component MOs as the one-particle basis. The SOC is then introduced at the CI level. A so-called SOCI can be realized either as a one- or two-step procedure. Evidently, one-step methods determine spin–orbit coupling and electron correlation simultaneously. In two-step procedures, typically different matrix representations of the electrostatic and magnetic Hamiltonians are chosen.

In a one-step SOCI, a Hamiltonian matrix is set up in the basis of CSFs or determinants. It contains simultaneously matrix elements of the usual electrostatic Hamiltonian and of the spin–orbit operator and is complex in general. Since spin–orbit coupling mixes spin and spatial degrees of freedom, there are fewer possibilities of blocking the Hamiltonian matrix than for a spin-independent operator. To this end, one makes use of double group symmetries and transforms the Cartesian basis functions to so-called Kramers pairs.[52,134] By a special choice of phases, the matrix can even be made real in some cases. In general, however, very large complex eigenvalue problems have to be solved in a SOCI.

Due to its dimensionality, a complete diagonalization of the CI matrix is not feasible. One usually looks only for a few roots by means of an iterative procedure. For quantum chemical purposes, the Davidson algorithm or modifications thereof have proven to be well suited.[135–137] In its original

formulation, the Davidson method is applicable to real symmetric matrices only. With some care, it can nevertheless be applied also to a complex eigenvalue problem. Every complex Hermitian matrix **C** can be written as

$$\mathbf{C} = \mathbf{A} + i\mathbf{B} \qquad [195]$$

where **A** and **B** are real matrices. Since **C** is Hermitian, **A** is a symmetric matrix, and **B** is skew symmetric.[g] Similarly, the eigenvectors $\vec{Z}_i$ of **C** can be divided into a real part $\vec{U}_i$ and an imaginary part $i\vec{V}_i$, where $\vec{V}_i$ itself is real. The complex eigenvalue problem

$$\mathbf{C}\vec{Z}_i = E_i\vec{Z}_i \qquad [196]$$

can then be replaced by a real symmetric eigenvalue problem of two times its original dimension.[138]

$$\begin{pmatrix} \mathbf{A} & -\mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{pmatrix} \begin{pmatrix} \vec{U}_i \\ \vec{V}_i \end{pmatrix} = E_i \begin{pmatrix} \vec{U}_i \\ \vec{V}_i \end{pmatrix} \qquad [197]$$

All eigenvalues E_i of the real eigenvalue problem [197] are doubly degenerate. This degeneracy is of purely technical origin and should not be confused with Kramers degeneracy,[139,140] which may occur in addition. (For instance, four degenerate roots are obtained for a doublet state, i.e., two for each Kramers level.) In addition to the transposed eigenvector $(\vec{U}_i\vec{V}_i)^t$, a second one with structure $(-\vec{V}_i\vec{U}_i)^t$ is obtained. One of these solutions can be discarded. Alternatively, root-homing procedures can be employed to avoid its evaluation right from the beginning.[136,137]

Older versions of SOCI programs are very I/O intensive because they used to store the Hamiltonian matrix on disk and read it in every iteration step.[52,141] Integral-driven direct methods for spin–orbit coupling came up in the mid 1980s[123,142] following the original fomulation of direct CI methods for spin-independent Hamiltonians.[143–146] Modern direct SOCI programs can easily handle several million determinants.[108,147–151]

The convergence of the iterative determination of eigenvalues and eigenvectors is accelerated appreciably if spin–orbit CI and quasi-degenerate perturbation theory procedures are combined. To this end, the perturbation matrix is set up in the basis of the most important *LS* contracted CI vectors $|\Psi_m^{(0,1)}\rangle$. The solutions of this small eigenvalue problem

$$|\Psi_k^{(1,1)}\rangle = \sum_m^{\text{sf CI states}} c_{km}^{(1,1)} |\Psi_m^{(0,1)}\rangle \qquad [198]$$

are then used to start up the Davidson iteration.

[g] A matrix **B** is said to be skew-symmetric if $B_{ij} = -B_{ji}$.

As the number of electrons increases, the dimension of the CI space becomes increasingly large, even if excitations are restricted to single and double replacements. Many approximate schemes to restrict the number of configurations in a SOCI have been devised. The selected intermediate coupling CI (SICCI) code is an extension of the original COLUMBUS DGCI program and utilizes electrostatic and magnetic interactions as a selection criterion.[152] The SPDIAG program[141] within the BNSOC package[153] is based on the MRD-CI approach. The latter makes use of a correlation energy criterion for configuration selection and estimates the contribution of the discarded configurations to the spin-free correlated energy $E_m^{(0,1)}$ by means of Epstein–Nesbet perturbation theory.[111] If one assumes that the expansion [198] represents a decent approximation to the SOCI solution, the MRD-CI extrapolation scheme can easily be extended to the spin–orbit coupled case.[154,155] Also other approaches toward a balanced treatment of spin–orbit interaction and electron correlation are based on a manipulation of the spin-free energies and wave functions. A pure shift of excitation energies in the Davidson[135] start-up matrix is not sufficient, because the original eigenvalues of $\hat{H}^{(0)}$ are restored during the iterative process.[141] In the so-called spin-free state shifted (SFSS) SOCI method, this problem is circumvented by introducing a projector on the set of $\Psi_m^{(0,1)}$.[156]

$$\hat{\mathscr{H}}_{\mathrm{SO}}^{\mathrm{sfss}} = \hat{\mathscr{H}}_{\mathrm{SO}} + \sum_{m}^{\text{sf CI states}} \epsilon_m |\Psi_m^{(0,1)}\rangle\langle\Psi_m^{(0,1)}| \qquad [199]$$

with

$$\epsilon_m = (E_m - E_{\mathrm{X}}) - (E_m^{(0,1)} - E_{\mathrm{X}}^{(0,1)}) \qquad [200]$$

The ϵ_m quantities shift the spin-free excitation energies $(E_m^{(0,1)} - E_{\mathrm{X}}^{(0,1)})$, calculated at a lower level of correlation treatment, to the exact values $(E_m - E_{\mathrm{X}})$ or at least to higher accuracy estimates. Herein, X denotes a common reference state, in general the electronic ground state.

Due to the structure of the spin–orbit Hamiltonian, off-diagonal blocks of the CI matrix are dominated by single excitations. On the other hand, for electron correlation effects (diagonal blocks) double and higher excitations are decisive. In many approximate schemes, this fact is exploited. One of the first of these approaches was the relativistic CI (RCI) algorithm by Balasubramanian.[157] This is a two-step procedure. In a first step, large-scale multiconfiguration SCF and multireference singles and doubles CI (MRSDCI) calculations are carried out for a spin-free Hamiltonian to determine a set of natural orbitals (NOs). In the second step, spin–orbit interaction integrals are transformed to the NO basis and added to the one-electron Hamiltonian matrix elements of a smaller, truncated MRSDCI in relativistic symmetries. DiLabio and Christiansen[158] studied the separability of spin–orbit and correlation energies for

sixth-row main group compounds. They propose to determine the energy shift D_s^{SO} due to spin–orbit coupling as the difference between spin-free and intermediate coupling single excitation CI calculations. The energy of the correlated spin–orbit coupled states is then estimated by summing D_s^{SO} into the spin-free correlated energies $E_{sd}^{\lambda\sigma}$. Rakowitz et al.[159] employed a SFSS SOCI approach on the Ir^+ ion. They demonstrated that the heavily spin–orbit perturbed spectrum of this ion can be obtained in good accord with experiment at the single excitation level, if higher level correlated electrostatic energies are used to determine the shifts ϵ_m. Recently Vallet et al.[160] extended the effective Hamiltonian approach, originally implemented in the CIPSO code at the level of quasi-degenerate perturbation theory,[41] to be used in intermediate coupling calculations. Instead of representing $\mathscr{H}_{\text{el}}^{\text{eff}} + \hat{\mathscr{H}}_{\text{SO}}$ in the basis of *LS* contracted states, the new effective and polarized spin–orbit CI (EPCISO) operates in a determinantal model space. This model space is chosen as the union of the separate reference spaces in nonrelativistic symmetries augmented by the most significant configurations contributing to either the correlation energy or spin–orbit coupling.

COMPARISON OF FINE-STRUCTURE SPLITTINGS WITH EXPERIMENT

Spectroscopic parameters of a molecule are derived from experimentally determined spectra by fitting term values to a properly chosen model Hamiltonian.[161] Usually, the model Hamiltonian is an effective one-state Hamiltonian that incorporates the interactions with other electronic states parametrically. In rare cases, experimentalists have used a multistate ansatz like the supermultiplet approach[162] to fit the rovibronic spectra of strongly interacting near-degenerate electronic states. The safest way of comparing theoretical data to experiment is to compute the spectrum and to fit the calculated term energies to the same model Hamiltonian as the experimentalists use.

The vibrational dependence of an effective molecular parameter A is usually expressed as

$$A_{v_i} = A_e + \sum_k \alpha_k \left(v_i + \frac{d_i}{2} \right)^k \qquad [201]$$

where A_e is the value of the property at the equilibrium distance, and d_i is the degeneracy of the vibrational mode v_i. The vibrational dependence is caused by the anharmonicity of the potential energy surface and by the variation of A with the vibrational coordinate Q. Such information can in practice be obtained from a quantum chemical treatment only for diatomic and triatomic molecules. To this end, A is fitted to a function of Q—mostly a polynomial or

a cubic spline—and vibrationally averaged, that is, $A(Q)$ is weighted with the probability density of a particular vibrational state $\chi_{v_i}(Q)$ and integrated over the vibrational coordinate Q

$$A_{v_i} = \langle \chi_{v_i}(Q) | A(Q) | \chi_{v_i}(Q) \rangle \qquad [202]$$

Alternatively, expectation values computed at the equilibrium geometry of a given electronic state can be directly compared with experimental parameters extrapolated from rovibrational branches.

In many instances, the "fine-structure" splitting caused by (first-order) spin–orbit coupling is considerably larger than the energy separation between adjacent rotational levels—at least for low values of the total angular momentum J—and rotational excitations can thus be neglected in a first approximation. Hydrides are exceptional in this respect because of their low reduced masses and the resulting large rotational constants B. Methods for computing rovibrational spectra of diatomic molecules from ab initio data including spin–orbit and rotational coupling were proposed among others by Yarkony[117] and by the author[163] employing Hund's case (b) and (a) basis functions, respectively.[161] For a review on the theoretical determination of rovibronic spectra of triatomic molecules, please refer to the work of Perić et al.[164] A general expression for the rotational dependence of a spectroscopic parameter cannot be given. Its functional form varies with the type of basis functions chosen for describing the actual rotation of the nuclear frame; the choice of basis functions in turn depends on the order in which the angular momenta are coupled (Hund's cases in linear molecules) and the type of rotor [spherical top, symmetric top (prolate, oblate), asymmetric top]. Looking at these particular cases in detail goes far beyond the scope of the present chapter. For additional reading, see, for example, Refs. 165–167.

First-Order Spin–Orbit Splitting

Only spatially degenerate states exhibit a first-order zero-field splitting. This condition restricts the phenomenon to atoms, diatomics, and highly symmetric polyatomic molecules. For a comparison with experiment, computed matrix elements of one or the other microscopic spin–orbit Hamiltonian have to be equated with those of a phenomenological operator. One has to be aware of the fact, however, that experimentally determined parameters are effective ones and may contain second-order contributions. Second-order SOC may be large, particularly in heavy element compounds. As discussed in the next section, it is not always distinguishable from first-order effects.

A phenomenological spin–orbit Hamiltonian, formulated in terms of tensor operators, was presented already in the subsection on tensor operators. Few experimentalists utilize an effective Hamiltonian of this form (see Eq. [159]). Instead, shift operators are used to represent space and spin angular

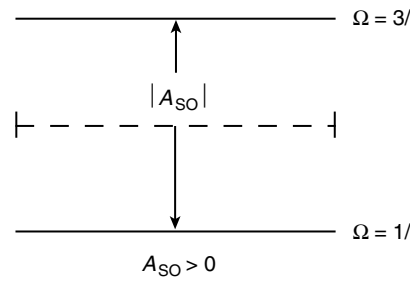

Figure 14 Spin–orbit splitting pattern of a (regular) $^2\Pi_r$ state. (The dashed line marks the position of the unperturbed state.)

momentum operators in their tensorial forms. Employing shift operator conventions and making use of the Russell–Saunders (*LS*) coupling scheme, the phenomenological spin–orbit Hamiltonian reads

$$A_{SO}\hat{\vec{\mathscr{L}}} \cdot \hat{\vec{\mathcal{S}}} = a_{\parallel}\hat{\mathscr{L}}_0\hat{\mathcal{S}}_0 + \frac{1}{2}a_{\perp}(\hat{\mathscr{L}}_+\hat{\mathcal{S}}_- + \hat{\mathscr{L}}_-\hat{\mathcal{S}}_+) \qquad [203]$$

For the sign of the spin–orbit parameter A_{SO}, the following conventions apply:

- A positive A_{SO} denotes a *regular* state (Figure 14). In regular states the sublevel with smallest J or Ω quantum number is lowest in energy.
- A negative A_{SO} denotes an *inverted* state (Figure 15). In inverted states the sublevel with largest J or Ω quantum number is lowest in energy.

Typically, states with less than half-filled shells (particle states) are regular, whereas states with more than half-filled shells (hole states) are inverted.

Atoms: The Landé Interval Rule

To determine the first-order splitting pattern of an atomic state in terms of the phenomenological spin–orbit parameter A_{SO} [Eq. 203], we utilize

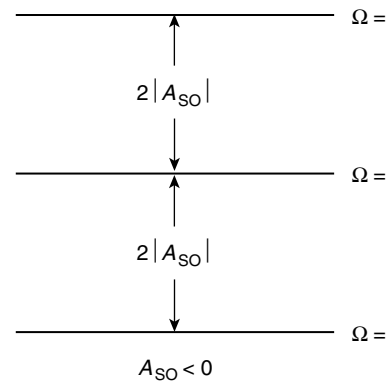

Figure 15 Spin–orbit splitting pattern of an (inverted) $^3\Delta_i$ state.

Russell–Saunders coupling, that is, $\vec{\hat{\mathcal{J}}} = (\vec{\hat{\mathcal{L}}} + \vec{\hat{\mathcal{S}}})$. From $\vec{\hat{J}}^2 = (\vec{\hat{\mathcal{L}}} + \vec{\hat{\mathcal{S}}})^2 = \vec{\hat{\mathcal{L}}}^2 + 2\hat{\mathcal{L}} \cdot \vec{\hat{\mathcal{S}}} + \hat{\mathcal{S}}$, we obtain

$$\vec{\hat{\mathcal{L}}} \cdot \vec{\hat{\mathcal{S}}} = \frac{1}{2}(\hat{\mathcal{J}}^2 - \hat{\mathcal{L}}^2 - \hat{\mathcal{S}}^2) \qquad [204]$$

Exploiting the fact that in first-order perturbation theory all sublevels exhibit identical $\hat{\mathcal{L}}^2$ and $\hat{\mathcal{S}}^2$ eigenvalues, respectively, yields a spin–orbit splitting of

$$E(J) - E(J-1) = A_{SO}J \qquad [205]$$

for two neighboring atomic fine-structure levels with quantum numbers J and $J-1$. This expression is the famous *Landé interval rule*.

As an example where this rule can be applied favorably to determine an atomic spin–orbit parameter, consider the first excited state of atomic copper, $^2D_g(d^9s^2)$. The ^2D_g state of copper is well separated from other electronic states that are allowed by symmetry to couple in second order. A ^2D state gives rise to two fine-structure levels with total angular momentum quantum numbers $J = \frac{3}{2}$ and $J = \frac{5}{2}$. The d^9s^2 configuration corresponds to a hole state, and we therefore expect an inverted splitting pattern (Figure 16). From Landé's interval rule, we determine a splitting of $\frac{5}{2}A_{SO}$. A_{SO} may thus be calculated directly from the experimentally determined splitting[168] of 2042 cm^{-1} by multiplying with $-\frac{2}{5}$, yielding $A_{SO} = -816.8$ cm^{-1}; this value is in good agreement with the theoretically determined spin–orbit parameter of $A_{SO} = -802.4$ cm^{-1}.[169]

Things are not as easy for the $^3D_g(d^9s^1)$ ground state of nickel. From a ^3D atomic state, three fine-structure levels originate with total angular momentum quantum numbers $J = 1$, $J = 2$, and $J = 3$ (Figure 17). The ratio of the measured fine-structure splittings $[E(2) - E(1)]/[E(3) - E(2)]$ amounts to 1.234 instead of $\frac{2}{3}$ as expected from the Landé interval rule. The main cause for this deviation is an energy shift of the $J = 2$ level due to the interaction with neighboring states of the same angular momentum quantum number. The $J = 3$ and $J = 1$ levels of ^3D_g are only weakly perturbed and their energy separation, $E(J = 3) - E(J = 1) = 5A_{SO}$, may be taken to extract an experimental value for $A_{SO} = -301.6$ cm^{-1}. The experimental spin–orbit constant is in excellent agreement with the theoretical value of -301 cm^{-1}.

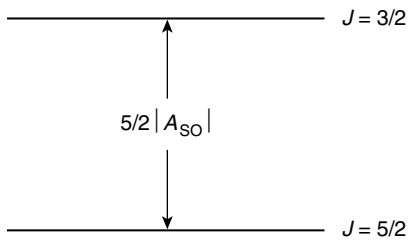

Figure 16 Spin–orbit splitting pattern of an inverted ^2D atomic state.

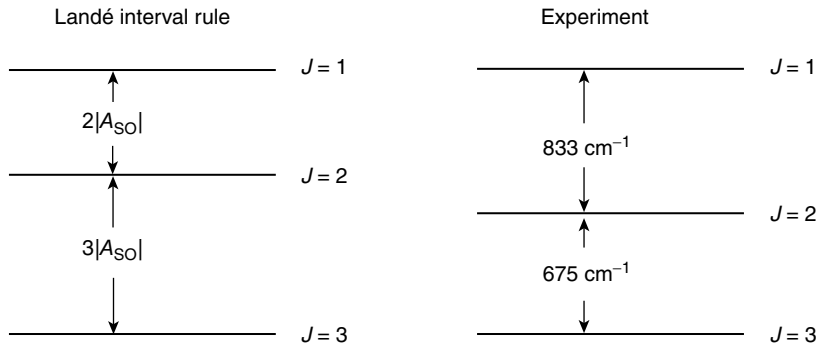

Figure 17 Spin–orbit splitting of the first excited ^3D atomic state of nickel. Experimental data (right) are from Ref. 168.

As the calculations show, the downward shift of the $J = 2$ level with respect to its ideal value is well described by taking into account the second-order interaction with a single perturber, the $^1D_g(d^9s^1)$ state.[116]

Unfortunately, the Landé rule is not obeyed very well in heavy atoms; extraction of an atomic spin–orbit parameter from purely experimental data may be tricky in these cases or even impossible.

Linear Molecules

For a linear molecule, only the z component of $\hat{\mathscr{H}}_{\mathrm{SO}}$ is involved in the evaluation of diagonal matrix elements. In a Hund's case (a) basis,[161] the fine-structure parameter A_e at the equilibrium geometry Q_e can be obtained from

$$A_e = \frac{1}{\Lambda\Sigma} \langle \Psi_n^{(0)}(\vec{r}, \tilde{Q}_e) | \hat{\mathscr{H}}_{\mathrm{SO}} | \Psi_n^{(0)}(\vec{r}, \tilde{Q}_e) \rangle_{\vec{r}} \qquad [206]$$

where the subscript $\vec{r}$ indicates that the integration is confined to the electronic degrees of freedom, and where Λ and Σ denote the projections of the total spatial and spin angular momenta on the internuclear axis, respectively.

Vibrational averaging is performed by multiplying the Q-dependent electronic integral by the appropriate vibrational wave functions $\chi_{v_i}(Q)$ and integrating over the vibrational coordinate Q, that is,

$$A_{v_i} = \frac{1}{\Lambda\Sigma} \langle \chi_{v_i}(Q) | \underbrace{\langle \Psi_n^{(0)}(\vec{r}, \tilde{Q}) | \hat{\mathscr{H}}_{\mathrm{SO}} | \Psi_n^{(0)}(\vec{r}, \tilde{Q}) \rangle_{\vec{r}}}_{A_{\mathrm{SO}}(Q)} | \chi_{v_i}(Q) \rangle_Q \qquad [207]$$

Typically, Hund's case (b)[161] applies to molecules with $\Lambda = 0$, to which spin–orbit effects do not contribute in first order. For states with a nonzero Λ value and a rotational parameter B_v of the same order-of-magnitdue as the spin–orbit parameter A_v, rotational coupling cannot be neglected. In this case, it

appears most convenient to calculate the rovibrational spectrum in a Hund's case (*a*) basis and to fit the fine-structure splittings to a function of the form $\gamma\vec{\mathcal{N}}\cdot\vec{\mathcal{S}}$, where $\vec{\mathcal{N}}$ represents the total angular momentum save for the spin and γ is a proportionality factor. Examples for such a procedure may be found in a study on the $a^3\Sigma^+$ state of NO^+ by Hutter et al.[170] and in the work by Kleinschmidt et al.[171,172] who treat the fine-structure splitting of the $X\ ^2\Pi$ state of the CH radical in this way. Alternatively, Hund's case (b) can be employed.[173]

The strong coupling limit, Hund's case (*c*)[161], cannot be dealt with in first order. Rather, it requires the inclusion of higher-order (at least second-order) spin–orbit coupling in the calculations.

Second-Order Spin–Orbit Splitting

Every electronic state of whatever spin or spatial symmetry may experience an energy shift due to second-order spin–orbit effects. Singlet states are merely shifted in one or the other direction depending on the relative energetic location of the perturber. In doublet states, second-order spin–orbit splitting just adds to the first-order effect—in an atom or a nonrotating molecule they cannot be distinguished experimentally. In triplets or higher multiplicity states, second-order SOC causes deviations from the regular first-order splitting patterns. These have been briefly addressed already in the section on Symmetry and the section on First-Order Spin–Orbit Splitting. Unfortunately, it is not possible to identify second-order SOC effects unambiguously solely on the basis of experimental data, because second-order SOC and first-order electronic spin–spin coupling (SSC) have identical tensorial structures. To see this, we rewrite the spin–orbit perturbation sum

$$\sum_{n=1}^{\infty}\frac{\langle\Psi_0^{(0)}|\hat{\mathscr{H}}_{\mathrm{SO}}|\Psi_n\rangle\langle\Psi_n|\hat{\mathscr{H}}_{\mathrm{SO}}|\Psi_0^{(0)}\rangle}{E_n-E_0}=\langle\Psi_0^{(0)}|\hat{\mathscr{H}}_{\mathrm{SO}}\left(\sum_{n=1}^{\infty}\frac{|\Psi_n\rangle\langle\Psi_n|}{E_n-E_0}\right)\hat{\mathscr{H}}_{\mathrm{SO}}|\Psi_0^{(0)}\rangle \qquad [208]$$

Further, $\hat{\mathscr{H}}_{\mathrm{SO}}$ may be expressed formally as $\hat{\mathscr{H}}_{\mathrm{SO}}=\sum_i\vec{\hat{\mathscr{D}}}_i^{(1)}\cdot\vec{\hat{s}}_i$ where the $\vec{\hat{s}}_i$ are the usual one-electron spin operators and the first rank tensor operators $\vec{\hat{\mathscr{D}}}_i^{(1)}$ denote the spatial part of $\hat{\mathscr{H}}_{\mathrm{SO}}$ related to electron *i*. (In the BP Hamiltonian (Eq. [104]), for example, $\vec{\hat{\mathscr{D}}}_i^{(1)}$ corresponds to the terms in braces.) One then obtains

$$\begin{aligned}&\langle\Psi_0^{(0)}|\sum_i\vec{\hat{s}}_i\vec{\hat{\mathscr{D}}}_i^{(1)}\left(\sum_{n=1}^{\infty}\frac{|\Psi_n\rangle\langle\Psi_n|}{E_n-E_0}\right)\sum_j\vec{\hat{\mathscr{D}}}_j^{(1)}\vec{\hat{s}}_j|\Psi_0^{(0)}\rangle\\&=\langle\Psi_0^{(0)}|\sum_i\sum_j\vec{\hat{s}}_i\underbrace{\vec{\hat{\mathscr{D}}}_i^{(1)}\left(\sum_{n=1}^{\infty}\frac{|\Psi_n\rangle\langle\Psi_n|}{E_n-E_0}\right)\vec{\hat{\mathscr{D}}}_j^{(1)}}_{\tilde{\mathscr{D}}_{ij}^{(2)}}\vec{\hat{s}}_j|\Psi_0^{(0)}\rangle\end{aligned} \qquad [209]$$

The effective operator in Eq. [209] has exactly the same structure as the second term of the Breit-Pauli SSC operator in Eq. [162]. The fine-structure tensor of rank two $\mathscr{D}^{(2)}$ has thus two major contributions

$$\mathscr{D}^{(2)} = \mathscr{D}^{(2)}_{\mathrm{SS}} + \mathscr{D}^{(2)}_{\mathrm{SO}} \qquad [210]$$

In spatially nondegenerate states of light molecules such as the X $^3\Sigma^-$ ground state of O_2, first-order SSC and second-order SOC have the same order of magnitude. However, spin–spin coupling involves only valence orbitals; contributions of closed-shell core orbitals cancel. This situation is not the case for spin–orbit coupling (see the section on Computational Aspects). As a consequence, SSC contributions remain nearly constant with growing nuclear charge, whereas SOC increases considerably. In heavy element components, second-order SOC by far outweighs first-order SSC.

For a comparison with experimentally determined parameters, the calculated total (SO and SS) $\mathscr{D}^{(2)}$ is equated with $\mathscr{D}^{(2)}_{\mathrm{SS}}$ as used in the phenomenological electronic spin–spin operator, Eq. [164].

In low-symmetry molecules, diagonal and off-diagonal matrix elements of the electronic dipolar coupling tensor may contribute to $\langle\Psi_0^{(0)}|\hat{\mathscr{H}}_{\mathrm{SS}}|\Psi_0^{(0)}\rangle$. Therefore, they are specified mostly in terms of their Cartesian components. If symmetry is C_{2v} or higher, the off-diagonal matrix elements of the tensor operator in Eq. [163] vanish (i.e., the principal axes diagonalizing the SCC tensor coincide with the inertial axes). For triplet and higher multiplicity states, one then obtains

$$\langle\Psi_0^{(0)}|\hat{\mathscr{H}}_{\mathrm{SS}}|\Psi_0^{(0)}\rangle = -D_{xx}\mathcal{S}_x^2 - D_{yy}\mathcal{S}_y^2 - D_{zz}\mathcal{S}_z^2 \qquad [211]$$

Reordering and making use of the traceless property $D_{xx} + D_{yy} + D_{zz} = 0$ gives

$$\langle\Psi_0^{(0)}|\hat{\mathscr{H}}_{\mathrm{SS}}|\Psi_0^{(0)}\rangle = D\left(\mathcal{S}_z^2 - \frac{1}{3}\mathcal{S}^2\right) + E(\mathcal{S}_x^2 - \mathcal{S}_y^2) \qquad [212]$$

where D and E are defined as

$$D = -\frac{3}{2}D_{zz} \qquad [213]$$

$$E = \frac{1}{2}(D_{yy} - D_{xx}) \qquad [214]$$

Usually, experimentalists fit spin–spin splittings in terms of these parameters D and E. For linear molecules, $E = 0$ by symmetry, since $D_{yy} = D_{xx}$. In the experimental literature, a different notation (λ_{SS}) is sometimes used in this

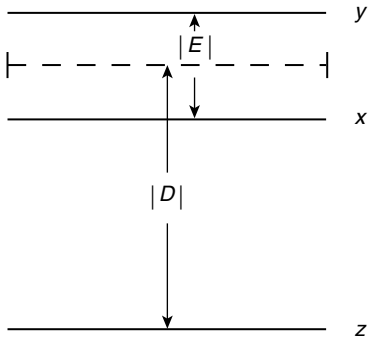

Figure 18 Fine-structure splitting ($D > 0,\ E > 0$) in a triplet state, spatially nondegenerate. (The dashed line is located half-way between T_x and T_y. It is used here merely to illustrate the definition of D. Its position does not indicate the location of the unperturbed triplet.)

case, related to D by $\lambda_{SS} = \frac{1}{2}D$. For similar reasons, atoms in S states do not show a fine-structure splitting caused by electronic SSC at all.

For triplets, it follows from Eq. [213] that D is positive/negative, if T_z is the lowest/highest component. Likewise E is positive/negative, if T_x is located energetically below/above T_y. Schematically, the fine-structure splitting of a spatially nondegenerate triplet state is shown in Figure 18.

SPIN-FORBIDDEN TRANSITIONS

Besides fine-structure splitting, the occurrence of spin-forbidden transitions is the most striking feature in which spin–orbit interaction manifests itself. Radiative spin-forbidden transitions in light molecules usually take place at the millisecond time scale, if the transition is dipole allowed. A dipole- and spin-forbidden transition is even weaker, with lifetimes of the order of seconds. Proceeding down the periodic table, spin-forbidden transitions become more and more allowed due to the increase of spin–orbit coupling. For molecules containing elements with principal quantum number 5 or higher (and the late first-row transition metals Ni and Cu), there is hardly any difference between transition probabilities of spin-allowed and spin-forbidden processes.

For radiationally long-lived excited states, other, nonradiative, depopulation mechanisms are often the lifetime determinating processes. Nonradiative transitions occur between states of similar energy. Under collision-free conditions—which we consider here—a nonradiative transition between different states requires a nonvanishing electronic coupling matrix element *and* a sufficient overlap of the vibrational and rotational wave functions. [Experiments are performed at finite pressure (gas phase) or dilution (solution), of course. Collision induced self-quenching rates are usually eliminated by linear extrapolation of measurements at finite concentrations to an infinitely dilute solution or zero pressure (Stern–Volmer plot[174]).] The overlap requirement is usually fulfilled, if the potential surfaces intersect or undergo an avoided

crossing, but this is not a necessary condition. Spin-allowed nonradiative transitions are called *internal conversion* (IC). They are mentioned here only for the sake of completeness. We rather focus on their spin-forbidden counterparts, that is, intersystem crossings (ISC).

An overview of the energetics and possible depletion mechanisms of excited electronic states is named a *Jablonski diagram*. Herein, singlet states are symbolized by S_0, S_1, S_2, and so on, and triplets by T_0, T_1, T_2, and so on, where the index labels their energetic order and should not be confused with tensor components. A typical Jablonski diagram for an organic molecule is shown in Figure 19.

Transition probabilities W for concurrent processes are additive and independent of the type of decay mechanism:

$$W_{\text{total}} = \frac{1}{g} \sum_{i}^{\text{channels}} W_i \qquad [215]$$

The number g is the degeneracy of the excited state, and the summation is over all possible ways of depletion. The probability W_i refers to any (radiative or

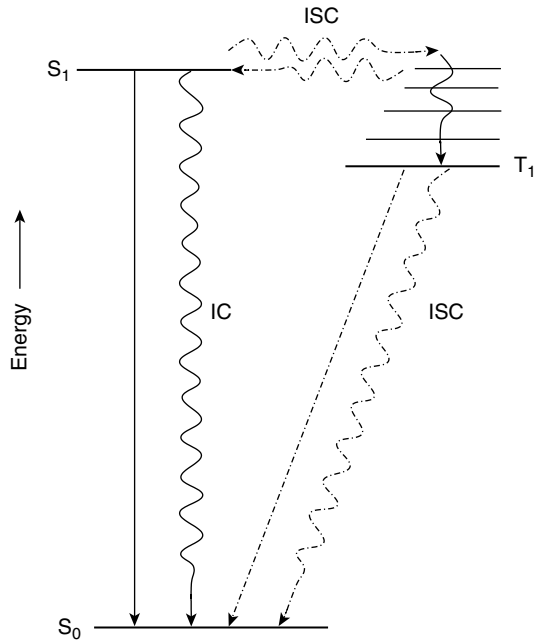

Figure 19 Jablonski diagram (schematic) showing the energetic location of the first excited singlet S_1 and triplet states T_1 with respect to the electronic singlet ground state S_0 and possible transitions between them. Radiative transitions are indicated by straight arrows, nonradiative processes by curly ones. Solid arrows represent spin-allowed transitions, dashed–dotted lines spin-forbidden ones.

nonradiative) depletion mechanism. Since lifetime and transition probabilities are related by $\tau = 1/W$, the total lifetime of an excited state is given by

$$\frac{1}{\tau_{\text{total}}} = g \sum_{i}^{\text{channels}} \frac{1}{\tau_i} \qquad [216]$$

This relationship means that the fastest process dominates the decay rate of an excited state, and the fastest is very often a nonradiative transition.

Whether transitions from the spin–orbit split components of an electronic state can be observed individually depends on the size of the fine-structure splitting and on the temperature. For large splittings and very low temperatures (usually below 10 K), transition probabilities are determined separately for the multiplet components. In this case, the components are treated as nondegenerate levels ($g = 1$). For closely spaced fine-structure levels, spin–lattice relaxation mechanisms cause a spin-depolarization at elevated temperatures. In the high-temperature limit—for organic molecules the temperature of liquid nitrogen (77 K) is usually sufficient—all fine-structure components are equally populated. Their individual transition probabilities have to be averaged then (degeneracy $g = 2S + 1$). In between, a Boltzmann distribution can be used to estimate the population of the individual levels.

Radiative Transitions

A typical spin–forbidden radiative transition is the phosphorescence from the first excited triplet state to the singlet ground state of an organic molecule. Another well-known example for a spin-forbidden radiative transition concerns singlet oxygen. Singlet oxygen is long-lived and can be considered as metastable with respect to the triplet ground state. Spin-forbidden radiative transitions may occur also between doublet and quartet states. Chemists are less aware of those transitions because only a few stable molecules such as NO exist with an unpaired number of electrons.

Spin–Orbit Perturbed Wave Functions

To work out the fundamentals, we employ Rayleigh–Schrödinger perturbation theory and restrict ourselves to a few terms in the (in principle) infinite perturbation sums [193] because things are more transparent in this way. In an actual calculation, it is more advantageous to use methods that avoid an explicit summation over states, such as spin–orbit CI,[41,52,117,141,147,160] variational perturbation theory,[119,120,123] or response theory.[118,124,175] Excellent reviews on the latter two subjects have been given by Yarkony[117] and by Ågren et al.[118] We therefore refrain from going into the details of these methods.

As an example, we investigate the spin-forbidden transition from an excited triplet (a) to a singlet electronic ground state (X). The resulting expressions can easily be extended to other cases, however. Let Ψ denote a spin–orbit coupled state and Φ a pure *LS* state. We express the unperturbed wave

function of the electronic ground state as a bra vector.

$$\langle \Psi_{\mathrm{X}}^{(0)}| = \langle {}^1\Phi_{\mathrm{X}}^{(0)}| \qquad [217]$$

Because of the spin–orbit selection rules, only triplet zeroth-order states contribute to the first-order perturbation correction of a singlet wave function. In Rayleigh–Schrödinger perturbation theory, the expansion coefficient a_i^* of a triplet zeroth-order state $\langle {}^3\Phi_i^{(0)}|$ is determined by its spin–orbit matrix element with the electronic ground state (in the numerator) and its energy difference with respect to the latter (in the denominator).

$$\langle \Psi_{\mathrm{X}}^{(1)}| = \sum_i^{\text{triplets}} a_i^* \langle {}^3\Phi_i^{(0)}| = \sum_i^{\text{triplets}} \frac{\langle {}^1\Phi_{\mathrm{X}}^{(0)}|\hat{\mathscr{H}}_{\mathrm{SO}}|{}^3\Phi_i^{(0)}\rangle}{{}^3E_i^{(0)} - {}^1E_{\mathrm{X}}^{(0)}} \langle {}^3\Phi_i^{(0)}| \qquad [218]$$

The three multiplet components of an excited triplet state are degenerate in zeroth order. We have therefore, in principle, the freedom of choosing these in their spherical or Cartesian forms. On the other hand, the spin–orbit split triplet levels will transform according to the irreps of the molecular point group. For a smooth variation of the wave function gradient with respect to the perturbation parameter λ, we employ *Cartesian triplet spin functions* also in the unperturbed case and express them as ket vectors:

$$|\Psi_{\mathrm{a}}^{(0)}, \mathrm{T}_x\rangle = |{}^3\Phi_{\mathrm{a}}^{(0)}, \mathrm{T}_x\rangle \qquad [219]$$

$$|\Psi_{\mathrm{a}}^{(0)}, \mathrm{T}_y\rangle = |{}^3\Phi_{\mathrm{a}}^{(0)}, \mathrm{T}_y\rangle \qquad [220]$$

$$|\Psi_{\mathrm{a}}^{(0)}, \mathrm{T}_z\rangle = |{}^3\Phi_{\mathrm{a}}^{(0)}, \mathrm{T}_z\rangle \qquad [221]$$

Accordingly, the first-order spin–orbit perturbation of a triplet wave function may be written as a linear combination of unperturbed singlet, triplet, and quintet states with expansion coefficients defined in a similar way as those in Eq. [218].

$$\begin{aligned} |\Psi_{\mathrm{a}}^{(1)}, \mathrm{T}_x\rangle &= \sum_i^{\text{singlets}} b_i |{}^1\Phi_i^{(0)}\rangle + \sum_j^{\text{triplets}} b_j' |{}^3\Phi_j^{(0)}\rangle + \sum_k^{\text{quintets}} b_k'' |{}^5\Phi_k^{(0)}\rangle \\ &= \sum_i^{\text{singlets}} \frac{\langle {}^1\Phi_i^{(0)}|\hat{\mathscr{H}}_{\mathrm{SO}}|{}^3\Phi_{\mathrm{a}}^{(0)}, \mathrm{T}_x\rangle}{{}^1E_i^{(0)} - {}^3E_{\mathrm{a}}^{(0)}} |{}^1\Phi_i^{(0)}\rangle + \cdots \end{aligned} \qquad [222]$$

$$\begin{aligned} |\Psi_{\mathrm{a}}^{(1)}, \mathrm{T}_y\rangle &= \sum_i^{\text{singlets}} c_i |{}^1\Phi_i^{(0)}\rangle + \sum_j^{\text{triplets}} c_j' |{}^3\Phi_j^{(0)}\rangle + \sum_k^{\text{quintets}} c_k'' |{}^5\Phi_k^{(0)}\rangle \\ &= \sum_i^{\text{singlets}} \frac{\langle {}^1\Phi_i^{(0)}|\hat{\mathscr{H}}_{\mathrm{SO}}|{}^3\Phi_{\mathrm{a}}^{(0)}, \mathrm{T}_y\rangle}{{}^1E_i^{(0)} - {}^3E_{\mathrm{a}}^{(0)}} |{}^1\Phi_i^{(0)}\rangle + \cdots \end{aligned} \qquad [223]$$

$$|\Psi_{\mathrm{a}}^{(1)}, \mathrm{T}_z\rangle = \sum_{i}^{\text{singlets}} d_i |{}^1\Phi_i^{(0)}\rangle + \sum_{j}^{\text{triplets}} d'_j |{}^3\Phi_j^{(0)}\rangle + \sum_{k}^{\text{quintets}} d''_k |{}^5\Phi_k^{(0)}\rangle$$

$$= \sum_{i}^{\text{singlets}} \frac{\langle {}^1\Phi_i^{(0)} | \hat{\mathscr{H}}_{\mathrm{SO}} | {}^3\Phi_{\mathrm{a}}^{(0)}, \mathrm{T}_z\rangle}{{}^1E_i^{(0)} - {}^3E_{\mathrm{a}}^{(0)}} |{}^1\Phi_i^{(0)}\rangle + \cdots \qquad [224]$$

Electric Dipole Transitions

In terms of the first-order perturbed wave functions, the matrix element for an electric dipole transition moment is given by

$$\mu_{\mathrm{el}} = \langle \Psi_{\mathrm{X}}^{(0)} + \Psi_{\mathrm{X}}^{(1)} | \sum_j e\vec{r}_j | \Psi_{\mathrm{a}}^{(0)} + \Psi_{\mathrm{a}}^{(1)}\rangle \qquad [225]$$

We disregard here the triplet and quintet contributions to the perturbed triplet wave functions $|\Psi_{\mathrm{a}}^{(1)}\rangle$, because they are not connected directly to the singlet ground state by dipole transition elements. (This approximation is not valid, if the perturbing triplet or quintet state is similar in energy to the perturbed triplet.) Further, the higher-order term $\langle \Psi_{\mathrm{X}}^{(1)} | \sum_j e\vec{r}_j | \Psi_{\mathrm{a}}^{(1)}\rangle$ is usually neglected. (Higher-order terms are of importance, if the perturbing state is energetically near degenerate with one of the perturbed states and has a large coefficient in the perturbation expansion. In this case, a sum-over-states method cannot be employed.) We now express the electric dipole transition moment

$$\mu_{\mathrm{el}} = \langle \Psi_{\mathrm{X}}^{(0)} | \sum_j e\vec{r}_j | \Psi_{\mathrm{a}}^{(1)}\rangle + \langle \Psi_{\mathrm{X}}^{(1)} | \sum_j e\vec{r}_j | \Psi_{\mathrm{a}}^{(0)}\rangle \qquad [226]$$

in terms of contributions from the zeroth-order states. For the T_x component of the excited triplet state, one finds

$$\mu_{\mathrm{el}}(\mathrm{a}_x, \mathrm{X}) = \sum_{i}^{\text{singlets}} b_i \langle {}^1\Phi_{\mathrm{X}}^{(0)} | \sum_j e\vec{r}_j | {}^1\Phi_i^{(0)}\rangle + \sum_{i}^{\text{triplets}} a_i^* \langle {}^3\Phi_i^{(0)}, \mathrm{T}_x | \sum_j e\vec{r}_j | {}^3\Phi_{\mathrm{a}}^{(0)}, \mathrm{T}_x\rangle$$

$$= \sum_{i}^{\text{singlets}} \frac{\langle {}^1\Phi_i^{(0)} | \hat{\mathscr{H}}_{\mathrm{SO}} | {}^3\Phi_{\mathrm{a}}^{(0)}, \mathrm{T}_x\rangle}{{}^1E_i^{(0)} - {}^3E_{\mathrm{a}}^{(0)}} \langle {}^1\Phi_{\mathrm{X}}^{(0)} | \sum_j e\vec{r}_j | {}^1\Phi_i^{(0)}\rangle$$

$$+ \sum_{i}^{\text{triplets}} \frac{\langle {}^1\Phi_{\mathrm{X}}^{(0)} | \hat{\mathscr{H}}_{\mathrm{SO}} | {}^3\Phi_i^{(0)}, \mathrm{T}_x\rangle}{{}^3E_i^{(0)} - {}^1E_{\mathrm{X}}^{(0)}} \langle {}^3\Phi_i^{(0)}, \mathrm{T}_x | \sum_j e\vec{r}_j | {}^3\Phi_{\mathrm{a}}^{(0)}, \mathrm{T}_x\rangle \qquad [227]$$

Similar expressions are obtained for the other two multiplet components. The restriction to T_x levels of perturbing triplet states in Eq. [227] is a consequence of the fact that the electric dipole operator does not contain spin-dependent

terms and thus may couple only equal spin components. This is different for magnetic dipole transitions as discussed below.

There is no restriction as to the phase of the coefficients: they can take positive or negative, real or imaginary values. Computing oscillator strengths f or transition probabilities W, one first has to perform the summation in Eq. [227] before multiplying μ_{el} by its complex conjugate.

$$W_{el}(a,X) = \frac{4e^2}{3c^3\hbar^4}(E_a - E_X)^3 \mu_{el}(a,X)\mu_{el}^*(a,X) \qquad [228]$$

Expressing W in units of reciprocal seconds (s^{-1}), ΔE in reciprocal centimeters (cm^{-1}), and μ_{el} in atomic units (ea_0), the numerical value of the prefactor becomes 2.0261×10^{-6}.

In the picture of spin–orbit perturbed Russell–Saunders states, the dipole transition moment of a spin-forbidden radiative transition is thus a sum of spin-allowed dipole transitions weighted by spin–orbit coupling coefficients (e.g., the expansion coefficients in Eq. [218]). The fact that the transition dipole moment of a spin-forbidden radiative transition is a weighted sum of spin-allowed dipole transition moments is exactly what experimentalists mean when they speak of *intensity borrowing*. The contribution of perturbing states to the oscillator strength can be positive or negative. In other words, perturbers can not only lend intensity to a spin-forbidden transition, they can also take it away.

One feature particular to spin-forbidden transitions is worth noting. The perturbation expansion of the excited triplet state components [222], [223], and [224] may include the unperturbed ${}^1\Phi_X^{(0)}$ ground state. Similarly, the unperturbed ${}^3\Phi_a^{(0)}$ first excited state may contribute to first-order perturbed singlet wave function [218]. For an electronic dipole transition, this means that in addition to off-diagonal matrix elements of $\sum_j e\vec{r}_j$, diagonal elements occur in Eq. [227]. In other words, dipole moments of the ground and excited electronic states are involved—their difference to be more precise. While in spin-allowed transitions, length and velocity forms of the electric dipole transition operator give identical results, save for the incompleteness of the wave function representation, one has to exercise care in spin-forbidden radiative transitions. All diagonal elements of the usual velocity form of the dipole transition operator vanish. This discrepancy was resolved by Lohr[176] and Goodman and Laurenzi.[177] Instead of using $\sum_i \hat{\vec{p}}_i$, the appropriate transition dipole operator in the velocity form is given in Breit-Pauli theory by

$$\hat{\vec{\Pi}} = \sum_i \left(\hat{\vec{p}}_i + \frac{e^2\hbar}{2m_e^2c^2} \hat{\vec{s}}_i \times \vec{\nabla}_i \left\{ \sum_\alpha \frac{Z_\alpha}{\hat{r}_{i\alpha}} - \sum_{j \neq i} \frac{1}{\hat{r}_{ij}} \right\} \right) \qquad [229]$$

where α denotes the nuclei and i and j the electrons. The operator $\hat{\vec{\Pi}}$ can be expressed by the commutator $[\hat{\mathscr{H}}_0 + \hat{\mathscr{H}}_{SO}, \vec{r}] = (i\hbar/m_e)\hat{\vec{\Pi}}$. Thus $e\vec{r}$ is a correct

dipole transition operator. Similar considerations apply to the no-pair operator. For the calculation of transition moments of spin-forbidden radiative dipole transitions, the length form of the dipole transition operator should therefore be employed.

An Example: The Phosphorescence of Dithiosuccinimide Many thiocarbonyls have photostable excited ($n \rightarrow \pi^*$) and ($\pi \rightarrow \pi^*$) states that tend to relax by photophysical rather than photochemical processes.[177,178] Recently, the electronic spectra of dithioimides have been under experimental and theoretical investigation.[179–181] The spin-forbidden radiative decay of the lowest-lying triplet state of dithiosuccinimide may serve as an example to illustrate the results of the previous sections. Experimentally a lifetime of 0.10 ± 0.01 ms was determined for the T_1 state.[179] This value has been corrected for solvent effects, but the transition may include radiative as well as nonradiative depletion mechanisms.

Dithiosuccinimide (Figure 20) exhibits a planar ground state equilibrium configuration. The ground state has $^1A_1$ symmetry. The first excited states result from $n \rightarrow \pi^*$ excitations. Meskers et al.[179] suggest that the observed fluorescence is caused by the spin-forbidden radiative transition from the a^3B_1 state to the X^1A_1 ground state. Indeed, multireference CI calculations at the ground-state equilibrium geometry find a^3B_1 to be the first excited state.[180] There are various other low-lying valence states of $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ type from which the spin-forbidden radiative transition $X^1A_1 \leftarrow a^3B_1$ may borrow intensity (Figure 21). The strongest measured fluorescence is the $\pi \leftarrow \pi^*$ transition $X^1A_1 \leftarrow 1^1B_2$.[179] Further, $n \leftarrow \pi^*$ transitions from 1^1A_2 (which is symmetry forbidden in the C_{2v} group) and 1^1B_1 have been observed.

Let us assume for the time being that the molecule retains a planar conformation in the a^3B_1 state. [This assumption was checked by performing a geometry optimization in the a^3B_1 and 1^3A_2 states. It is known from other (thio)ketones that $n \rightarrow \pi^*$ excited triplet states exhibit a pyramidal configuration at the keto carbon. The a^3B_1 state has a planar equilibrium structure, whereas the geometry of 1^3A_2 is slightly distorted toward C_2 symmetry (CCCC dihedral angle about 10°; out-of-plane angles at the thiocarbons less than 1°).] Irreducible representations of the triplet spin functions and dipole operators in C_{2v} symmetry are listed in Table 11. We employ the spin–orbit operator in its Cartesian form [160] and remember that the compound space

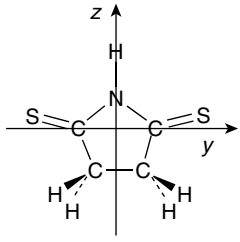

Figure 20 Dithiosuccinimide. The five-membered ring is chosen to lie in the *yz* plane.

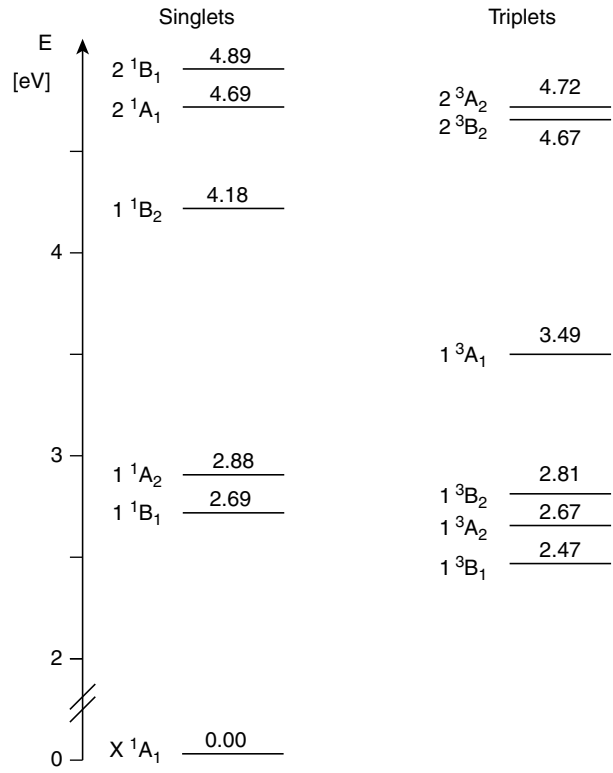

Figure 21 Calculated vertical electronic spectrum of dithiosuccinimide. [Adapted from Ref. 180.]

and spin operator is totally symmetric. Direct product representations for the singlet and triplet states of dithiosuccinimide are displayed in Table 15.

The spin and space parts of the $^1A_1$ ground-state wave function are totally symmetric. As a consequence, only states of overall symmetry A_1 can contribute to the first-order perturbed ground-state wave function

$$\langle \Psi_{\mathrm{X}}^{(1)}| = a_1^* \langle 1^3 A_2^{(0)}, \mathrm{T}_z| + a_2^* \langle \mathrm{a}^3 B_1^{(0)}, \mathrm{T}_y| + a_3^* \langle 1^3 B_2^{(0)}, \mathrm{T}_x| + \cdots \qquad [230]$$

The product symmetries of the excited a^3B_1 multiplet components are A_2, A_1, and B_2. The spin–orbit perturbed excited state wave functions are therefore given by

$$\begin{aligned} |\Psi_{\mathrm{a}}^{(1)}, \mathrm{T}_x\rangle &= b_1 |1^1 A_2^{(0)}\rangle + \cdots \\ |\Psi_{\mathrm{a}}^{(1)}, \mathrm{T}_y\rangle &= c_1 |\mathrm{X}^1 A_1^{(0)}\rangle + \cdots \\ |\Psi_{\mathrm{a}}^{(1)}, \mathrm{T}_z\rangle &= d_1 |1^1 B_2^{(0)}\rangle + \cdots \end{aligned} \qquad [231]$$

Table 15 Direct Product Representations for Space and Spin Parts of Singlet and Triplet Wave Functions in C_{2v} Symmetry

State	Spatial Symmetry	Spin Symmetry	Spatial ⊗ Spin Symmetry
$^1A_1$	A_1	A_1	A_1
$^3A_1$,T_x	A_1	B_2	B_2
,T_y		B_1	B_1
,T_z		A_2	A_2
$^1A_2$	A_2	A_1	A_2
$^3A_2$,T_x	A_2	B_2	B_1
,T_y		B_1	B_2
,T_z		A_2	A_1
$^1B_1$	B_1	A_1	B_1
$^3B_1$,T_x	B_1	B_2	A_2
,T_y		B_1	A_1
,T_z		A_2	B_2
$^1B_2$	B_2	A_1	B_2
$^3B_2$,T_x	B_2	B_2	A_1
,T_y		B_1	A_2
,T_z		A_2	B_1

Note that there is no contribution from the near-degenerate first excited singlet state because states of equal spatial symmetries do not interact via SOC in C_{2v} and higher point groups.

The evaluation of the transition moment is straightforward now. Even though the energy difference between a^3B_1 and 1^1A_2 is small and the T_x level is strongly perturbed, the dipole transition to the ground state is forbidden as long as the molecule remains planar because of the dipole selection rules: this transition would require an operator of A_2 symmetry, but $\hat{x}$, $\hat{y}$, and $\hat{z}$ transform like B_1, B_2, and A_1, respectively. The transition may gain some intensity due to second-order spin–vibronic interactions, however.

$$\langle \mathrm{X}| \sum_j e\vec{r}_j |\mathrm{a}, \mathrm{T}_x\rangle = 0 \qquad \text{(in } C_{2\mathrm{v}} \text{ symmetry)} \tag{232}$$

$$\langle \mathrm{X}| \sum_j e\vec{r}_j |\mathrm{a}, \mathrm{T}_y\rangle = a_2^* \langle \mathrm{a}^3B_1^{(0)}| \sum_j e\hat{z}_j |\mathrm{a}^3B_1^{(0)}\rangle + c_1 \langle \mathrm{X}^1A_1^{(0)}| \sum_j e\hat{z}_j |\mathrm{X}^1A_1^{(0)}\rangle + \cdots \tag{233}$$

$$\langle \mathrm{X}| \sum_j e\vec{r}_j |\mathrm{a}, \mathrm{T}_z\rangle = d_1 \langle \mathrm{X}^1A_1^{(0)}| \sum_j e\hat{y}_j |1^1B_2^{(0)}\rangle + a_1^* \langle 1^3A_2^{(0)}| \sum_j e\hat{y}_j |\mathrm{a}^3B_1^{(0)}\rangle + \cdots \tag{234}$$

A careful inspection of the spin-perturbation coefficients a_2^* and c_1 shows that their absolute values are equal but they differ in sign, that is, $a_2^* = -c_1$. It is

thus the difference of ground and first excited state dipole moments that enters in Eq. [233]. It is typically the case that the difference of the dipole moments is decisive and not particular to dithiosuccinimide.

The actual lifetime calculations on dithiosuccinimide were performed by means of quasi-degenerate perturbation theory using multireference CI wave functions as zeroth-order states.[181] Spin–orbit matrix elements were evaluated in the no-pair one-center mean-field approximation. The computed radiative lifetimes amount to 62.5 and 3.3 ms for the T_y and T_z levels, respectively. This yields a high-temperature value of 9.4 ms, about two orders-of-magnitude longer than the experimental $\tau = 0.1$ ms. Although both the accurate calculation and measurement of phosphorescence lifetimes are difficult tasks, this large discrepancy between theory and experiment is unacceptable. There are two possible explanations: (1) The 1^3B_2 state undergoes an effective intersystem crossing with the ground state such that the measured lifetime is mainly due to nonradiative depletion. (2) The observed transition originates from the 1^3A_2 state for which Tatchen et al.[181] computed a phosphorescence lifetime of only 0.6 ms—the same order-of-magnitude as the measured 0.1 ms. As before, most intensity is borrowed from the $X^1A_1 \leftarrow 1^1B_2$ fluorescence, in accord with the measured polarization of the band. Further investigations are required to decide upon these two alternative interpretations.

Magnetic Dipole Transitions

In a multipole expansion of the interaction of a molecule with a radiation field, the contribution of the magnetic dipole is in general much smaller than that of the electric dipole. The prefactor for a magnetic dipole transition probability differs from the one for an electric dipole by $\alpha^2/4 \approx 1.3 \times 10^{-5}$. Magnetic dipoles may play an important role, however, when electric dipole transitions are symmetry-forbidden as, e.g., in homonuclear diatomics.

Usually, the magnetic dipole transition operator is found in the literature in a form that reads (neglecting constants)

$$\vec{\mu}_{\mathrm{L}} = \mu_{\mathrm{B}} \sum_j \vec{\ell}_j \qquad [235]$$

In spin-forbidden transitions there is a second, nonnegligible term that describes a spin flip. Instead of Eq. [235], one has to employ

$$\vec{\mu}_{\mathrm{mag}} = \mu_{\mathrm{B}} \sum_j (\vec{\ell}_j + g_e \vec{s}_j) \qquad [236]$$

The spin-flip term

$$\frac{\langle b^1\Sigma_g^+ | \hat{\mathscr{H}}_{\mathrm{SO}} | X^3\Sigma_g^-, T_0 \rangle}{E_b - E_X} \langle X^3\Sigma_g^-, T_0 | \sum_j g_e \vec{s}_j | X^3\Sigma_g^-, T_{\pm 1} \rangle \qquad [237]$$

was shown to make the major contribution to the transition probability $b^1\Sigma_g^+ \rightarrow X^3\Sigma_g^-$ in O_2.[182,183]

Again expressing W in units of reciprocal seconds (s^{-1}), ΔE in reciprocal centimeters (cm^{-1}), and μ_{mag} in atomic units ($\hbar$), the probability of a magnetic dipole transition from an initial state i to the final state f is given by

$$W_{mag}(i,f) = 2.6973 \times 10^{-11}(E_i - E_f)^3 \mu_{mag}(i,f)\mu^*_{mag}(i,f) \qquad [238]$$

Nonradiative Transitions

Bound electronic states exhibit a discrete spectrum of rovibrational eigenstates below the dissociation energy. The interaction between discrete levels of two bound electronic states may lead to perturbations in their rovibrational spectra and to nonradiative transitions between the two potentials. In the case of an intersystem crossing, this process is often followed by a radiative depletion. Above the dissociation energy and for unbound states, the energy is not quantized, that is, the spectrum is continuous. The coupling of a bound state to the vibrational continuum of another electronic state leads to predissociation.

Apart from the selection rules for the electronic coupling matrix element, spin-forbidden and spin-allowed nonradiative transitions are treated completely analogously. Nonradiative transitions caused by spin–orbit interaction are mostly calculated in the basis of pure spin Born–Oppenheimer states. With respect to spin–orbit coupling, this implies a *diabatic behavior*, meaning that curve crossings may occur in this approach. The nuclear Schrödinger equation is first solved separately for each electronic state, and the rovibronic states are spin–orbit coupled then in a second step.

In principle, static or dynamic approaches can be employed to describe the nonradiative transition between two states, although dynamic approaches are generally preferable. In a time-independent formalism, the coupling manifests itself in a mixing between vibrational levels of the involved electronic states. The relative probability of finding the molecule in a particular state is proportional to its squared expansion coefficient. In the language of time-dependent perturbation theory, the coupling is expressed as a propagation back and forth between vibronic states. The equivalence between the static and dynamic picture holds as long as there are no other fast depletion mechanisms. In particular, the radiative lifetimes of the interacting states have to be long enough for an equilibration. If one of the states is short-lived, the nonradiative transition occurs in essence only in one direction. The theory and computational treatment of the coupling between multidimensional potential energy surfaces have recently been reviewed by Köppel and Domcke.[184]

Here, we confine the discussion to cases that can be reduced to one dimension. As far as bound-continuum interactions are concerned, we restrict ourselves to weak interactions. This condition is mostly fulfilled for

spin-forbidden processes. Otherwise the full apparatus of scattering theory has to be invoked, which is far beyond the scope of this chapter.

Bound–Bound Interactions

Perturbations in the spectrum involving a pair of vibrationally bound states occur only occasionally, for instance, when the vibrational ladders of the electronic states approximately match in energy. The mixing coefficients can be calculated from the potential energy curves of the states involved and the electronic coupling matrix element for various values of the vibrational coordinate. As noted previously, the Schrödinger equation for the vibrational motion is solved separately for each of the bound states. Their solutions are taken as the basis for a matrix representation of the interaction Hamiltonian. To ease the integration over the vibrational coordinate Q, the electronic coupling matrix element X is often approximated by an analytic function of Q, mostly a polynomial or a cubic spline function. The matrix elements of the perturbation matrix are then calculated in a way similar to the vibrational averaging of spectroscopic parameters described in the previous section on the Comparison of Fine-Structure Splittings with Experiment. Instead of employing the same vibronic level on both sides, now

$$H_{ij} = \langle \chi_{v_i}(Q) | X(Q) | \chi_{v'_j}(Q) \rangle_Q \qquad [239]$$

is evaluated, where $X(Q)$ is the off-diagonal electronic coupling, and v_i and v'_j belong to different electronic states. Diagonalization of this matrix directly yields energies and mixing coefficients of the levels.

In practical applications, one often combines experimental and theoretical information. Potential energy curves, deduced from experiment with high accuracy, if available, are employed together with theoretically determined Q dependent coupling matrix elements. As the degree of mixing between vibronic states strongly depends on their energetic separation, even uncertainties in the electronic excitation energy of some 10 wavenumbers—which is a small error by present-day standards—are often sufficient to change the order-of-magnitude for a transition rate. Coupling matrix elements, on the other hand, can be computed with high accuracy, whereas they are not really available from experimental data.

From Eq. [239], it is apparent that the size of a particular H_{ij} is not only determined by the magnitude of the electronic coupling matrix element but also by the overlap of the vibrational wave functions v_i and v'_j. Squared overlap integrals of the type $|\langle \chi_{v_i}(Q) | \chi_{v'_j}(Q) \rangle_Q|^2$ are frequently called *Franck–Condon* (FC) factors. In contrast to radiative processes, FC factors for nonradiative transitions become particularly unfavorable if two states differing considerably in their electronic energies exhibit similar shapes and equilibrium coordinates of their potential curves. Due to the near-degeneracy requirement, an upper state vibrational wave function, with just a few nodes

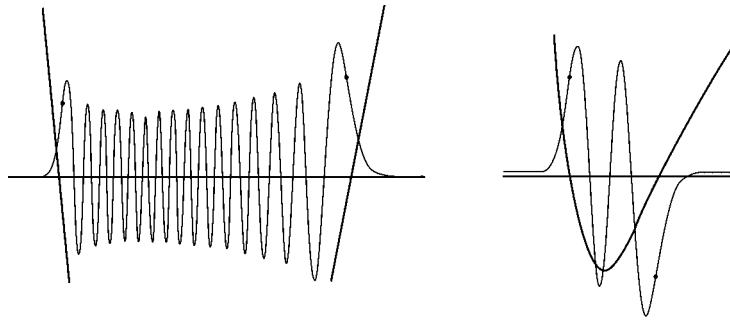

Figure 22 Schematic drawing of nuclear wave functions with 30 (to the left) and 3 (to the right) vibrational quanta. The little dots indicate the vibrational wave function amplitude at the classical turning points of the potentials. Portions of the potential energy curves are shown as thick lines. The horizontal lines indicate the energy of the vibronic (vibrational plus electronic) states.

(Figure 22, right), has to interact with a highly oscillatory vibrational wave function of the ground state (Figure 22, left). As a consequence, low vibrational levels of excited triplet states are often metastable with respect to nonradiative depletion despite a considerable electronic coupling matrix element. An example is the $a^3\Pi$ state of carbon monoxide. Although its $v = 3$ level is nearly degenerate with the $v' = 30$ level of the $X^1\Sigma^+$ electronic ground state and the spin–orbit matrix element amounts to about 35 cm^{-1}, there is hardly any coupling between the states.[185]

In the case of a curve crossing, FC factors profit from the fact that the amplitudes of highly excited states are large at the turning points. Levels that lie energetically close to the intersection will therefore have a nonnegligible overlap. Even weak spin–orbit matrix elements (as occur, e.g., in organic molecules) are often sufficient to make an intersystem crossing an effective depletion channel for excited triplet states.

Bound–Continuum Interactions

Continuum wave functions are spatially extended and are not normalizable in the usual spatial sense. Instead an energy normalization is chosen.[186]

In the weak interaction approximation, the Wentzel–Rice formula can be applied to calculate the predissociation line width Γ and thereby the lifetime for the nonradiative decay $\tau = \hbar/\Gamma$.[187–190]

$$\Gamma = \frac{2\pi}{\hbar}|\langle\chi_{v_i}(Q)|X(Q)|\chi_E(Q)\rangle_Q|^2 \qquad [240]$$

Here $X(Q)$ represents the electronic matrix element of any operator coupling the potentials under consideration. The Wentzel–Rice approximation restricts the coupling of a bound initial vibrational state χ_{v_i} to a single continuum state with energy $E = E_i$. Schematically, the interaction between a bound

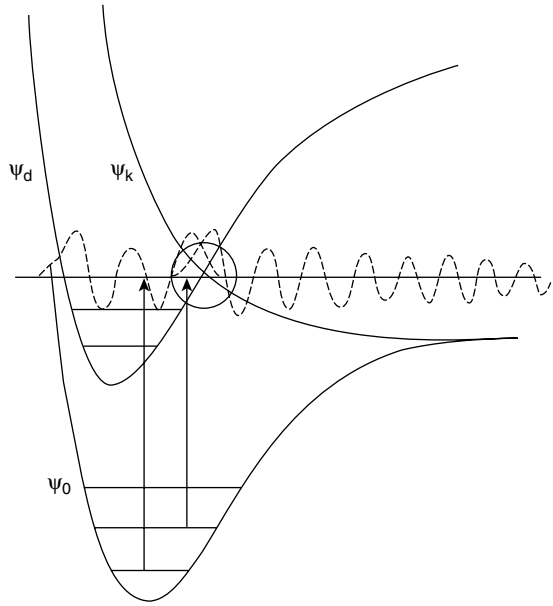

Figure 23 Predissociation of the $v = 2$ vibrational level of the bound electronic state Ψ_d by a vibrational continuum wave function of the dissociative electronic state Ψ_k after radiative excitation (arrows) from the electronic ground state Ψ_0. The circle around the potential curve crossing point indicates an area of large overlap between the vibrational wave functions.

vibrational wavefunction and a continuum wave function of equal energy is depicted in Figure 23.

If the electronic matrix element does not vary significantly with the vibrational coordinate Q, nuclear and electronic factors can be separated in the sense of a Franck–Condon approximation. The line width Γ may then be estimated by

$$\Gamma \approx \frac{2\pi}{\hbar} |\langle \chi_{v_i}(Q) | \chi_{v_E}(Q) \rangle_Q|^2 \cdot |\langle X(Q_x) \rangle|^2 \qquad [241]$$

where $\langle X(Q_x) \rangle$ is the electronic coupling matrix element at the crossing point.

The variation of FC factors (FCFs) with vibrational quantum number and the variation's dependence on the relative slopes of the potential energy curves has been discussed in detail by Murrell and Taylor[191] for the Schumann–Runge bands of O_2. Assuming a constant electronic coupling, the line widths depend on the FCFs only. Murrell and Taylor chose three model cases similar to the ones sketched in Figure 24. In case of a potential energy curve crossing of type (**a**), only one or two vibrational states with energies close to the energy at the crossing point have a marked overlap with the continuum.[191]

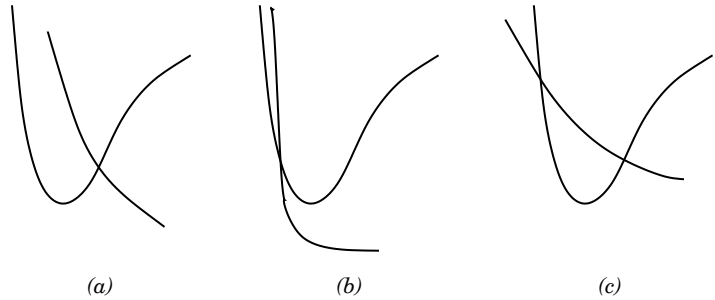

Figure 24 Model cases for the potential energy curve crossing between a bound and dissociative state. (**a**) The potential energy curves cross approximately at right angles. This is often the case when the dissociative state intersects the bound state on its outer limb, i.e., at bond distances longer than its equilibrium internuclear separation. (**b**) Bound and dissociative state exhibit similar slopes and cross on the inner limb of the bound state. (**c**) The dissociative state crosses the bound potential both on the inner and outer limbs.

A Franck–Condon-type approximation might be appropriate then, even if $\langle X(Q)\rangle$ varies strongly with the vibrational coordinate.[192,193] In contrast, many bound vibrational levels will be predissociated in case (**b**). The FCFs will change slowly from one vibrational level to the next one. A strong variation of the electronic interaction matrix element precludes the possibility of applying the FC approximation. Case (**c**) is in between. The FCFs do not vary as strongly with vibrational quantum number as in case (**a**), but their distribution shows marked maxima in the proximity of the two crossing points. An extension to crossings of multidimensional curves interacting via spin–orbit coupling was discussed by Schön and Köppel.[194]

In practical applications, the continuum is often approximated by a discrete spectrum. To this end, one conveniently introduces a potential wall at long internuclear separations and solves for the artifically bound states.[171,172] Alternatively, basis set expansion techniques can be employed.[195,196] In either case, the density of states depends on external conditions, that is, the size of the box or the number of basis functions. This dependence on external conditions has to be accounted for by the energy normalization. Instead of employing a single continuum wave function with proper energy E in Eq. [240], one samples over the discrete levels with energy E_j

$$\Gamma(E_j) = \frac{4\pi}{\hbar(E_{j+1} - E_{j-1})} |\langle \chi_{v_i}(Q)|X(Q)|\chi_{E_j}(Q)\rangle_Q|^2 \qquad [242]$$

and determines $\Gamma(E)|_{E=E_i}$ by interpolation.[192,195] In this expression, the mean energy separation between neighboring states has been used for normalization.

Even for states of equal multiplicity spin–orbit coupling may be the rate-determining process for predissociation. The $b^4\Sigma_g^-$ state of O_2^+ (see Figure 25)

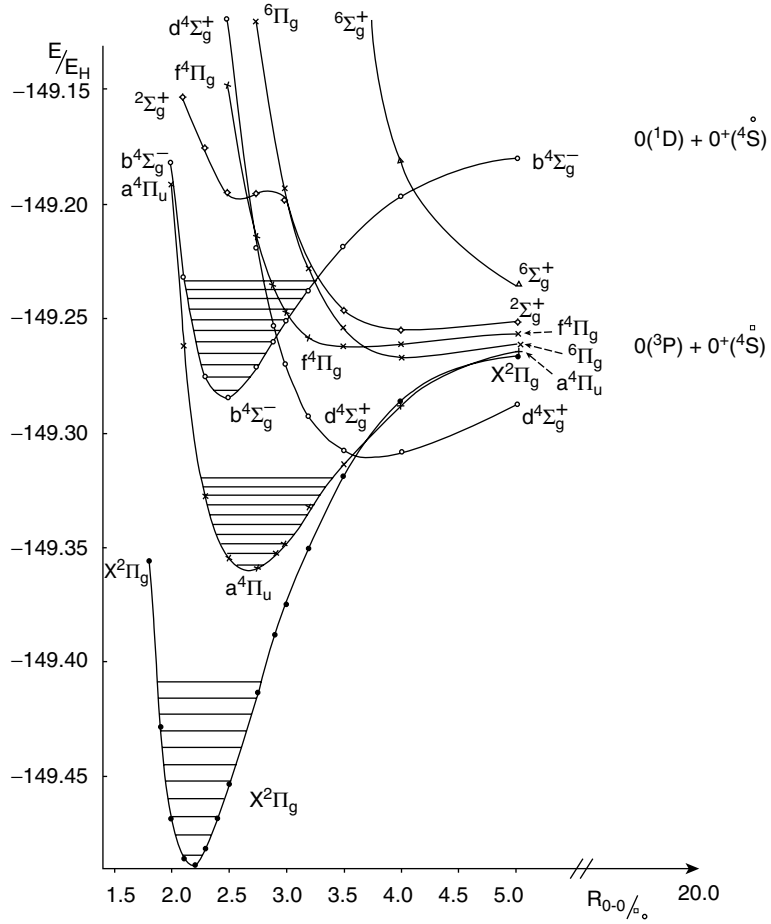

Figure 25 Potential energy curves of O_2^+ computed by multireference CI methods. Based on Ref. 27.

can undergo a spin-allowed radiative transition to the metastable $a^4\Pi_u$ and nonradiative transitions to the states correlating with the ground state dissociation channel $O(^3P_g) + O^+(^4S_u)$.[192] For the vibrational levels $v_0 \ldots v_3$, the spin-allowed fluorescence to the $a^4\Pi_u$ state appears to be the fastest depletion mechanism ($\tau_{rad} \approx 10^{-6}$ s). Vibrational levels v_4 and v_5 are strongly predissociated by $d^4\Sigma_g^+$, v_6 and v_7 by $f^4\Pi_g$. The predominant coupling mechanism is spin–orbit coupling in both cases. Rotational predissociation plays only a minor role here. It is noteworthy that molecular rates for predissociation by spin–orbit interaction are larger by three to four orders-of-magnitude than spin-allowed radiative transition rates.

SUMMARY AND OUTLOOK

Spin–orbit interaction plays an important role in many areas of chemistry, physics, and biology. It causes two essential effects. The first one is a splitting of multiplet levels, referred to as *zero-field splitting* or *fine-structure splitting*. The energetic separation between the multiplet components increases strongly with growing nuclear charge. In heavy element compounds, spin–orbit splittings are of a size comparable to excitation energies between different electron configurations. Second, spin–orbit coupling mixes electronic states of different multiplicities, thus allowing radiative transitions (phosphorescence) and nonradiative transitions (intersystem crossing) between them. Phosphorescence and intersystem crossing rates determine, for instance, the lifetime of the first electronically excited triplet state of an organic molecule. The availibility of molecules in this state plays a key role in photochemical reactions, photosynthesis, the photobiological activity of drugs, luminescence, and so on.

The first spin–orbit coupling Hamiltonian was derived by Pauli in the 1920s. As was shown later, the Breit-Pauli Hamiltonian is not bounded from below. Strictly, it may only be employed in first-order perturbation theory. Neither the Douglas–Kroll transformed no-pair Hamiltonians [Eq. 105] nor the zero-order regular approximation spin–orbit (ZORA) operators suffer from being unbounded. They are suitable for applications in which heavy elements are involved and can safely be utilized in variational procedures such as the spin–orbit configuration interaction approach. To ease the computational effort connected with the evaluation of two-electron spin–orbit integrals, several efficient one-electron operators have been devised. Among these, the most reliable appear to be the all-electron mean-field Hamiltonian and semilocal model Hamiltonians parameterized to be used with effective core potentials.

The tensorial structure of the spin–orbit operators can be exploited to reduce the number of matrix elements that have to be evaluated explicitly. According to the Wigner–Eckart theorem, it is sufficient to determine a single (nonzero) matrix element for each pair of multiplet wave functions; the matrix element for any other pair of multiplet components can then be obtained by multiplying the reduced matrix element with a constant. These vector coupling coefficients, products of $3j$ symbols and a phase factor, depend solely on the symmetry of the problem, not on the particular molecule. Furthermore, selection rules can be derived from the tensorial structure; for example, within an *LS* coupling scheme, electronic states may interact via spin–orbit coupling only if their spin quantum numbers S and S' are equal or differ by 1, i.e., $S = S'$ or $S = S' \pm 1$.

Symmetry considerations are instrumental in a qualitative discussion of spin–orbit effects. Qualitatively, a phenomenological Hamiltonian of the form $A_{SO}\hat{\vec{\mathscr{L}}} \cdot \hat{\vec{\mathcal{S}}}$ suffices. In actual calculations, however, this operator must not be utilized. The operators $\hat{\vec{\mathscr{L}}}$ and $\hat{\vec{\mathcal{S}}}$ form a basis for a matrix representation of the usual molecular point group. The same is true for the spatial and spin wave

functions of a molecule with an even number of electrons. On the other hand, a group theoretical treatment of similarity transformations of spin functions with an odd number of electrons requires the introduction of additional symmetry operations. Together with the molecular point group operations, they form the elements of the so-called molecular double group. Selection rules for the interaction of molecular electronic states via the spin–orbit coupling operator can thus be derived from their direct product representations of the molecular double group.

A phenomenological Hamiltonian is also useful for comparison with experiment, as experimentalists determine spectroscopic parameters by fitting spectral data to such a model Hamiltonian. Only spatially degenerate states experience a first-order spin–orbit splitting. A second-order spin–orbit shift may occur in any electronic state. For triplets and higher multiplicity states, a first-order spin–spin interaction cannot be distinguished experimentally from a second-order spin–orbit interaction. This indistinguishability has to be kept in mind when comparing theoretical and experimental fine-structure splittings, in particular for spatially nondegenerate states of light element compounds.

For the evaluation of probabilities for spin-forbidden electric dipole transitions, the length form is appropriate. The velocity form can be made equivalent by adding spin-dependent terms to the momentum operator. A sum-over-states expansion is slowly convergent and ought to be avoided, if possible. Variational perturbation theory and the use of spin–orbit CI expansions are conventional alternatives to elegant and more recent response theory approaches.

Rates for nonradiative spin-forbidden transitions depend on the electronic spin–orbit interaction matrix element as well as on the overlap between the vibrational wave functions of the molecule. Close to intersections between potential energy surfaces of different space or spin symmetries, the overlap requirement is mostly fulfilled, and the intersystem crossing is effective. Interaction with vibrationally unbound states may lead to predissociation.

The classical field for the application of spin–orbit Hamiltonians in electronic structure calculations is spectroscopy. Fine-structure splittings in the spectra of small light molecules have been predicted with large confidence since the 1970s. Recent developments in relativistic electronic structure theory have pushed the line further down the periodic table and have made heavy element compounds accessible. Spin independent relativistic effects are easily incorporated in effective core potentials or through modified one-electron integrals in all-electron calculations. For either case, corresponding variationally stable spin–orbit operators have been developed. The event of reliable effective one-electron spin–orbit operators has extended the applicability of two-component methods to medium-sized molecules, with 10–20 nonhydrogenic atoms. Methods for a reliable determination of off-diagonal spin–orbit coupling matrix elements in larger molecules still are to be developed.[197]

One of the grand challenges of the 1990s was the merger of spin–orbit and electron correlation effects. There has been good progress, but there is still room for improvement. The reasons for complications in the accurate simultaneous determination of spin–orbit and electron correlation are manifold. Spin–orbit interaction is dominated by single excitations, whereas these hardly contribute at all to electron correlation corrections. For the latter, double and higher excitations are required. Further, the amount of electron correlation obtained in a quantum chemical calculation converges slowly with the expansion length. The number of configurations per electronic state that can taken into account for describing electron correlation effects in a spin–orbit calculation is usually much smaller than for a spin independent Hamiltonian. One reason is the much less efficient symmetry blocking of the Hamiltonian matrix due the coupling of electronic states with different spin and spatial symmetries. Additionally, the matrix elements of the spin–orbit operator are complex numbers in general, thus doubling the dimension of the eigenvalue problem. Another cause of complications is the use of a common set of orbitals for all states under consideration. The treatment of orbital relaxation effects through a configuration expansion requires large reference spaces with many different active orbitals and thus prevents the use of highly efficient concepts, such as the unitary group approach, for evaluating the Hamiltonian matrix elements. One way out of these problems appears to be the usage of so-called dressed Hamiltonians[160] that incorporate major dynamic correlation effects at the spin-free level. This procedure allows the number of explicitly varied expansion coefficients in a spin–orbit calculation to be kept at a moderate size at little or no loss of correlation contributions. This latter discussion shows that quantum chemical approaches that include spin–orbit coupling effects are far from the stage of black-box programs.

ACKNOWLEDGMENTS

This chapter was derived from lecture notes on a course about spin-dependent interactions in molecules at the University of Bonn. Continuing discussions with many colleagues and students have been helpful in pointing out problems and eliminating errors. Research on this subject was financially supported by the German Research Council (DFG) through SFB334 and SPP "Relativistische Effekte". Traveling funds by the European Science Foundation in the framework of the ESF program on "Relativistic Effects" and the German Academic Exchange Service (DAAD) have been essential for establishing scientific contacts outside Germany. Major parts of this chapter have been written while holding a position at the Heinrich-Heine-University in Düsseldorf.

REFERENCES

1. W. Gerlach and O. Stern, *Z. Phys.*, **8**, 110 (1922). Der experimentelle Nachweis des magnetischen Moments des Silberatoms.
2. W. Gerlach and O. Stern, *Z. Phys.*, **9**, 349 (1922). Der experimentelle Nachweis der Richtungsquantelung im Magnetfeld.

3. S. Goudsmit and G. Uhlenbeck, *Naturwissenchaften,* **13,** 953 (1925). Ersetzung der Hypothese vom unmechanischen Zwang durch eine Forderung bezüglich des inneren Verhaltens jedes einzelnen Elektrons.
4. S. Goudsmit and G. Uhlenbeck, *Nature* (*London*), **117,** 264 (1926). Spinning Electrons and the Structure of Spectra.
5. R. Feynman, R. B. Leighton, and M. L. Sands, *The Feynman Lectures on Physics, Vol. III,* California Institute of Technology, Pasadena, CA, 1989.
6. A. Landé, *Z. Phys.,* **5,** 231 (1921). Über den anomalen Zeemaneffekt (Teil I).
7. A. Landé, *Z. Phys.,* **7,** 398 (1921). Über den anomalen Zeemaneffekt (Teil II).
8. A. Landé, *Phys. Z.,* **22,** 417 (1921). Anomaler Zeemaneffekt und Seriensysteme bei Ne und Hg.
9. W. Pauli, *Z. Phys.,* **31,** 373 (1925). Über den Einfluβ der Geschwindigkeitsabhängigkeit der Elektronenmasse auf den Zeemaneffekt.
10. W. Pauli, *Z. Phys.,* **31,** 765 (1925). Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren.
11. P. A. M. Dirac, *Proc. R. Soc. A,* **117,** 610 (1928). The Quantum Theory of the Electron.
12. L. D. Landau and E. M. Lifschitz, *Lehrbuch der Theoretischen Physik*, Vol. IVa, Akademie Verlag, Berlin, 1975.
13. E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra*, Cambridge University Press, Cambridge, UK, 1967.
14. F. Gross, *Relativistic Quantum Mechanics and Field Theory,* Wiley, New York, 1993.
15. G. Breit, *Phys. Rev.,* **34,** 553 (1929). The Effect of Retardation on the Interaction of Two Electrons.
16. A. Gaunt, *Philos. Trans. R. Soc. A,* **228,** 151 (1929). IV. The Triplets of Helium.
17. B. A. Hess, C. M. Marian, and S. D. Peyerimhoff, in *Advanced Series in Physical Chemistry, Vol.2, Modern Structure Theory, Part I,* C.-Y. Ng and D. R. Yarkony, Eds., World Scientific, Singapore, 1995, pp. 152–278. Ab Initio Calculation of Spin–Orbit Effects in Molecules Including Electron Correlation.
18. J. Almlöf and O. Gropen, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1996, Vol. 8, pp. 203–244. Relativistic Effects in Chemistry.
19. B. A. Hess and C. M. Marian, in *Computational Molecular Spectroscopy,* P. Jensen and P. R. Bunker, Eds., Wiley, Chichester, 2000, pp. 152–278. Relativistic Effects in the Calculation of Electronic Energies.
20. K. G. Dyall, *J. Chem. Phys.,* **100,** 2118 (1994). An Exact Separation of the Spin-Free and Spin-Dependent Terms of the Dirac-Coulomb-Breit Hamiltonian.
21. K. G. Dyall, *Int. J. Quantum Chem.*, **78,** 412 (2000). Relativistic Electric and Magnetic Property Operators for Two-Component Transformed Hamiltonians.
22. W. Pauli, *Z. Phys.,* **43,** 601 (1927). Zur Quantenmechanik des magnetischen Elektrons.
23. L. L. Foldy and S. A. Wouthuysen, *Phys. Rev.,* **78,** 29 (1950). On the Dirac Theory of Spin 1/2 Particles and Its Non-Relativistic Limit.
24. M. Douglas and N. M. Kroll, *Ann. Phys. (N.Y.),* **82,** 89 (1974). Quantum Electrodynamical Corrections to the Fine Structure of Helium.
25. B. A. Hess, *Phys. Rev. A,* **33,** 3742 (1986). Relativistic Electronic-Structure Calculations Employing a Two-Component No-Pair Formalism with External-Field Projection Operators.
26. G. Jansen and B. A. Hess, *Phys. Rev. A,* **39,** 6016 (1989). Revision of the Douglas–Kroll Transformation.
27. C. M. Marian, *Berechnung von Matrixelementen des Spin-Bahn- und Spin–Spin-Kopplungsoperators mit MRD-CI-Wellenfunktionen,* Doctoral Thesis, University of Bonn, 1981.

28. B. A. Hess, *Matrixelemente zwischen Molekülwellenfunktionen zur Berechnung relativistischer Effekte,* Diploma Thesis, University of Bonn, 1977.

29. J. S. Cohen, W. R. Wadt, and P. J. Hay, *J. Chem. Phys.,* **71,** 2955 (1979). Spin–Orbit Coupling and Inelastic Transitions in Collisions of O ^1D with Ar, Kr, and Xe.

30. C. Heinemann, W. Koch, and H. Schwarz, *Chem. Phys. Lett.,* **245,** 509 (1995). An Approximate Method for Treating Spin–Orbit Effects in Platinum.

31. C. Ribbing, M. Odelius, A. Laaksonen, J. Kowalewski, and B. Roos, *Int. J. Quantum Chem., Quantum Chem. Symp.,* **24,** 295 (1990). Simple Non-Empirical Calulations of the Zero-Field Splitting in Transition Metal Systems. I. The Ni(II)-Water Complexes.

32. C. Ribbing, M. Odelius, and J. Kowalewski, *Mol. Phys.,* **74,** 1299 (1991). Simple Non-Empirical Calulations of the Zero-Field Splitting in Transition Metal Systems. II. NiF_6^{4-} and Ni(II) Complexes with Mixed Ligands.

33. M. Blume and R. E. Watson, *Proc. R. Soc. A,* **270,** 127 (1962). Theory of Spin–Orbit Coupling in Atoms. I. Derivation of the Spin–Orbit Coupling Constant.

34. M. Blume and R. E. Watson, *Proc. R. Soc. A,* **271,** 565 (1963). Theory of Spin–Orbit Coupling in Atoms. II. Comparison of Theory with Experiment.

35. T. R. Cundari, M. T. Benson, M. L. Lutz, and S. O. Sommerer, in *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1996, Vol. 8, pp. 145–202. Effective Core Potential Approaches to the Chemistry of the Heavier Elements.

36. G. Frenking, I. Antes, M. Boehme, S. Dapprich, A. W. Ehlers, V. Jonas, A. Neuhaus, M. Otto, R. Stegmann, A. Veldkamp, and S. F. Vyboishchikov, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1996, Vol. 8, pp. 63–143. Pseudopotential Calculations of Transition Metal Compounds: Scope and Limitations.

37. L. Seijo and Z. Barandiarán, in *Computational Chemistry: Reviews of Current Trends,* J. Leszcynski, Ed., World Scientific, Singapore, Vol. 4, 1999, pp. 55-152. The Ab Initio Model Potential Method: A Common Strategy for Effective Core Potential and Embedded Cluster Calculations.

38. M. Dolg, in *Modern Methods and Algorithms of Quantum Chemistry,* J. Grotendorst, Ed., NIC Series, Jülich, Germany, 2000, Vol. 1, pp. 479–501. Effective Core Potentials.

39. Y. S. Lee, W. C. Ermler, and K. S. Pitzer, *J. Chem. Phys.,* **67,** 5861 (1977). Ab Initio Effective Core Potentials Including Relativistic Effects. I. Formalism and Applications to the Xe and Au Atoms.

40. P. Hafner and W. H. E. Schwarz, *J. Phys. B,* **11,** 217 (1977). Pseudo-Potential Approach Including Relativistic Effects.

41. C. H. Teichteil, M. Pélissier, and F. Spiegelmann, *Chem. Phys.,* **81,** 273 (1983). Ab Initio Molecular Calculations Including Spin-Orbit Coupling. I. Method and Atomic Tests.

42. L. F. Pacios and P. A. Christiansen, *J. Chem. Phys.,* **82,** 2664 (1985). Ab Initio Relativistic Effective Potentials with Spin–Orbit Operators. I. Li through Ar.

43. M. M. Hurley, L. F. Pacios, P. A. Christiansen, R. B. Ross, and W. C. Ermler, *J. Chem. Phys.,* **84,** 6840 (1986). Ab Initio Relativistic Effective Potentials with Spin–Orbit Operators. II. K through Kr.

44. L. A. LaJohn, P. A. Christiansen, R. B. Ross, T. Atashroo, and W. C. Ermler, *J. Chem. Phys.,* **87,** 2812 (1988). Ab Initio Relativistic Effective Potentials with Spin–Orbit Operators. III. Rb through Xe.

45. R. B. Ross, J. M. Powers, T. Atashroo, W. C. Ermler, L. A. LaJohn, and P. A. Christiansen, *J. Chem. Phys.,* **93,** 6654 (1990). Ab Initio Relativistic Effective Potentials with Spin–Orbit Operators. IV. Cs through Rn.

46. R. B. Ross, S. Gayen, and W. C. Ermler, *J. Chem. Phys.,* **100,** 8145 (1994). Ab-Initio Relativistic Effective Potentials with Spin–Orbit Operators. V. Ce through Lu.

47. W. C. Ermler, R. B. Ross, and P. A. Christiansen, *Int. J. Quantum Chem.,* **40,** 829 (1991). Ab Initio Relativistic Effective Potentials with Spin–Orbit Operators. VI. Fr through Pu.

48. C. S. Nash, B. E. Bursten, and W. C. Ermler, *J. Chem. Phys.,* **106,** 5133 (1997). Ab Initio Relativistic Effective Potentials with Spin–Orbit Operators. VII. Am through Element 118.
49. S. A. Wildman, G. A. DiLabio, and P. A. Christiansen, *J. Chem. Phys.,* **107,** 9975 (1997). Accurate Relativistic Effective Potentials for the Sixth-Row Main Group Elements.
50. A. V. Titov and N. S. Mosyagin, *Int. J. Quantum Chem.,* **71,** 359 (1999). Generalized Relativistic Effective Core Potential: Theoretical Grounds.
51. W. C. Ermler, R. B. Ross, and P. A. Christiansen, *Adv. Quantum Chem.,* **19,** 139 (1988). Spin–Orbit Coupling and Other Relativistic Effects in Atoms and Molecules.
52. R. M. Pitzer and N. W. Winter, *J. Chem. Phys.,* **92,** 3061 (1988). Electronic-Structure Methods for Heavy-Atom Molecules.
53. M. Dolg, H. Stoll, and H. Preuss, *J. Chem. Phys.,* **90,** 1730 (1989). Energy-Adjusted Ab Initio Pseudopotentials for the Rare Earth Elements.
54. D. Andrae, U. Häussermann, M. Dolg, H. Stoll, and H. Preuss, *Theor. Chim. Acta,* **77,** 123 (1990). Energy-Adjusted Ab Initio Pseudopotentials for the Second and Third Row Transition Elements.
55. W. Küchle, M. Dolg, H. Stoll, and H. Preuss, *Mol. Phys.,* **74,** 1245 (1991). Ab Initio Pseudopotentials for Hg to Rn. I. Parameter Sets and Atomic Calculations.
56. M. Dolg and H. Stoll, in *Handbook on the Physics and Chemistry of Rare Earths,* K. A. Gscheidner Jr. and L. Eyring, Eds., Elsevier, Amsterdam, 1996, Vol. 22 pp. 607–729. Electronic Structure Calculations for Molecules Containing Lanthanide Atoms.
57. T. Leininger, A. Berning, A. Nicklass, H. Stoll, H.-J. Werner, and H.-J. Flad, *Chem. Phys.,* **217,** 19 (1997). Spin–Orbit Interaction in Heavy Group 13 Atoms and TlAr.
58. F. M. Dolg, *Quasirelativistische und relativistische energiekonsistente Pseudopotentiale für quantentheoretische Untersuchungen der Chemie schwerer Elemente,* Habilitation Thesis, Universität Stuttgart, 1997.
59. L. Seijo, *J. Chem. Phys.,* **102,** 8078 (1995). Relativistic Ab Initio Model Potential Calculations Including Spin–Orbit Effects through the Wood–Boring Hamiltonian.
60. B. A. Hess, C. M. Marian, U. Wahlgren, and O. Gropen, *Chem. Phys. Lett.,* **251,** 365 (1996). A Mean-Field Spin–Orbit Method Applicable to Correlated Wavefunctions.
61. D. Danovich, C. M. Marian, T. Neuheuser, S. D. Peyerimhoff, and S. Shaik, *J. Phys. Chem. A,* **102,** 5923 (1998). Spin–Orbit Coupling Patterns Induced by Twist and Pyramidalization Modes in C_2H_4: A Quantitative Study and a Qualitative Analysis.
62. J. Tatchen and C. M. Marian, *Chem. Phys. Lett.,* **313,** 351 (1999). On the Performance of Approximate Spin–Orbit Hamiltonians in Light Conjugated Molecules: The Fine–Structure Splitting of HC_6H^+, NC_5H^+, and NC_4N^+.
63. W. G. Richards, H. P. Trivedi, and D. L. Cooper, *Spin–Orbit Coupling in Molecules,* Clarendon Press, Oxford, UK, 1981.
64. B. Schimmelpfennig, *Atomic Spin–Orbit Mean-Field Integral Program AMFI*, developed at Stockholms Universitet, 1996.
65. C. M. Marian and U. Wahlgren, *Chem. Phys. Lett.,* **251,** 357 (1996). A New Mean-Field and ECP-Based Spin–Orbit Approach. Applications to Pt and PtH.
66. B. Schimmelpfennig, L. Maron, U. Wahlgren, C. Teichteil, H. Fagerli, and O. Gropen, *Chem. Phys. Lett.,* **286,** 261 (1998). On the Efficiency of an Effective Hamiltonian in Spin–Orbit CI Calculations.
67. B. Schimmelpfennig, L. Maron, U. Wahlgren, C. Teichteil, H. Fagerli, and O. Gropen, *Chem. Phys. Lett.,* **286,** 267 (1998). On the Combination of ECP-Based CI Calculations with All-Electron Spin–Orbit Mean-Field Integrals.
68. G. F. Koster, J. O. Dimmock, R. G. Wheeler, and H. Statz, *Properties of the Thirty-Two Point Groups,* MIT Press, Cambridge, MA, 1963.
69. G. Herzberg, *Molecular Spectra and Molecular Structure. III. Electronic Spectra and Electronic Structure of Polyatomic Molecules,* Van Nostrand, Princeton, NJ, 1966.

70. R. McWeeny, *J. Chem. Phys.,* **42,** 1717 (1965). On the Origin of Spin-Hamiltonian Parameters.
71. B. L. Silver, *Irreducible Tensor Methods*, Academic Press, New York, 1976.
72. M. Rotenberg, R. Bivins, N. Metropolis, and J. K Wooten, *The 3j and 6j Symbols,* MIT Press, Cambridge, MA, 1959.
73. W. Weltner Jr., *Magnetic Atoms and Molecules,* Van Nostrand Reinhold, New York, 1983.
74. R. N. Zare, *Angular Momentum: Understanding Spatial Aspects in Chemistry and Physics,* Wiley, New York, 1988.
75. E. P. Wigner, *Gruppentheorie*, Vieweg, Braunschweig, 1930.
76. C. Eckart, *Rev. Mod. Phys.,* **2,** 305 (1930). The Application of Group Theory to the Quantum Dynamics of Monoatomic Systems.
77. M. Weissbluth, *Atoms and Molecules,* Academic Press, New York, 1978.
78. C. C. J. Roothaan, *Int. J. Quantum Chem., Quantum Chem. Symp.,* **27,** 13 (1993). New Algorithms for Calculating $3n - j$ Symbols.
79. C. C. J. Roothaan, *Int. J. Quantum Chem.,* **63,** 57 (1997). Calculation of $3n - j$ Symbols by Labarthe's Method.
80. I. L. Cooper and J. I. Musher, *J. Chem. Phys.,* **57,** 1333 (1972). Evaluation of Matrix Elements of Spin-Dependent Operators for N-Electron System One-Body Operators.
81. H. M. Quiney, H. Skaane, and I. P. Grant, *Adv. Quantum Chem.,* **32,** 1 (1999). Ab Initio Relativistic Quantum Chemistry: Four-Components Good, Two-Components Bad!
82. L. Visscher, W. A. de Jong, O. Visser, P. J. C. Aerts, H. Merenga, and W. C. Nieuwpoort, in *MOTECC-95,* E. Clementi and G. Corongiu, Eds., STEF, Cagliari, Sardinia, 1995, pp. 219–241. Relativistic Quantum-Chemistry. The MOLFDIR Program Package.
83. C. M. Marian, *Chem. Phys. Lett.,* **173,** 175 (1990). On the Dependence of Correlation and Relativity: The Electron Affinity of the Copper Atom.
84. M. R. A. Blomberg and U. I. Wahlgren, *Chem. Phys. Lett.,* **145,** 393 (1988). On the Effect of Core Orbital Relaxation in First-Order Relativistic Calculations.
85. J. G. Snijders and P. Pyykkö, *Chem. Phys. Lett.,* **75,** 5 (1980). Is the Relativistic Contraction of Bond Lengths an Orbital-Contraction Effect?
86. C. M. Marian, *Theoretische Spektroskopie zweiatomiger Übergangsmetallverbindungen*, Habilitation Thesis, University of Bonn, 1991.
87. R. M. Pitzer, C. W. Kern, and W. N. Lipscomb, *J. Chem. Phys.,* **37,** 267 (1962). Evaluation of Molecular Integrals by Solid Spherical Harmonic Expansions.
88. T. E. H. Walker and W. G. Richards, *Symp. Faraday Soc.,* **2,** 64 (1968). Ab Initio Computatation of Spin–Orbit Coupling Constants in Diatomic Molecules.
89. R. L. Matcha, C. W. Kern, and D. M. Schrader, *J. Chem. Phys.,* **51,** 2152 (1969). Fine-Structure Studies of Diatomic Molecules: Two-Electron Spin–Spin and Spin–Orbit Integrals.
90. R. L. Matcha and C. W. Kern, *J. Chem. Phys.,* **55,** 469 (1971). Evaluation of Three- and Four-Center Integrals for Operators Appearing in the Breit–Pauli Hamiltonian.
91. C. W. Kern and M. Karplus, *J. Chem. Phys.,* **43,** 415 (1965). Gaussian-Transform Method for Molecular Integrals. II. Evaluation of Molecular Properties.
92. S. R. Langhoff and C. W. Kern, in *Modern Theoretical Chemistry*, H. F. Schaefer III, Ed., Plenum, New York, 1977, pp. 381–437. Molecular Fine Structure.
93. P. W. Abegg, *Mol. Phys.,* **30,** 579 (1975). Ab Initio Calculation of Spin–Orbit Coupling Constants for Gaussian Lobe and Gaussian-Type Wave Functions.
94. C. M. Marian, *Berechnung von Spin–Spin- und Spin–Bahn–Wechselwirkungsintegralen über Atomfunktionen und deren Anwendung,* Diploma Thesis, University of Bonn, 1977.
95. H. F. King and T. F. Furlani, *J. Comput. Chem.,* **9,** 771 (1985). Computation of One and Two Electron Spin–Orbit Integrals.

96. H. Ito and Y. J. I'Haya, *Mol. Phys.,* **24,** 1103 (1972). Evaluation of Molecular Spin–Orbit Integrals by a Gaussian Expansion Method.

97. O. Matsuoka, *Int. J. Quantum Chem.,* **7,** 365 (1973). Molecular Integrals of Relativistic Effects with Gaussian-Type Orbitals.

98. G. L. Bendazzoli and P. Palmieri, *Int. J. Quantum Chem.,* **8,** 941 (1974). Spin–Orbit Interaction in Polyatomic Molecules: Ab Initio Computations with Gaussian Orbitals.

99. J. Breulet, *J. Comput. Chem.,* **2,** 244 (1981). Ab Initio Calculation of Spin–Orbit Interaction in Polyatomic Molecules Using Gaussian-Type Wavefunctions.

100. P. Chandra and R. J. Buenker, *J. Chem. Phys.,* **79,** 358 (1983). Relativistic Integrals over Breit–Pauli Operators Using General Cartesian Gaussian Functions. I. One-Electron Interactions.

101. P. Chandra and R. J. Buenker, *J. Chem. Phys.,* **79,** 366 (1983). Relativistic Integrals over Breit–Pauli Operators Using General Cartesian Gaussian Functions. II. Two-Electron Interactions.

102. R. Samzow, *Die Zweielektronenterme des no-pair-Hamiltonoperators,* Doctoral Thesis, University of Bonn, 1991.

103. R. Samzow and B. A. Hess, *Chem. Phys. Lett.,* **184,** 491 (1991). Spin–Orbit Effects in the Br Atom in the Framework of the No-Pair Theory.

104. M. J. Bearpark, N. C. Handy, P. Palmieri, and R. Tarroni, *Mol. Phys.,* **80,** 479 (1993). Spin–Orbit Interactions from Self-Consistent-Field Wave-Functions.

105. M. Sjøvoll, H. Fagerli, O. Gropen, J. Almlöf, B. Schimmelpfennig, and U. Wahlgren, *Theor. Chem. Acc.,* **99,** 1 (1998). An Efficient Treatment of Kinematic Factors in Pseudo-Relativistic Calculations of Electronic Structure.

106. R. C. Raffenetti, *J. Chem. Phys.,* **58,** 4452 (1973). General Contraction of Gaussian Atomic Orbitals: Core, Valence, Polarization, and Diffuse Basis Sets; Molecular Integral Evaluation.

107. J. Almlöf and P. R. Taylor, *J. Chem. Phys.,* **86**(7), 4070 (1987). General Contraction of Gaussian Basis Sets. I. Atomic Natural Orbitals for First- and Second-Row Atoms.

108. R. J. Buenker, A. B. Alekseyev, H. P. Liebermann, R. Lingott, and G. Hirsch, *J. Chem. Phys.,* **108,** 3400 (1998). Comparison of Spin–Orbit Configuration Interaction Methods Employing Relativistic Effective Core Potentials for the Calculation of Zero-Field Splittings of Heavy Atoms with a $^2P^0$ Ground State.

109. E. R. Davidson, in *The World of Quantum Chemistry,* R. Daudel and B. Pullman, Eds., Reidel, Dordrecht, 1974, pp. 17–30. Configuration Interaction Description of Electron Correlation.

110. B. Huron, J. Malrieu, and P. Rancurel, *J. Chem. Phys.,* **58,** 5745 (1973). Iterative Perturbation Calculations of Ground and Excited Energies from Multiconfigurational Zeroth-Order Wavefunctions.

111. R. J. Buenker and S. D. Peyerimhoff, *Theor. Chim. Acta,* **35,** 33 (1974). Individualized Configuration Selection in CI Calculations with Subsequent Energy Extrapolation.

112. P. E. M. Siegbahn, *Int. J. Quantum Chem.,* **23,** 1869 (1983). The Externally Contracted CI Method Applied to N_2.

113. W. Meyer, in *Modern Theoretical Chemistry,* H. F. Schaefer III, Ed., Plenum, New York, 1977, pp. 413–446. Configuration Expansion by Means of Pseudonatural Orbitals.

114. H.-J. Werner and P. J. Knowles, *J. Chem. Phys.,* **89,** 5803 (1988). An Efficient Internally Contracted Multiconfigurational Interaction Method.

115. J. A. Pople, R. Seeger, and R. Krishnan, *Int. J. Quantum Chem., Quantum Chem. Symp.,* **11,** 149 (1977). Variational Configuration Interaction Methods and Comparison with Perturbation Theory.

116. C. M. Marian, *J. Chem. Phys.,* **93,** 1176 (1990). Quasirelativistic Calculation of the Vibronic Spectra of NiH and NiD.

117. D. R. Yarkony, *Int. Rev. Phys. Chem.,* **11,** 195 (1992). Spin-Forbidden Chemistry Within the Breit–Pauli Approximation.

118. H. Ågren, O. Vahtras, and B. Minaev, *Adv. Quantum Chem.,* **27,** 71 (1996). Response Theory and Calculations of Spin–Orbit Coupling Phenomena in Molecules.
119. B. A. Hess, R. J. Buenker, C. M. Marian, and S. D. Peyerimhoff, *Chem. Phys.,* **71,** 79 (1982). Ab-Initio Calculation of the Zero-Field Splittings of the $X^3\Sigma_g^-$ and $B^3\Pi_{g,i}$ States of the S_2 Molecule.
120. S. J. Havriliak and D. R. Yarkony, *J. Chem. Phys.,* **83,** 1168 (1985). On the Use of the Breit–Pauli Approximation for Evaluating Line Strengths for Spin-Forbidden Transitions: Application to NF.
121. W. Kutzelnigg, *Einführung in die Theoretische Chemie, Bd. 1, Quantenmechanische Grundlagen,* Verlag Chemie, Weinheim, 1974.
122. E. A. Hylleraas, *Z. Phys.,* **65,** 209 (1930). Über den Grundterm der Zweielektronenprobleme für H^-, He, Li^+, Be^{++} usw.
123. D. R. Yarkony, *J. Chem. Phys.,* **84,** 2075 (1986). On the Use of the Breit–Pauli Approximation for Evaluating Line Strengths for Spin-Forbidden Transitions. II. The Symbolic Matrix Element Method.
124. O. Christiansen, J. Gauss, and B. Schimmelpfennig, *Phys. Chem. Chem. Phys.,* **2,** 965 (2000). Spin–Orbit Coupling Constants from Coupled-Cluster Response Theory.
125. C. Chang, M. Pélissier, and P. Durand, *Phys. Scr.,* **34,** 394 (1986). Regular Two-Component Pauli-Like Effective Hamiltonians in Dirac Theory.
126. J. L. Heully, I. Lindgren, E. Lindroth, S. Lundquist, and A. M. Mårtensson-Pendrill, *J. Phys. B,* **19,** 2799 (1986). Diagonalisation of the Dirac Hamiltonian as a Basis for a Relativistic Many-Body Procedure.
127. E. van Lenthe, E. J. Baerends, and J. G. Snijders, *J. Chem. Phys.,* **99,** 4597 (1993). Relativistic Regular 2-Component Hamiltonians.
128. M. C. Kim, S. Y. Lee, and Y. S. Lee, *Chem. Phys. Lett.,* **253,** 216 (1996). Spin–Orbit Effects Calculated by a Configuration Interaction Method Using Determinants of Two-Component Molecular Spinors: Test Calculations on Rn and TlH.
129. H.-S. Lee, Y.-K. Han, M. C. Kim, C. Bae, and Y. S. Lee, *Chem. Phys. Lett.,* **293,** 97 (1998). Spin–Orbit Effects Calculated by Two-Component Coupled-Cluster Methods: Test Calculations on AuH, Au_2, TlH and Tl_2.
130. L. Visscher, O. Visser, P. J. C. Aerts, H. Merenga, and W. C. Nieuwpoort, *Comput. Phys. Commun.,* **81,** 120 (1994). Relativistic Quantum-Chemistry—the MOLFDIR Program Package.
131. N. Rösch, S. Krüger, M. Mayer, and V. A. Nasluzov, in *Recent Developments and Applications of Modern Density Functional Theory,* J. M. Seminario, Ed., Elsevier, Amsterdam, 1996, pp. 497–566. The Douglas–Kroll–Hess Approach to Relativistic Density Functional Theory: Methodological Aspects and Applications to Metal Complexes and Clusters.
132. E. van Lenthe, E. J. Baerends, and J. G. Snijders, *J. Chem. Phys.,* **101,** 9783 (1994). Relativistic Total-Energy Using Regular Approximations.
133. T. Nakajima, T. Suzumura, and K. Hirao, *Chem. Phys. Lett.,* **304,** 271 (1999). A New Relativistic Scheme in Dirac–Kohn–Sham Theory.
134. L. Visscher, *Chem. Phys. Lett.,* **253,** 20 (1996). On the Construction of Double Group Molecular Symmetry Functions.
135. E. R. Davidson, *J. Comput. Phys.,* **17,** 87 (1975). The Iterative Calculation of a Few of the Lowest Eigenvalues and Corresponding Eigenvectors of Large Real-Symmetric Matrices.
136. W. Butscher and W. Kammer, *J. Comput. Phys.,* **20,** 313 (1976). Modification of Davidson's Method for the Calculation of Eigenvalues and Eigenvectors of Large Real-Symmetric Matrices: "Root Homing Procedure".
137. B. Liu, in *Numerical Algorithms in Chemistry: Algebraic Methods,* C. Moler and I. Shavitt, Eds., Lawrence Berkeley Laboratory, CA, 1978, Vol. LBL-8158, pp. 49–53. The Simultaneous Expansion Method for the Iterative Solution of Several of the Lowest Eigenvalues and Corresponding Eigenvectors of Large Real-Symmetric Matrices.

138. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran 77: The Art of Scientific Computing*, Cambridge University Press, Cambridge, UK, 1992.

139. H. A. Kramers, *Z. Phys.*, **53**, 422 (1929). Zur Aufspaltung von Multiplett S-Termen in zweiatomigen Molekülen. I.

140. H. A. Kramers, *Z. Phys.*, **53**, 429 (1929). Zur Aufspaltung von Multiplett S-Termen in zweiatomigen Molekülen. II.

141. S. J. Hutter, *Methodologische Aspekte der Behandlung der Spin-Bahn–Kopplung zweiatomiger Moleküle*, Doctoral Thesis, University of Bonn, 1994.

142. M. Esser, *Int. J. Quantum Chem.*, **304**, 313 (1984). Direct MRCI Method for the Calculation of Relativistic Many-Electron Wavefunctions. I. General Formalism.

143. B. O. Roos and P. E. M. Siegbahn, in *Modern Theoretical Chemistry*, H. F. Schaefer III, Ed., Plenum, New York, 1977, pp. 277–318. The Direct Configuration Interaction Method from Molecular Integrals.

144. J. Paldus, *Int. J. Quantum Chem., Quantum Chem. Symp.*, **9**, 165 (1975). A Pattern Calculus for the Unitary Group Approach to the Electronic Correlation Problem.

145. I. Shavitt, *Int. J. Quantum Chem., Quantum Chem. Symp.*, **11**, 131 (1977). Graph Theoretical Concepts for the Unitary Group Approach to the Many-Electron Correlation Problem.

146. B. Liu and M. Yoshimine, *J. Chem. Phys.*, **74**, 612 (1981). The Alchemy Configuration Interaction Method. I. The Symbolic Matrix Element Method for Determining Elements of Matrix Operators.

147. M. Sjøvoll, O. Gropen, and J. Olsen, *Theor. Chem. Acc.*, **97**, 301 (1997). A Determinantal Approach to Spin–Orbit Configuration Interaction.

148. S. Yabushita, Z. Zhang, and R. M. Pitzer, *J. Phys. Chem. A* , **103**, 5791 (1999). Spin–Orbit Configuration Interaction Using the Graphical Unitary Group Approach and Relativistic Core Potential and Spin–Orbit Operators.

149. T. Fleig, J. Olsen, and C. M. Marian, *J. Chem. Phys.*, **114**, 4775 (2000). The Generalized Active Space (GAS) Concept for the Relativistic Treatment of Electron Correlation. I. Kramers Restricted Two-Component CI.

150. A. Berning, H.-J. Werner, M. Schweizer, P. J. Knowles, and P. Palmieri, *Mol. Phys.*, **98**, 823 (2000). Spin–Orbit Matrix Elements for Internally Contracted Multireference Configuration Interaction Wavefunctions.

151. M. Kleinschmidt, C. M. Marian, M. Waletzke, and S. Grimme, to be published.

152. G. A. DiLabio and P. A. Christiansen, *Chem. Phys. Lett.*, **277**, 473 (1997). Low-Lying O^+ States of Bismuth Hydride.

153. BNSOC is a spin–orbit coupling program package developed at the University of Bonn, Germany, by B. A. Hess and C. M. Marian, with contributions from P. Chandra, S. Hutter, M. Kleinschmidt, F. Rakowitz, R. Samzow, B. Schimmelpfennig, and J. Tatchen, 1999.

154. F. Rakowitz and C. M. Marian, *Chem. Phys.*, **225**, 223 (1997). An Extrapolation Scheme for Spin–Orbit Configuration Interaction Energies Applied to the Ground and Excited Electronic States of Thallium Hydride.

155. F. Rakowitz, *Entwicklung, Implementierung und Anwendung effizienter Methoden in der relativistischen Elektronenstrukturtheorie*, Doctoral Thesis, University of Bonn, 1999. URL www.thch.uni-bonn.de/tc/.

156. R. Llusar, M. Casarrubios, Z. Barandiarán, and L. Seijo, *J. Chem. Phys.*, **105**, 5321 (1996). Ab Initio Model Potential Calculations on the Electronic Spectrum of Ni^{2+}-Doped MgO Including Correlation, Spin–Orbit and Embedding Effects.

157. K. Balasubramanian, *J. Chem. Phys.*, **89**, 5731 (1988). Relativistic Configuration Interaction Calculations for Polyatomics: Applications to PbH_2, SnH_2, and GeH_2.

158. G. A. DiLabio and P. A. Christiansen, *J. Chem. Phys.*, **108**, 7527 (1998). Separability of Spin–Orbit and Correlation Energies for the Sixth-Row Main Group Hydride Ground States.

159. F. Rakowitz, M. Casarrubios, L. Seijo, and C. M. Marian, *J. Chem. Phys.,* **108,** 7980 (1998). Ab Initio Spin-Free-State-Shifted Spin–Orbit Configuration Interaction Calculations on Singly Ionized Iridium.

160. V. Vallet, L. Maron, C. Teichteil, and J.-P. Flament, *J. Chem. Phys.,* **113,** 1391 (2000). A Two-Step Uncontracted Determinantal Effective Hamiltonian-Based SO–CI Method.

161. H. Lefebvre-Brion and R. W. Field, *Perturbations in the Spectra of Diatomic Molecules,* Academic Press, Orlando, 1986.

162. R. W. Field, *Ber. Bunsenges. Phys. Chem.,* **86,** 771 (1982). Diatomic Molecule Electronic Structure Beyond Simple Molecular Constants.

163. C. M. Marian, *Ber. Bunsenges. Phys. Chem.,* **99,** 254 (1995). An Approach to the Calculation of Ω-Splittings in Diatomic Molecules with Strongly Coupled Electronic States and Its Application to NiH and NiD.

164. M. Perić, B. Engels, and S. D. Peyerimhoff, in *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy,* S. R. Langhoff, Ed., Kluwer, Dordrecht, 1995, pp. 261–356. Theoretical Spectroscopy on Small Molecules: Ab Initio Investigations of Vibronic Structure, Spin–Orbit Splittings, and Magnetic Hyperfine Effects in the Electronic Spectra of Triatomic Molecules.

165. L. Engelbrecht and J. Hinze, *Adv. Chem. Phys.,* **44,** 1 (1980). Molecular Properties Observed and Computed.

166. A. Carrington, D. H. Levy, and T. A. Miller, *Adv. Chem. Phys.,* **18,** 149 (1970). Electron Resonance of Gaseous Diatomic Molecules.

167. A. Carrington, *Microwave Spectroscopy of Free Radicals,* Academic Press, London, UK, 1974.

168. C. E. Moore, *Atomic Energy Levels, NBS Circular 467,* National Bureau Standard, Washington, DC, 1949.

169. C. M. Marian, *J. Chem. Phys.,* **94,** 5574 (1991). Theoretical Study of the Spectra of CuH and CuD.

170. S. Hutter, B. A. Hess, C. M. Marian, and R. Samzow, *J. Chem. Phys.,* **100,** 5617 (1994). Theoretical Description of the $b^3\Pi$ to $a^3\Sigma^+$ Transition of NO^+.

171. M. Kleinschmidt, *Drehimpulskopplungen in CH/CD und PbH. Eine Studie mit Allelektronen-ab-initio-Methoden,* Diploma Thesis, University of Bonn, 1999. URL www.thch.uni-bonn.de/tc/.

172. M. Kleinschmidt, T. Fleig, and C. M. Marian, *J. Mol. Spectrosc.,* submitted for publication. Kramers-Type Splitting in the $X^2\Pi$ and $a^4\Sigma^-$ States of CH and CD Calculated in a Hund's Case (a) Basis.

173. H. Hettema and D. Yarkony, *J. Chem. Phys.,* **100,** 8991 (1994). On the Radiative Lifetime of the $(a^4\Sigma^-, v, N, F_i)$ Levels of the CH Radical: An Ab Initio Treatment.

174. A. Maciejewski and R. P. Steer, *Chem. Rev.,* **93,** 67 (1993). The Photophysics, Physical Photochemistry, and Related Spectroscopy of Thiocarbonyls.

175. O. Vahtras, H. Ågren, P. Jørgensen, H. J. A. Jensen, T. Helgaker, and J. Olsen, *J. Chem. Phys.,* **96,** 2118 (1992). Spin–Orbit Coupling Constants in a Multiconfiguration Linear Response Approach.

176. L. L. Lohr Jr., *J. Chem. Phys.,* **45,** 1362 (1966). Spin-Forbidden Electric–Dipole Transition Moments.

177. L. Goodman and B. J. Laurenzi, *Adv. Quantum Chem.,* **4,** 153 (1968). Probability of Singlet–Triplet Transitions.

178. P. Avouris, W. M. Gelbert, and M. A. El-Sayed, *Chem. Rev.,* **77,** 793 (1977). Nonradiative Electronic Relaxation Under Collision-Free Conditions.

179. S. C. J. Meskers, T. Polonski, and H. P. J. M. Dekkers, *J. Phys. Chem.,* **99,** 1134 (1995). Polarized Absorption and Phosphorescence Spectra and Magnetic Circular Dichroism of Dithioimides: Assignment of the Lower $^1n\pi^*$ and $^3n\pi^*$ States.

180. J. Tatchen, *Anwendung von Mean-Field Operatoren zur Beschreibung von Spin-Bahn–Effekten in organischen Molekülen,* Diploma Thesis, University of Bonn, 1999. URL www.thch.uni-bonn.de/tc/.

181. J. Tatchen, C. M. Marian, M. Waletzke, and S. Grimme, to be published. A Quantum Chemical Study of Spin-Forbidden Transitions in Dithiosuccinimide.

182. B. Minaev, *Int. J. Quantum Chem.*, **17,** 367 (1980). Intensities of Spin-Forbidden Transitions in Molecular Oxygen and Selective Heavy-Atom Effects.

183. R. Klotz, C. M. Marian, S. D. Peyerimhoff, B. A. Hess, and R. J. Buenker, *Chem. Phys.,* **89,** 223 (1984). Calculation of Spin-Forbidden Radiative Transitions Using Correlated Wave Functions: Lifetimes of $b^1\Sigma^+$, $a^1\Delta$ States in O_2, S_2 and SO.

184. H. Köppel and W. Domcke, in *Encyclopedia of Computational Chemistry,* P. v. R. Schleyer, Ed., Wiley, Chichester, 1997, pp. 152–278. Vibronic Dynamics of Polyatomic Molecules.

185. B. Gemein and S. D. Peyerimhoff, *Chem. Phys. Lett.,* **184,** 45 (1991). Theoretical Study of the Vibronic Interactions in the Ground and First Excited $a^3\Pi$ and $a'^3\Sigma^+$ States of the CO Molecule.

186. J. R. Taylor, *Scattering Theory: The Quantum Theory on Nonrelativistic Collisions,* Wiley, New York, 1972.

187. G. Wentzel, *Z. Phys.,* **43,** 524 (1927). Über strahlungslose Quantensprünge.

188. G. Wentzel, *Phys. Z.,* **29,** 321 (1928). Die unperiodischen Vorgänge in der Wellenmechanik.

189. O. K. Rice, *Phys. Rev.,* **33,** 748 (1929). Perturbations in Molecules and the Theory of Predissociation and Diffuse Spectra.

190. O. K. Rice, *Phys. Rev.,* **35,** 1551 (1930). Perturbations in Molecules and the Theory of Predissociation and Diffuse Spectra. II.

191. J. N. Murrell and J. M. Taylor, *Mol. Phys.,* **16,** 609 (1969). Predissociation in Diatomic Spectra with Special Reference to the Schumann–Runge Bands of O_2.

192. C. M. Marian, R. Marian, S. D. Peyerimhoff, B. A. Hess, R. J. Buenker, and G. Seger, *Mol. Phys.,* **46,** 779 (1982). Ab-Initio Study of O_2^+ Predissociation Phenomena Induced by a Spin–Orbit Coupling Mechanism.

193. D. Edvardsson, S. Lunell, F. Rakowitz, C. M. Marian, and L. Karlsson, *Chem. Phys.,* **229,** 203 (1998). Calculation of Predissociation Rates in O_2^{2+} by Ab Initio MRD–CI Methods.

194. J. Schön and H. Köppel, *J. Chem. Phys.,* **108,** 1503 (1998). Geometric Phases and Quantum Dynamics in Spin–Orbit Coupled Systems.

195. J. Römelt and R. Runau, *Theor. Chim. Acta,* **54,** 171 (1980). Franck–Condon Matrix Elements for Bound-Continuum Vibrational Transitions Calculated by Numerical Integration and Basis Set Expansion Techniques.

196. D. O. Harris, G. G. Engerholm and W. D. Gwinn, *J. Chem. Phys.,* **43,** 1515 (1965). Calculation of Matrix Elements for One-Dimensional Quantum-Mechanical Problems and the Application to Anharmonic Oscillators.

197. Note Added in Proofs: For the evaluation of spin–orbit matrix elements, standard basis sets of at least double-zeta plus polarization quality should be employed. See, e.g., R. Klotz, C. M. Marian, S. D. Peyerimhoff, B. A. Hess, and R. J. Buenker, *Chem. Phys.*, **76,** 367 (1983). Study of the Dependence of Spin–Orbit Matrix Elements on AO Basis Set Composition for Inner and Valence Shells: Results for the Multiplet Spiltting of $X^3\sum^-$ and $C^3\Pi$ of SO and $X^2\Pi$ in SO^+. A proper representation of the inner and outer nodes of the valence shell orbitals is essential. The representation of the inner core orbitals was found to be less critical. Scalar relativistic all-electron calculations require a modification of standard Gaussian basis set contractions. For elements H–Kr, scalar relativistic recontractions are available from http://www.emsl.pnl.gov:2080/forms/basisform.html. For the Douglas–Kroll recontracted basis sets, see: W. A. de Jong, R. J. Harrison, and D. A. Dixon, *J. Chem. Phys.*, **114,** 48 (2001). Parallel Douglas–Kroll Energy and Gradients in NWChem: Estimating Scalar Relativistic Effects Using Douglas–Kroll Contracted Basis Sets.

Reviews in Computational Chemistry, Volume 17. Edited by Kenny B. Lipkowitz, Donald B. Boyd

ISBNs: 0-471-39845-4 (Hardcover); 0-471-22441-3 (Electronic)

CHAPTER 4

Cellular Automata Models of Aqueous Solution Systems

Lemont B. Kier,* Chao-Kun Cheng,[†] and Paul G. Seybold[‡]

**Department of Medicinal Chemistry, Virginia Commonwealth University, Richmond, Virginia 23298, [†]Department of Mathematical Sciences, Virginia Commonwealth University, Richmond, Virginia 23298, and [‡]Department of Chemistry, Wright State University, Dayton, Ohio 45435*

INTRODUCTION

For the most part, experimental observations made by chemists—of phase changes, physical and chemical properties, chemical reactions, and so on—are averaged outcomes resulting from the behaviors of a very large number of interacting particles, such as molecules, ions, or rare gas atoms. To make sense of these observations, two general approaches have been pursued.

In one approach, the focus is on the individual particles or molecules themselves, with the hope that from an understanding of these basic units in isolation one might be able to infer their behavior en masse. Aided by the rapid development of both chemical theory and its principal tool, the modern digital computer, this approach has been highly successful in many instances. Quantum chemical calculations of molecular structures and electronic distributions have greatly enhanced knowledge of the physical and chemical properties of substances, and studies of reaction transition states have significantly expanded our understanding of a wide variety of chemical reactions. Yet there

are clear limits to this approach, especially when one seeks to explore many of the less detailed, broader aspects of the observations, such as the variations in species populations with time and the statistical and kinetic details of the phenomenon in question.

In the second general approach to this problem, an attempt is made to examine in some manner the overall behavior of the entire ensemble of interacting units. By far, the most common approach here, and the one normally taught from textbooks, is to represent the kinetic behavior of a particular system in terms of an applicable set of coupled differential rate equations. These equations, with their associated rate constants, summarize the bulk behaviors of the ingredients involved in an averaged way. For example, the simple two-step transformation A → B → C, can be characterized by the set of rate equations:

$$d[\mathrm{A}]/d\mathrm{t} = -k_1[\mathrm{A}] \quad [1]$$

$$d[\mathrm{B}]/d\mathrm{t} = k_1[\mathrm{A}] - k_2[\mathrm{B}] \quad [2]$$

$$d[\mathrm{C}]/d\mathrm{t} = k_2[\mathrm{B}] \quad [3]$$

where k_1 and k_2 are the rate constants for the two steps. This formulation can be referred to as the "traditional" approach. Solution of these equations yields the time-dependent populations of the species involved. Although analytical solutions may be impossible to obtain for some very complicated systems of equations, powerful approximate numerical methods make such exact, analytical solutions unnecessary. There are, however, several important—and often unrecognized—limitations of this traditional approach, in addition to its occasional cumbersomeness. The first limitation, which is more a formal than a practical problem, is that despite its widespread employment and obvious utility, such an approach has no solid theoretical justification; its use rests on empiricism.[1] A more practical limitation is that a deterministic approach of this nature is truly applicable only at the limit in which an exceedingly large number of interacting particles is involved. Furthermore, this approach gives no information on the nature of the inherent, and often important, fluctuations that can be expected to arise in the behaviors of real, finite systems. Since recent advances in analytical technology have pushed levels of experimental observations to very small samples and even to the single molecule level, the ability to deal with finite samples has become increasingly important.

Stochastic analysis presents an alternative avenue for dealing with the inherently probabilistic and discontinuous microscopic events that underlie macroscopic phenomena. Many processes of chemical and physical interest can be described as random Markov processes.[1,2] Unfortunately, solution of a stochastic master equation can present an extremely difficult mathematical challenge for systems of even modest complexity. In response to this difficulty, Gillespie[3–5] developed an approach employing numerical Monte Carlo

techniques that accurately describes the stochastic features of such processes, and several applications have demonstrated the usefulness of this approach.[6,7]

There is an intermediate level of consideration of chemical phenomena laying somewhere between the intense scrutiny of a single molecule and the averaged, possibly superficial, treatment of a bulk sample with an effectively infinite population. This intermediate level is the level of chemical events for which the phenomena of interest can be modeled or simulated with acceptable accuracy by including a finite number of molecules (greater than one but far less than Avogadro's number). For example, the behavior of a solution cannot be modeled accurately with just one or two or even a few of molecules. Ideally, however, if we could examine or model a few tens, a hundred, or a thousand ingredients, then some significant insight into the nature of the solution might be achieved. This intermediate level model might, then, be reasonably simpler to create and still give meaningful insights into the phenomena of interest.

The past three decades have witnessed the development of three broad techniques—molecular dynamics (MD), Monte Carlo (MC), and cellular automata simulations—that approach the study of molecular systems by simulating submicroscopic chemical events at this intermediate level. All three methods focus attention on a modest number of molecules and portray chemical phenomena as being dependent on dynamic, and interactive events (a portrayal consistent with our scientific intuition and a characteristic not intrinsic to either thermodynamics or the traditional deterministic approach based on differential equations). These techniques lend themselves to a visual portrayal of the evolution of the configurations of the systems under study. Because each approach has its own particular advantages and shortcomings, one must take into consideration the pros and cons of each, especially in light of the nature of the problem to be solved.

Both MD and MC methods require either quantum mechanical or force field calculations to determine the energies. Normally, for large systems one must begin with a choice of a force field or potential functions (namely, a set of equations representing the electrostatic, steric, and other interactions that can be expected to apply within the system studied). Regardless of how these energies are computed, in an MD calculation Newton's laws of motion are applied in a linearized form using a very short (e.g., 0.5 femtosecond) time step. Positions of the atoms are determined at each time step, and the evolution of the system is followed over the time period of the simulation. In an MC calculation, small random changes are made in the positions of the particles in the system. The configurations produced in this way are evaluated and weighted according to their energies as determined by a suitable quantum method or by the potential function chosen for the system. The probability of a change from one configuration to another can thus be determined; some changes can be accepted, and others rejected. Sampling and evaluation of the normally very large number (e.g., 10^7) of configurations produced in this way allows

estimation of the thermodynamic properties of the system studied. More detailed descriptions of these techniques can be found in recent reviews.[8–11]

A question arises as to the relationship of cellular automata models to the MC and MD models. Put briefly, cellular automata models are usually much simpler both in concept and implementation than are MD or MC models. In results and procedures, cellular automata have much in common with the integrated MC technique.[11] All three techniques provide visual expressions of the processes involved as well as information on fluctuations in the system's behavior. The cellular automata models, however, omit the detailed molecular structures and energy calculating methods that are commonly employed in MC and MD simulations, substituting in their place heuristic rules for these features. As a result, cellular automata models are far less demanding computationally and much faster than either of the other techniques, but they do not directly reveal information about the energetics of the process or system. However, the use of rules in place of forces does have the special advantage of providing specific features that can be varied individually, and the consequences of those changes can be queried independent of other variations that are intimately coupled to them. For example, in a cellular automaton model one can independently vary the temperature of a solvent while keeping the temperature of its solute constant in order to assess the effect of environmental temperature, or one can increase one specific reaction or transition rate while holding others constant so as to isolate the influence of a particular process on the overall dynamics of a system.

In the following sections, we describe what cellular automata are, how they work, and then we provide examples of their uses. Because much of chemistry is carried out in condensed phases, especially in water, we limit our applications to this medium but point out that extensions to other types of systems are possible.

CELLULAR AUTOMATA

Historical Background

Cellular automata were first proposed by the mathematician Stanislaw Ulam and the mathematical physicist John von Neumann a half century ago,[12–14] although related ideas were put forth even earlier, in the 1940s, by the German engineer Konrad Zuse.[15–17] von Neumann's interest was in the construction of "self-reproducing automata".[18] His original idea was to construct a series of mechanical devices or "automata" that would gather and integrate the ingredients that could reproduce themselves. A suggestion by Ulam[13] led von Neumann to consider grids with moving ingredients, operating with rules. The first such system was made up of square cells in a matrix, each with a state, operating with a set of rules in a two-dimensional grid. The system existed with up to 29 different states. With the development of modern

digital computers, it became increasingly clear that these fairly abstract ideas could in fact be usefully applied to the examination of real physical systems.[19,20] As described by Wolfram,[21] cellular automata have five fundamental defining characteristics:

They consist of a discrete lattice of cells.

They evolve in discrete time steps.

Each site takes on a finite number of possible values.

The value of each site evolves according to the same deterministic rules.

The rules for the evolution of a site depend only on a local neighborhood of sites around it.

As we shall see, the fourth characteristic can be modified to include probabilistic rules as well as deterministic rules. An important feature sometimes observed in the evolution of these computational systems was the development of unanticipated patterns of ordered dynamical behavior, or "emergent properties". As Kauffman has expressed it,[22] "Studies of large, randomly assembled cellular automata ... have now demonstrated that such systems can spontaneously crystallize enormously ordered dynamical behavior. This crystallization hints that hitherto unexpected principles of order may be found, [and] that the order observed may have significant explanatory import in [biology and physics]." This proposal has borne considerable fruit, not only in biology and physics, but also in chemistry. Readers are referred to reviews for applications in physics[23–26] and biology;[27] selected physical and chemical applications have been reviewed by Chopard and co-workers.[28]

The General Structure

The simulation of a dynamic system using cellular automata requires several parts that make up the process. The cell is the basic model of each ingredient, molecule, or whatever constitutes the system. These cells may have several shapes as part of the matrix or grid of cells. The grid may have boundaries or be part of a topological object that eliminates boundaries. The cells may have rules that apply to all of the edges, or there may be different rules for each edge. This latter plan may impart more detail to the model, as needed for a more detailed study.

The Cells

Cellular automata have been designed for one (1D), two (2D), or three-dimensional (3D) arrays. The most commonly used is the 2D grid. The cells may be triangles, squares, hexagons, or other shapes in the 2D grid. The square cell has been the one most widely used over the past 40 years. Each cell in the grid is endowed with a primary state, namely, whether it is empty or occupied with a particle, object, molecule, or whatever the system requires

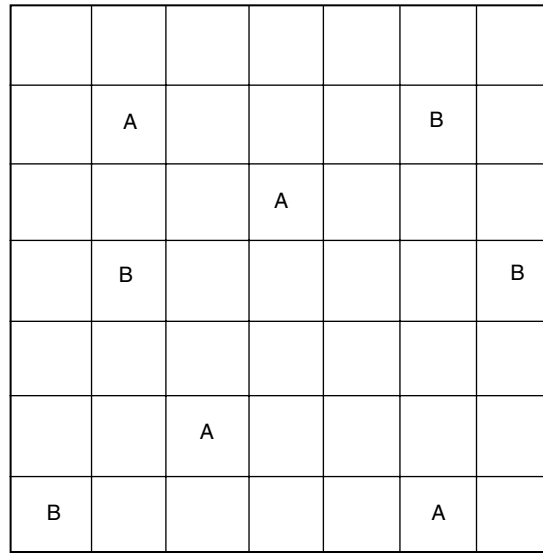

Figure 1 A cellular automata grid showing occupied cells of different states A and B, and unoccupied cells (blank).

to study the dynamic event (see Figure 1). Secondary information is contained in the state description that encodes the differences among cell occupants in a study.

The Cell Shape

The choice of the cell shape is based on the objective of the study. In studies of water and solution phenomena, a square cell is appropriate because the water molecule is quadravalent to hydrogen bonding to other water molecules or solutes. A water molecule donates two hydrogens and two lone-pair electrons in forming the tetrahedral structure that characterizes the liquid state. The four faces of a square cell thus correspond to the bonding opportunities of a water molecule.

Grid Boundary Cells

The ''moving cell'' may encounter an edge or boundary during its movements. The boundary cell may be treated as any other occupied cell, following rules that permit joining or breaking. A more common practice is to assume that the grid is simulating a small segment of a large dynamic system. In this model, the boundaries do come into play in the results. The grid is then considered to be the surface of a torus (Figure 2); the planar projection of this surface would reveal the movement of a cell off of the edge and reappearing from the opposite edge onto the grid (see Figure 3). In some cases, it is necessary to establish a vertical relationship among occupants, which establishes a gravity

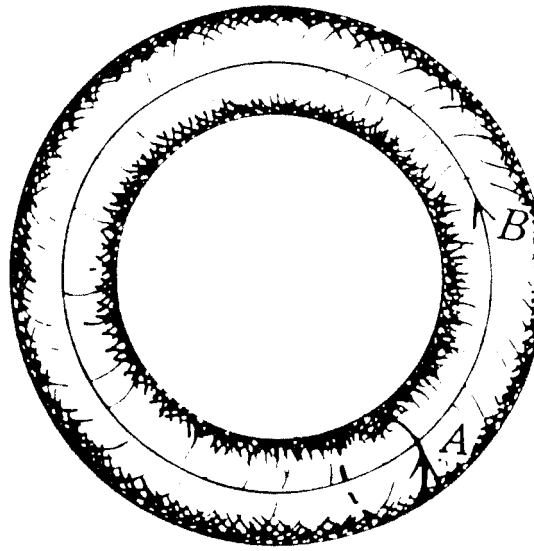

Figure 2 The grid located on the surface of a torus. Column A of the grid is represented by line A on the torus; row B on the grid is represented by line B on the torus.

effect. In this situation, the grid is on the surface of a cylinder with a boundary condition at the top and bottom, which is an impenetrable boundary.

Variegated Cell Types

Until recently, all of the cellular automata models assumed that each edge of a cell bore the same state and movement rules. Recent work has

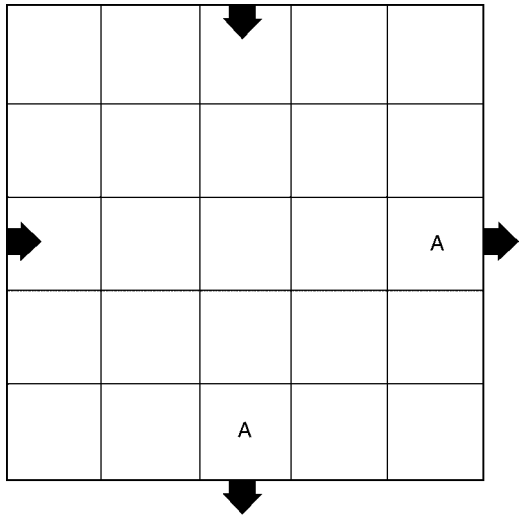

Figure 3 Movement of cell occupants at the boundaries of the grid on a torus. The cell occupant A may move right and appear at the opposite side of the grid. The other cell occupant A may move down and appear at the top of the grid.

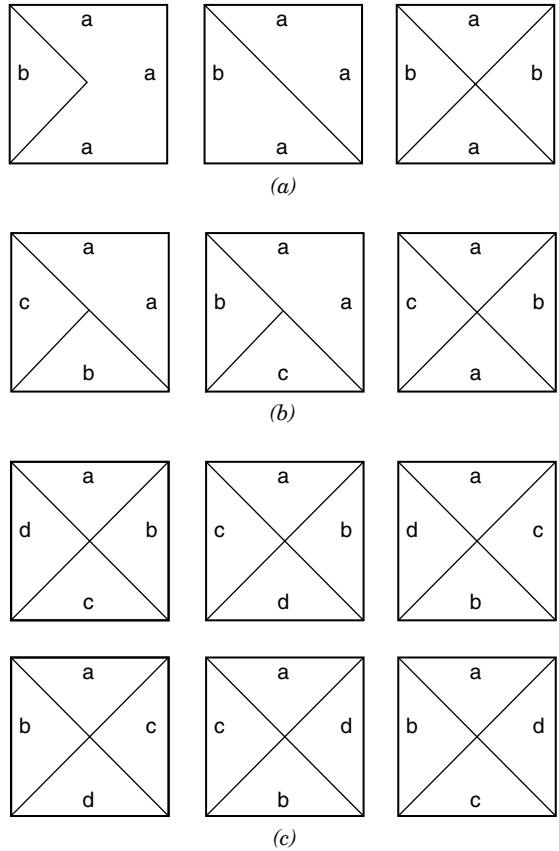

Figure 4 The possible variegated cells. (**a**) Possible occupants with two different types of edges allowing for different rules for the edges. (**b**) Possible occupants with three different types of edges hence possibly different rules for the edges. (**c**) Possible occupants with two different types of edges hence different rules for the edges.

employed a variegated cell where each edge may have its own state and set of movement rules. Examples of some variegated cells are shown in Figure 4.

Cell Movement

The dynamic character of cellular automata is developed by the simulation of movement of the cells, which may be a simultaneous process or each cell, in turn, may execute a movement. Each cell's movement is based on rules derived from the states of other nearby cells. These nearby cells constitute a neighborhood. The rules may be deterministic or stochastic, the latter process being driven by probabilities of certain events occurring. These processes are elaborated upon in this section.

Synchronous–Asynchronous Movement

When we speak of movement of a cell or the movement of cell occupants, we are speaking of the simulation of a movement from one cell to another. Thus a water molecule or some object is postulated to move across space, appearing in a new location at time $t + 1$. In the cellular automata models, the actual situation is the exchange of state between two adjacent cells. If we are modeling the movement of a molecule from place A to the adjacent place B, then we must exchange the states of cells A and B. Initially, at time t, cell A has a state corresponding to an occupant molecule, while adjacent cell B is devoid of a molecule (i.e., it is empty). At time $t + 1$, the states of the cells A and B have exchanged. Cell A is empty, and cell B has the state of the occupant molecule. This exchange gives the illusion, and the practical consequences, of a movement from A to B. We speak of the movement of cells or of the movement of cell occupants; either way we are describing the process of simulating a movement as stated above.

Cell movement of all occupants in the grid may occur simultaneously (*synchronous*) or it may occur sequentially (*asynchronous*). When all cells in the grid have computed their state and have executed their movement (or not) it is called one *iteration*, a unit of time in the cellular automata simulation. Each cell is identified in the program and is selected randomly for the choice of movement or not. The question of which type of movement to use depends on the system being modeled and the information sought from the model, but it should reflect reality. If the system being studied is a slow process, then synchronous motion may best represent the process. In contrast, if the system is very fast, like proton hopping among water molecules where the cellular automata is using a few thousand cells, then an asynchronous model is desirable. A synchronous execution of the movement rules leads to possible competition for a cell from more than one occupant. A resolution scheme must thus be in place to resolve the competition; otherwise this may interfere with the validity of the model.

Neighborhoods

Cell movement is governed by rules called *transition functions*. The rules involve the immediate environment of the cell called the *neighborhood*. The most common neighborhood used in 2D cellular automata is called the *von Neumann neighborhood* after the pioneer of the method. A cell, i, is in the center of four cells, j, adjoining the four faces of i, [see Figure 5(**a**) for this pattern]. Another common neighborhood is the *Moore neighborhood* [Figure 5(**b**)], where cell i is completely surrounded by j cells. Other neighborhoods include the extended von Neumann neighborhood shown in Figure 5(**c**), where the k cells beyond j are identified and allowed to participate in movements.

Deterministic/Probabilistic Movement Rules

The rules governing cell movement may be deterministic or probabilistic. Deterministic cellular automata use a fixed set of rules, the values of which are

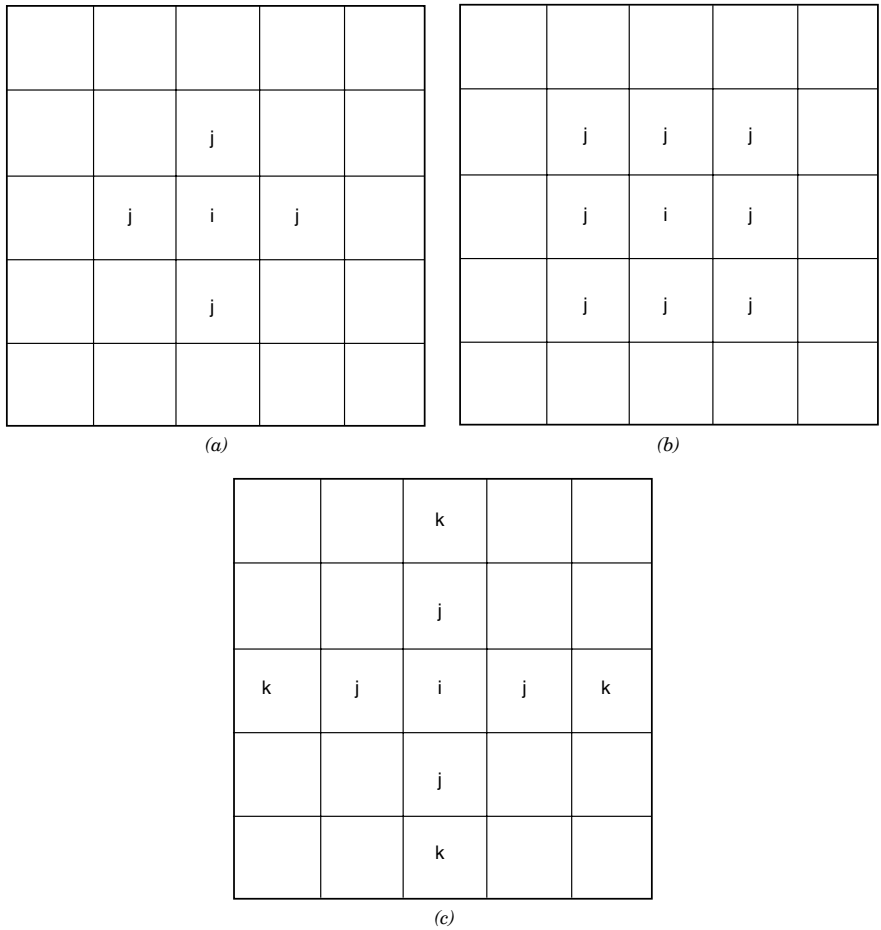

Figure 5 The cellular automata neighborhoods associated with cell *i*. (**a**) The von Neumann neighborhood. (**b**) The Moore neighborhood. (**c**) The extended von Neumann neighborhood.

constant and uniformly applied to the cells of each type. In probabilistic cellular automata, the movement of *i* is based on a probability-chosen rule where a certain probability to move or not to move is established for each type of *i* cell at its turn. Its state (empty or occupied) is determined, then its attribute as an occupant is determined. The probability of movement is next determined by a random number selection between two predefined limits. As an example, the random choice limits can be 0 to 1000, and a choice of numbers between 0 and 200 is designated as a move rule, whereas the remaining set, 201–1000, is a no-move rule. This case represents a probabilistic rule of 20% movement. Each cell then chooses a random number and behaves according to the rule corresponding to that numerical value.

Movement (Transition) Rules

The movement of cells is based upon rules governing the events inherent in cellular automata dynamics. The rules describe the probabilities of two adjacent cells separating, two cells joining at a face, two cells displacing each other in a gravity simulation, or a cell with different designated edges rotating in the grid. These events are the essence of the cellular automata dynamics and produce configurations that may possibly mirror physical events. The following sections develop these topics, and the Appendix provides technical detail about the hardware and software used and defines the directional moving probability, $MP(d)$, for each direction d.

General Conditions for Water Models

Several probabilistic rules are used to govern the trajectories of molecules moving in the grid. The first set of these rules, called transition functions, govern the movement of the occupied cells across the grid. The rules respond to the immediate neighborhood of the occupied cell, thus all events are local. A probabilistic or stochastic set of rules operating in a grid on the surface of a torus is most appropriate for the simulation of water and solution phenomena. This arrangement provides a near-infinitely large system with no boundary conditions.

The Free Movement Probability

The first rule is the movement probability P_m. This rule involves the probability that an occupant in an unbound cell, i, will move to one of four adjacent cells, j, if that space is unoccupied. An example is cell i in Figure 6 that may move (in its turn) to any unoccupied cell, j. If it moves to a cell whose neighbor is an occupied cell, k, then a bond will form between cells i and k. As

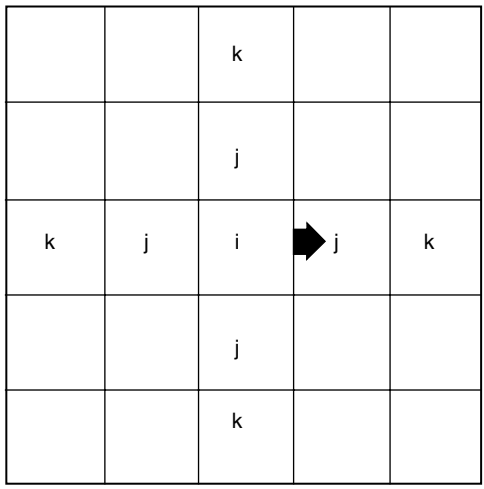

Figure 6 Possible movement of cell i occupant to unoccupied cell j.

a matter of course, this probability, P_m, is usually set at 1.0, which means that this event always happens (a rule).

Breaking Rules

Just as two cells can join together, so their tessellated state can be broken. The first of two trajectory or interaction rules is the breaking probability P_B. To comprehend this, it is convenient to refer to the occupant of a cell as a molecule. The P_B(AB) rule is the probability for a molecule A, bonded to molecule B, to break away from B, as shown in Figure 7(**a**). The value for P_B lies between 0 and 1. If molecule A is bonded to two molecules, B and B, the simultaneous probability of a breaking away event from both B and B is $P_B(AB)^*P_B(AB)$, as shown in Figure 7(**b**). If molecule A is bound to three other molecules, B, B, and B, the simultaneous breaking probability of molecule A is $P_B(AB)^*P_B(AB)^*(P_B(AB)$, shown in Figure 7(**c**). If molecule A is surrounded by four molecules, it cannot move on this 2D grid.

Joining Parameter

A joining trajectory parameter, J(AB), describes the movement of a molecule at A to join with a molecule at B, when an intermediate cell is vacant (see Figure 8). This rule follows the rule to move or not to move described above. The parameter J is a nonnegative real number. When $J = 1$, molecule A has the same probability of movement toward or away from B as for the case when the B cell is empty. When $J > 1$, molecule A has a greater probability of movement toward an occupied cell B than when cell B is empty. When $0 < J < 1$, molecule A has a lower probability of such movement. When $J = 0$, molecule A cannot make any movement.

Relative Gravity

The simulation of a "gravity" effect has been introduced into the cellular automata paradigm to model separating phenomena such as the demixing of immiscible liquids or the flow of solutions in a chromatographic separation. To accomplish this effect, a boundary condition is imposed at the upper and lower edges of the grid to simulate a vertical versus a horizontal relationship. The differential effect of gravity is simulated by introducing two new rules governing the preferences of two cells of different composition to exchange positions when they are in a vertically joined state. When molecule A is on top of a second molecule B, then two new rules are actuated. The first rule, G_D(AB), is the probability that molecule A will exchange places with molecule B, assuming a position below B. The other gravity term is G_D(BA), which expresses the probability that molecule B will occupy a position beneath molecule A. These rules are illustrated in Figure 9. In the absence of any strong evidence to support the model that the two gravity rules are complementary to each other in general, the treatment described above reflects the situation in which the gravity effects of A and B are two separate random events. Based

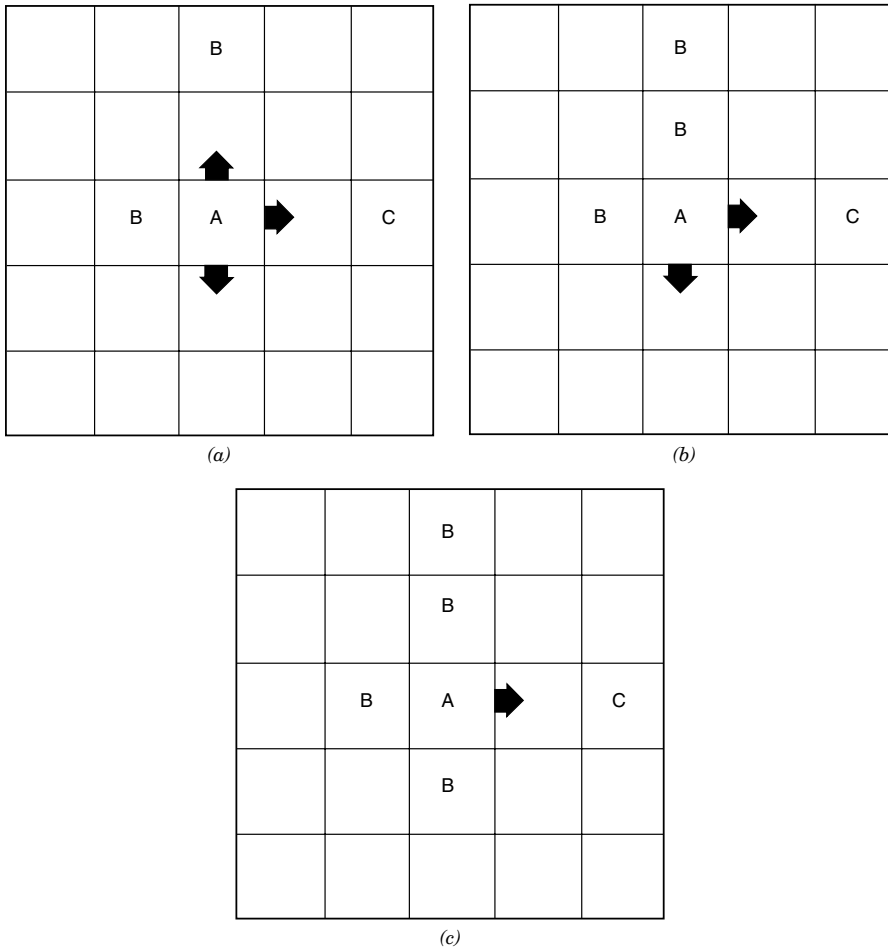

Figure 7 Illustration of the breaking parameter P_B. (**a**) The breaking away of cell occupant A from cell B and movement to an unoccupied cell with probabilities: north 0.266, east 0.266, south 0.266, and west 0.000. (**b**) Occupant A breaking away from cell B and moving to an unoccupied cell with probabilities: north 0.000, east 0.320, south 0.320, and west 0.000. (**c**) Occupant A breaking away from cell B and moving to an unoccupied cell B with probabilities: north 0.000, east 0.512, south 0.000, and west 0.000. The parameters used are $P_B(AB) = 0.8,\ J(AB) = J(AC) = 1.0,\ absG(A) = 0,$ and $G_D(AB) = 0$.

on an assumption of complementarity, the equality $G_D(AB) = 1 - G_D(BA)$ may be employed in the gravity simulation. Once again, the rules are probabilities of an event occurring. The choice of actuating this event made by each cell, in turn, is based on a random number selection within the boundaries used for that particular event. This process was mentioned in the earlier section on grid boundary cells.

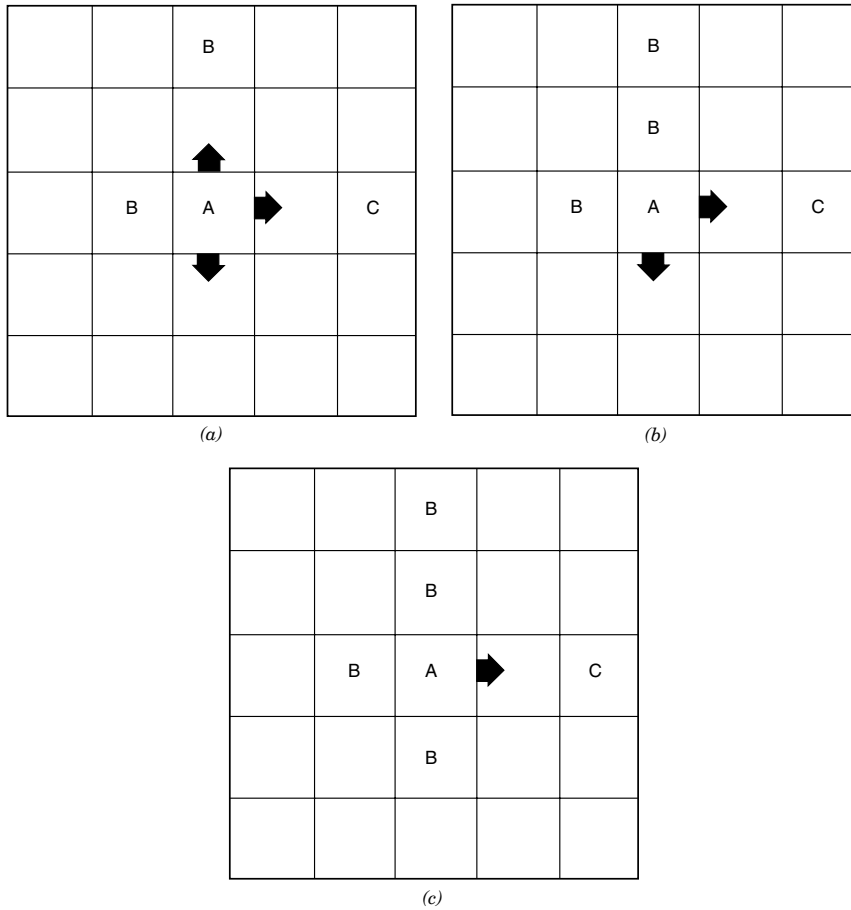

Figure 8 Illustration of the joining parameter J. (**a**) Movement of cell occupant A in three directions with probabilities: north 0.153, east 0.421, south 0.266, and west 0.000. (**b**) Movement of A in two directions with probabilities: north 0.000, east 0.484, south 0.266, and west 0.000. (**c**) Movement of A in one direction with probabilities: north 0.000, east 0.677, south 0.000, and west 0.000. The parameters used are $P_B(AB) = 0.8,\ J(AB) = 0.5,\ J(AC) = 2.0,\ absG(A) = 0,$ and $G_D(AB) = 0$.

Absolute Gravity

The absolute gravity measure of a molecule A, denoted by $absG(A)$, is a nonnegative number. It is the adjustment needed in the computation to determine movement if A is to have a bias to move down (or up). This movement is shown in Figure 10.

Cell Rotation

In cases where an unsymmetrical cell is used (see Figure 4, for examples), it is necessary to insure there exists a uniform representation of all possible

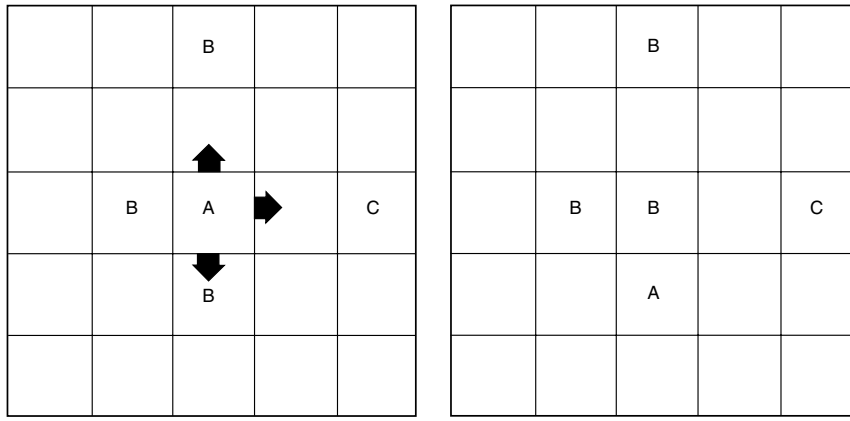

Figure 9 Illustration of the relative gravity parameter. (**a**) Occupant A may move to an empty cell with probabilities: north 0.174, east 0.445, south 0.379, or west 0.000, or it may also exchange positions with cell B with probability 0.379. (**b**) The grid configuration after cell A exchanges with cell B. The parameters used are $P_B(AB) = 0.8$, $J(AB) = 0.5$, $J(AB) = 2.0$, $absG(A) = 0$, and $G_D(AB) = 1.5$.

rotational states of that cell in the grid. To accomplish this, the unsymmetrical cells are randomly rotated 90° every iteration after beginning the run. This process is illustrated for four iterations in Figure 11.

Collection of Data

A cellular automata simulation of a dynamic system provides two classes of information. The first, a visual display, may be very informative of the char-

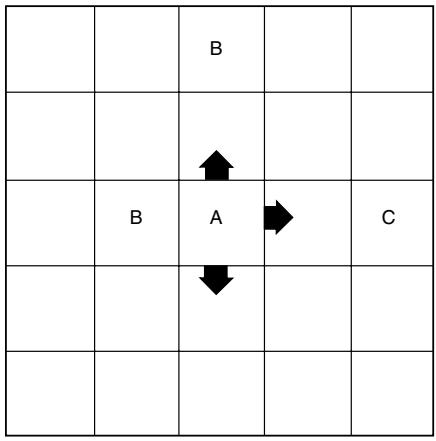

Figure 10 Illustration of the absolute gravity parameter. Occupant A may move to an empty cell with probabilities: north 0.153, east 0.421, south 0.421, or west 0.000. The parameters used are $P_B(AB) = 0.8$, $J(AB) = 0.5$, $J(AC) = 2.0$, $absG(A) = 0$, and $G_D(AB) = 0$.

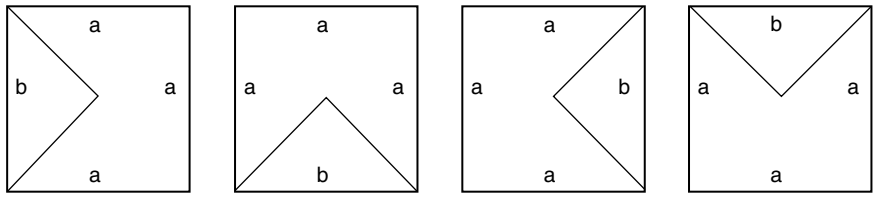

Figure 11 Illustration of rotation of variegated cells each iteration.

acter of a system as it develops from initial conditions. This visualization can be a dramatic portrayal of a process that opens the door to greater understanding of systems in the minds of students. The second source of information is the count of cells in different configurations such as those in isolation or those that are joined to another cell. This pattern of cells is called the *configuration* and is a rich source of information from which understanding of a process and the prediction of unforeseen events may be derived. These are illustrated in the following sections.

Number of Runs

It is customary to collect data from several runs, averaging the counts over those runs. The number of iterations performed depends on the system under study. The data collection may be over several iterations following the achievement of a stable or equilibrium condition. This stability is reckoned as a series of values that exhibit a relatively constant average value over a number of iterations. In other words, there is no trend observed toward a higher or lower average value.

Types of Data Usually Collected

From typical simulations used in the study of aqueous systems, several attributes are customarily recorded and used in comparative studies with properties. These attributes used singly or in sets are useful for analyses of different phenomena. Examples of the use and significance of these attributes will be offered in a later part of this chapter. The designations are

f_0 fraction of cells not bound to other cells.

f_1 fraction of cells bound to only one other cell.

f_2 fraction of cells bound to two other cells.

f_3 fraction of cells bound to three other cells.

f_4 fraction of cells bound to four other cells.

f_H free hydrogen fraction (in water) $= \frac{1}{2}$ of fraction of unbound cell faces.

N_{HB} number of hydrogen bonds per water cell = average count of joined faces per cell.

In addition, the average distance that a cell travels may be another datum collected.

AQUEOUS SOLUTION SYSTEMS

Water as a System

It has long been recognized that water plays an essential part in chemical events especially those associated with drug phenomena including drug–receptor encounters. Water is viewed as an active participant in the complex system composed of drug–receptor–water. Along with these encounters, there are a host of pharmacodynamic events occurring in aqueous solution including drug absorption, distribution, metabolism, and elimination (ADME). It is essential then to add to the understanding of this essential fluid and the complex systems it forms when solutes enter its embrace.

When solution ingredients interact, there emerges in the system a set of properties not clearly recognizable as additive contributions from the ingredients. There is formed a complex system characterized by new properties. The subjects of complexity and emergent properties in drug research have been reviewed by Kier and Testa.[29] The complex nature of water and solutions is recognized and has prompted some investigators to derive models based on nonlinear combinations of ingredients. In particular, we have witnessed the growth of MC and MD simulations of water that have added to our understanding of its complex character.[30–34] However, the large amount of computer time required by MD and MC coupled with assumptions of specific force fields produces certain limitations. An alternative to these two simulation methods that reduces some of these problems is the cellular automata model of dynamic synthesis.

A prominent model of water is that of an extended network of hydrogen-bonded molecules that lack a single, identifiable, long-lived structure.[35] The hydrogen bonds continually form and break producing a constantly changing mosaic when viewed at the molecular system level. This model lends itself to simulation using dynamic methods such as cellular automata. Kier and Cheng[36] created such a model of liquid water using rules governing the joining and breaking of water-designated cells. This model provided the computational basis for further studies of aqueous solution phenomena, described in this part of the chapter.

The Molecular Model

We want to make clear just what the cells, the configurations generated, and the cellular automata models represent. This effort will help with understanding the results of a simulation and limiting misunderstanding based on

direct comparisons with molecular methods. A cell with a state value encoding occupation by a particular object is not a model of a molecule with specified electronic and steric features. These attributes are considered to be subsumed into the rules. In studies on water and solution phenomena, reference to a cell with a designated state and trajectory rules is made as a "molecule" for convenience.

In the models used in studies of aqueous phenomena, the trajectory of a liquid water molecule is assumed to follow the connections of the hexagonal ice lattice (Figure 12). Each vertex in that figure denotes a water molecule, and each edge denotes a bonding relationship. This 3D network can be dissected into a contiguous series of vertices arranged tetrahedrally around a central molecule (Figure 13). Some or all of the vertices in each fragment may be representative of a water molecule. The trace of each fragment may be mapped onto a 2D grid [Figure 5(**a**)], which is equated with the mapping of a cellular automaton von Neumann neighborhood. The cellular automata transition functions (i.e., the rules) operate randomly and asynchronously on the central cell *i* in each von Neumann neighborhood. Consequently, the new configuration for each cell *i* and its neighborhood is derived independently of all other cells outside of this neighborhood. The configuration of the system achieved after all cells respond in random order to the rules constitutes one iteration. This configuration is a composite of the collective configurations achieved in all of the von Neumann neighborhoods. Each of these neighborhoods is a 2D mapping of a tetrahedral fragment of the original 3D model. The model is a representation of the configuration of a 3D system on the basis of it being an

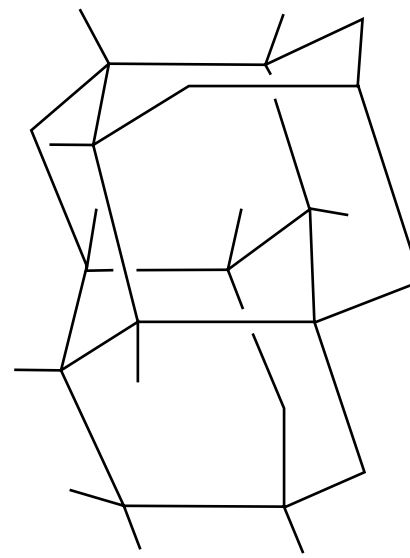

Figure 12 The hexagonal pattern of water molecules in a cluster, used as a model for the trajectory of a single water molecule.

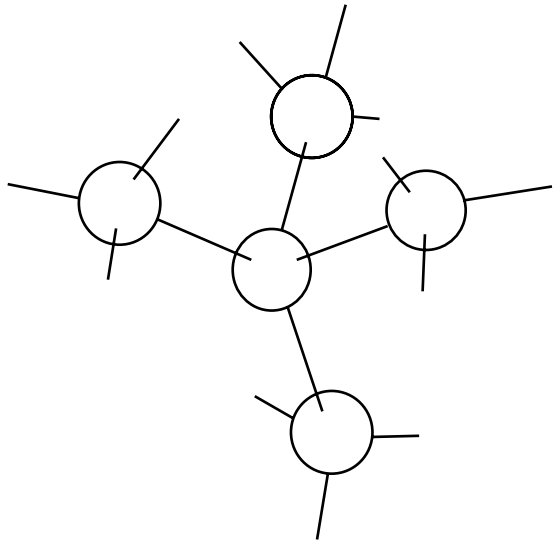

Figure 13 The tetrahedral pattern of a fragment of five water molecules in a cluster.

ensemble of discrete, orthogonal events occurring within that system. Other studies using this approximation have been reported.[37,38]

Significance of the Rules

The breaking and joining rules described above have a physical parallel in studies of water and solution phenomena. The breaking probability, $P_B(W)$, governs the self-affinity of a water molecule, W. This probability has a relationship to the boiling point, described by the equation:

$$P_B(W) = 0.01T_B(^\circ C) \quad [4]$$

The companion rule, $J(W)$, reflects the affinity of two W molecules. For water, it is observed that the joining and breaking parameters are related as

$$\log J(W) = -1.50P_B(W) + 0.60 \quad [5]$$

These rules produce conditions in the cellular automata configuration that simulate physical properties.

The relationship of a water molecule, W, and a solute molecule, S, is governed by the parameters $P_B(WS)$ and $J(WS)$. Higher values of $P_B(WS)$ and lower values of $J(WS)$ reflect a weak interaction between water and the solute. The opposite pattern of parameters characterize a closer physical

relationship between water and solute. These patterns ultimately translate into simulations of solubility, the hydrophobic effect, and diffusion preferences.

STUDIES OF WATER AND SOLUTION PHENOMENA

A Cellular Automata Model of Water

In a study of water, Kier and Cheng[36] found relationships between the average cluster size and viscosity, and between the fraction of unbound cells, f_0, and the vapor pressure. The number of water-designated cells was chosen to be 69% of the cells in the grid, this value giving a concentration of unbound water molecules in agreement with earlier predictions and experiments.[30,39] More recent studies have examined all the f_x attributes from the rules P_B and J, whereby the P_B(W) value was systematically varied. A profile of water cells, described by the fraction of each bonding type, f_0–f_4, for each "temperature", is shown in Figure 14. The values of f_0 correspond closely to the values used by Walrafen[39] and Haggis et al.[40] to predict the dielectric constant and heat capacity of water. This finding led to the conjecture that the rules chosen give rise to configurations mirroring physical reality. A test of this possibility is reported in Table 1 where a set of equations relating several physical properties of water to attributes from the cellular automata simulations are shown. In addition to the examples in the table, the free hydrogen fraction, f_H, is close to the value of about 15% reported from infrared (IR) evidence of the free OH content.[41] These correlations between simulated water attributes and various physical properties suggest that the water model created with cellular automata has validity.[36]

The Hydrophobic Effect

The *hydrophobic effect* is a term describing the influence of relatively nonpolar (lipophilic) substances on the collective behavior of water molecules in their vicinity. The common expression is that water is "more structured" or organized when in contact with a lipophilic solute. This behavior was observed in a cellular automata model of a solute in water,[42,43] which led to a study in more detail.[44] The hydrophobic effect was modeled by systematically increasing the breaking probability, P_B(WS), value, encoding an increasing probability of a solute molecule, S, not to associate with water.

It was observed that low P_B(WS) values, modeling a polar molecule, produced configurations in which the solute molecules were extensively surrounded by water molecules, a pattern simulating hydration or electrostriction. Conversely, with high values of P_B(WS) most of the solute molecules were found outside of the water clusters and within the cavities. This configuration leaves the water clusters relatively free of solute; hence they are more

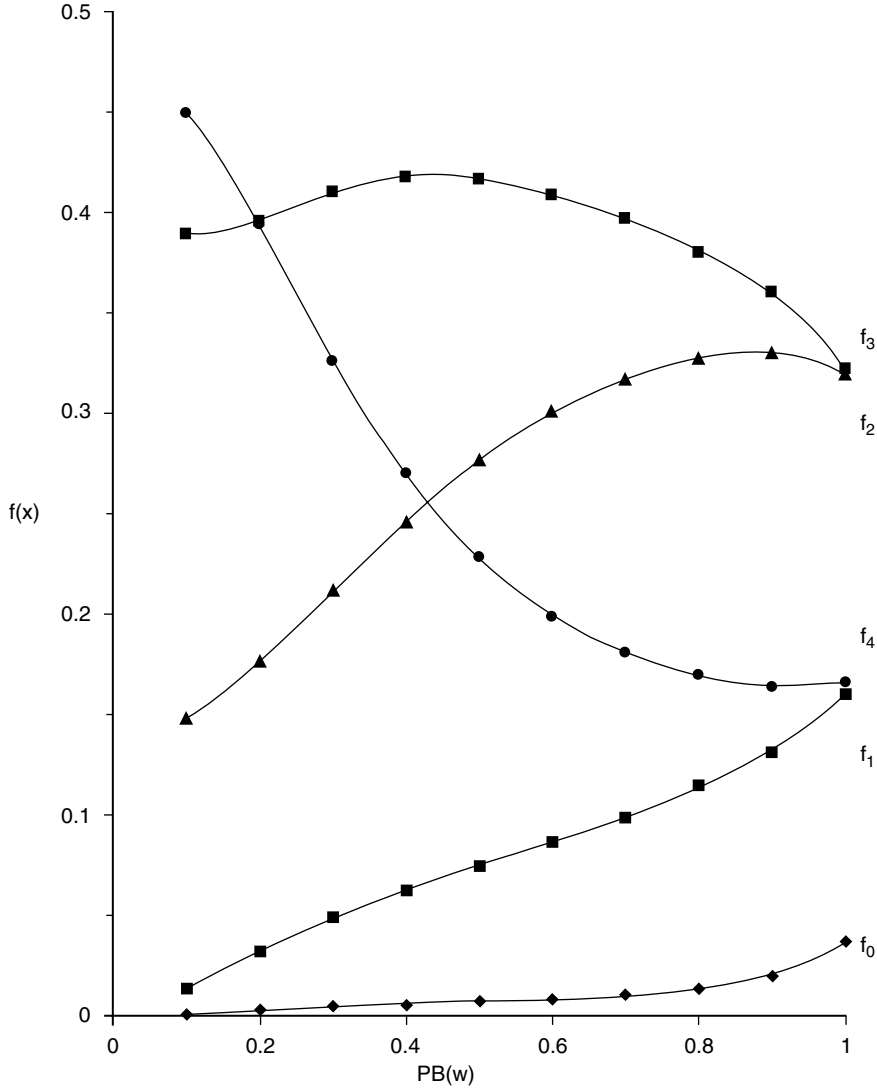

Figure 14 The fractions of water molecules in various levels of aggregation as a function of the P_B(W) value that correlates with the water temperature (Eq. [5]).

structured or organized. This arrangement is a model of the molecular level condition present in the hydrophobic effect. A typical configuration from this study is shown in Figure 15. These results agree with MD simulations,[31,45] and the interpretation agrees with evidence and models proposed for the hydrophobic effect.[46–48] The relationship between the P_B(WS) rules and the

Table 1 Water Properties Related to Cellular Automata Attributes

Property	Equation	r^2 Correlation
Vapor pressure	$\log P_v(\text{mm Hg}) = 13.77(f_0 + f_1) + 0.795$	0.987
Dielectric constant	$\varepsilon = -224f_1 + 86.9$	0.989
Viscosity	$\eta(\text{centipoise}) = 3.165f_4 - 0.187$	0.989
Ionization	$-\log K_w = -20.94f_H + 16.43$	0.999
Surface tension	$\gamma(\text{dyn/cm}) = 16.07N_{HB} + 22.35$	0.970
Compressibility	$\kappa(\times 10^6/\text{bar}) = -53.82f_3 + 66.66$	0.953

relative lipophilicity of a solute molecule is a useful rule for further studies on solution phenomena.

Solute Dissolution

Cellular automata simulations of the dissolution process have been described.[42] The solute molecules, S, started in a solid block of cells at the center of the grid. They are endowed with rules $P_B(S)$, $J(S)$, $P_B(WS)$, and $J(WS)$. The attributes recorded from the dynamics were the $f_0(S)$, fraction of solutes unbound to other solute molecules, plus the average number of solute–solute joined faces, $T(S)$, and the average distance that solute molecules have traveled from the center of the block at some specific iteration, $D(S)$. The $f_0(S)$ values were interpreted to represent the extent of dissolution of the solute. The decrease in the $T(S)$ values characterize the extent of disruption of the solute block, whereas the $D(S)$ values quantify the extent of diffusion of solutes into the surrounding water.

The effects of the four joining and breaking parameters, $P_B(WS)$, $J(WS)$, $P_B(S)$, and $J(S)$, were studied using a high and a low numerical value for each. From the results several conclusions were drawn about the influence of the rules and about some aspects of the process. The extent and rate of the solute cluster disruption, which is reflected by $T(S)$, is primarily a function of the $P_B(S)$ rule with secondary influence from the $P_B(WS)$ rule. A high value of $P_B(S)$ is a rule governing a high probability of solute molecules separating from each other; hence the rule may be viewed as the effective melting process relating to crystal disruption. The dissolved state was considered to exist when an isolated solute cell is present. The f_0 fraction is a measure of the extent of the modeled solubility. The extent of dissolution depends on both the $P_B(S)$ and the $P_B(WS)$ rules. High values of $P_B(S)$ and low values of $P_B(WS)$ promote an extensive degree of dissolution. A low value of $P_B(WS)$ characterizes a solute that is relatively polar and hydrophilic. The studies also showed that solutes with high $P_B(WS)$ values diffuse more rapidly than those with low values. The interpretation is that the more lipophilic the solute molecule, the greater its diffusion rate. It was also observed that simulations of higher temperatures led to faster disruption of the block of solute, more f_0 molecular

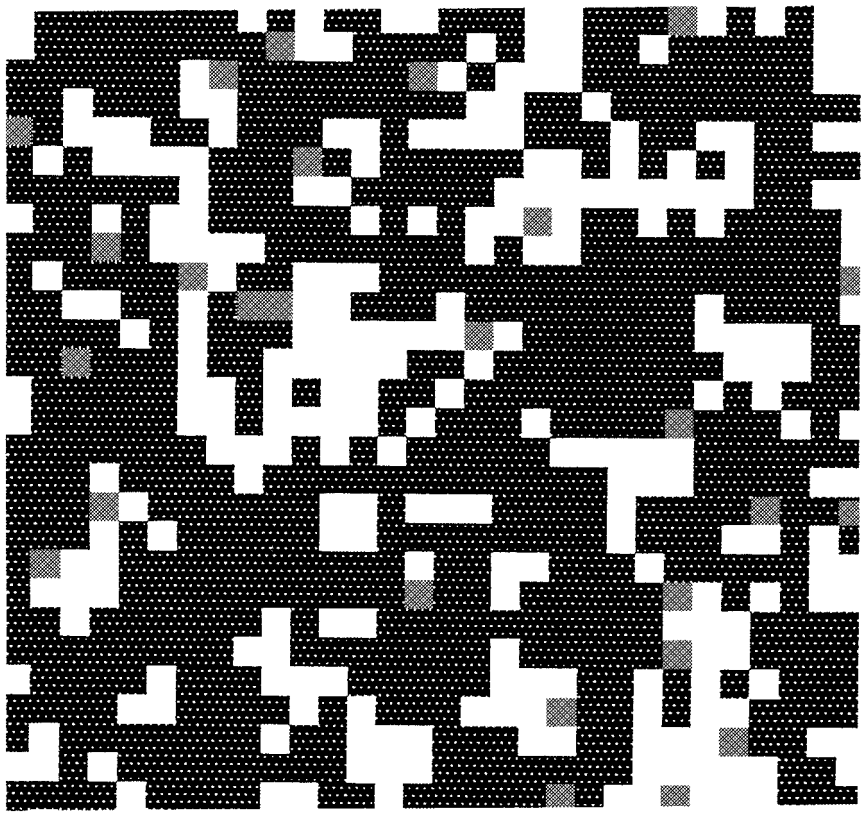

Figure 15 A model of the hydrophobic effect. The dark cells represent water, the white cells are cavities, and the gray cells are hydrophobic solute molecules.

configurations simulating greater dissolution, and more extensive diffusion of these molecules through the bulk water. This profile is the expected pattern of effects characteristic of increased temperature.

An interesting and unexpected observation arises from the graphical display of the disintegration of the solute block. The early stages of the disruption occurred as a series of intrusions of cavities rather than water molecules into the block. The cavities roamed throughout the block, behaving as "particles". The entrance of water into the block structure appeared much later, after there had occurred significant disruption and loss of solute molecules from the block. An image of this behavior is depicted in Figure 16. This observation was found for all parameter sets used in this study. The cellular automata model of the dissolution process is in substantial agreement with experimental evidence. This study produced some results suggesting ideas concerning a familiar process.

Figure 16 A model of the disruption of a crystal in water. The dark cells represent water, the white cells are cavities, and the gray cells are solute molecules.

Aqueous Diffusion

A series of studies have been reported modeling the diffusion process in water.[49] Using the rules previously defined, we examined several characteristics of a system to determine their influence on diffusion. The first study revealed that solutes of high lipophilicity (low polarity) diffuse faster than those of low lipophilicity (high polarity). This result is not commonly considered or reported. Diffusion studies are numerous in the literature, but comparisons with solute polarity are very scarce. Two such studies, however, support the cellular automata model of this phenomenon.[50,51]

A model of relative diffusion rates as a function of water temperature produced the expected result of greater diffusion with higher temperature. Additional studies[49] were conducted in an attempt to model the influence of solution characteristics on solute diffusion. A series of dilute solutions were

modeled whereby the relative lipophilicity of a solute S_1 was varied. A second solute, S_2, was introduced at the center of the grid. The dynamics showed that the solute S_2 diffused faster when there is no cosolute and when the polarity of S_2 is low. The presence of the cosolute revealed that the diffusion of S_2 was fastest when the cosolvent S_1 was lipophilic.

In another study,[49] the grid was divided into two halves, the upper half, containing a solution of a relatively nonpolar solute, while the lower half contained a solution of a relatively polar solute. Between these two halves, a thin layer of cells simulated a solution made up of a solute of intermediate polarity. The dynamics created a model in which the solute in the middle layer preferentially diffused into the upper layer containing the nonpolar solute. A second study simulated an aqueous solution of intermediate polarity in the middle of the grid. Surrounding this solution were regions with highly polar, intermediately polar, and nonpolar solutes in solution. A fourth region around the center was pure water. The solute in the central section of the grid was allowed to diffuse freely. It diffused more rapidly into the pure water; however, it secondarily preferred the more nonpolar solution quadrant. This finding is consistent with the results from the studies described above. The diffusion of a solute is modeled to be faster when it is lipophilic and when cosolutes are lipophilic.

Immiscible Liquids and Partitioning

Cheng and Kier modeled the separation of immiscible liquids and the partitioning of a solute between them.[52,53] For this study, the cellular automata grid had to be modified to create a model of the differential effect on the ingredients due to "gravity". To accomplish this, boundary conditions were imposed at the upper and lower edges of the grid to contain the liquids in a closed system. The differential effect of gravity was modeled by introducing the rules described earlier.

The water cells were ruled to have high self-affinities and low affinities with the other solvent. A $P_B(WS_1)$ rule was chosen with a high value to endow the two liquids with markedly different polarities. The emerging configuration can thus be described as being an immiscible, two-phase system. Each experiment began with a random distribution of equal numbers of water and S_1 cells, the sum totaling 69% of the grid space. The breaking and joining parameters were chosen to create significant self-affinity between each type of ingredient and a weak affinity between different types. The gravity terms were chosen to favor the water cell moving to the lower position relative to the other solvent.

At a very early iteration time, there formed small, vertically oriented stacks of water cells in the upper half of the grid and similar stacks of S_1 in the lower half of the grid. These stacks steadily enlarged and moved toward the central part of the grid. Aggregation of each common type of cell stack occurred as they moved toward the central section of the grid. An interface formed with a greater concentration of water in the lower half while the

second liquid dominated the upper half. The interface was unorganized with large "fingers" of cells from each half projecting into the other half of the grid. The configuration at the interface changed with each iteration but retained a roughly similar pattern once a relatively constant configuration had been reached. Some water and S_1 cells wander in and out of the opposite phases, producing an incomplete separation of phases. Figure 17 shows the interface in this model. This interface structure and behavior has also been observed from MD and MC simulations.[54–59] The simulated events of the demixing process present an intriguing model of a phenomenon that might be difficult to examine experimentally.

A companion study[52,53] tested the ability of this model to simulate the partitioning of solute molecules between the two phases, governed by their relative lipophilicity. The addition of a small number of solute molecules was made to the initial, random mixture. As the dynamics proceeded, it was

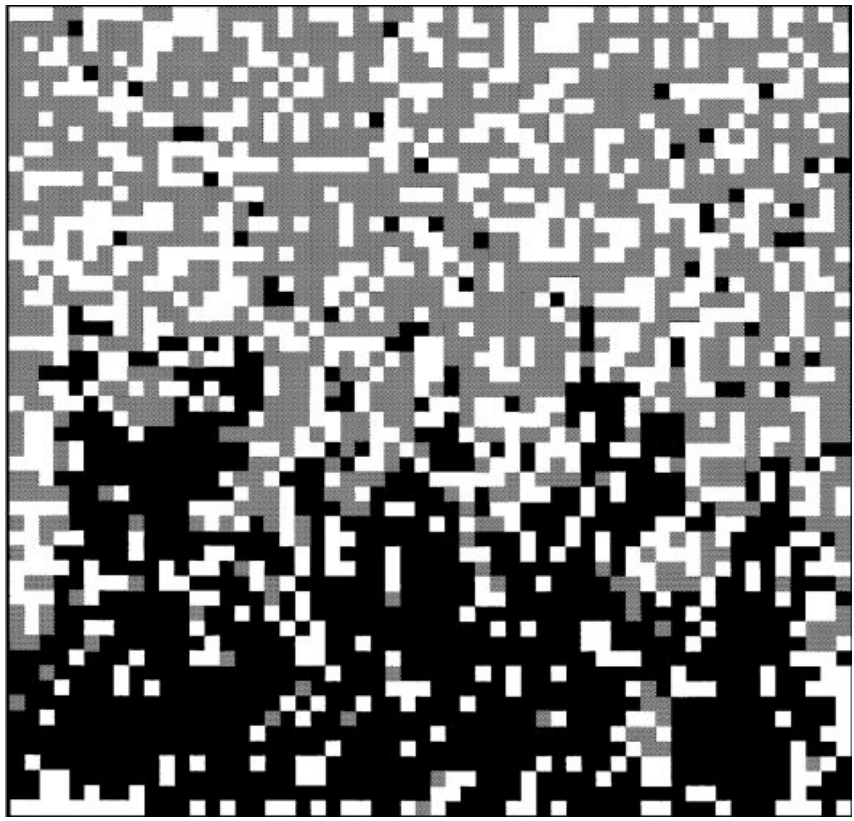

Figure 17 A model of two immiscible liquids, black and gray cells, forming an interface.

observed that the solute was associating with the patches of solvent to which it had the closest parameter-governed affinity. As a relatively stable configuration developed, the ratio of the solute molecules among the two phases became relatively constant. This ratio is the partition coefficient of the solute between the two phases. The dominant rules influencing the partition coefficient were $P_B(WS_2)$ and $P_B(S_1S_2)$, where the latter parameter reflects the affinity of the solute for the two liquids W and S_1, where S_2 is the "solute" molecule.

Observing the course of the dynamics, we see a constantly changing pattern from the random configuration at the outset to the eventual formation of a disturbed interface and separated compartments of the two solvents. The solute molecules moved rapidly to the patches in which the rules have ordained an affinity. The solute molecules partitioned themselves among the patches long before the two phases and the interface have formed.

Micelle Formation

The organization in a system of two immiscible liquids results from a general pattern of behavior found among two molecules with little attraction to each other. Another example of this can be found in the behavior of a molecule that has two substructures with significantly different polarities. A term applied to these molecules is *amphiphile*, denoting a molecule with a dual polarity character. Under certain conditions, these molecules may interact with one another to form large, long-lasting, functional structures called micelles.[60,61] These are organized and capable of entrapping solute molecules and forming membranes in biological systems. A micelle is a structure formed from the close interaction of the lipophilic fragments of amphiphiles plus the electrostatic encounters of the polar end of the amphiphile with the surrounding water. Typically, the micelles assume a spherical structure with the nonpolar fragments in the interior and the polar fragments on the periphery, interacting with the aqueous solution.

The formation of micellular structures is a dynamic process that has been modeled using cellular automata.[62] The model of an amphiphile was created by treating each face of a square automaton cell as an independent structure, as described in Figure 4. Each face of this variegated cell can have its own set of $P_B(X)$ and $J(X)$ values. Three of the faces were considered as equivalent and were endowed with rules modeling a lipophilic or nonpolar part of the amphiphile. The other face was treated as a polar fragment of the molecule and assigned characteristic rules. The outcome of the dynamics was the creation of structures in which the nonpolar fragments were in the interior of an aggregation of cells while the polar fragment lay on the periphery as seen in Figure 18. The interpretation of these organized clusters is that they model a micelle. The dominant influence on the formation of these structures is the extent of nonpolar character of the designated three sides of the cell. Of secondary influence is the polarity of the remaining face of the cell. If this is too

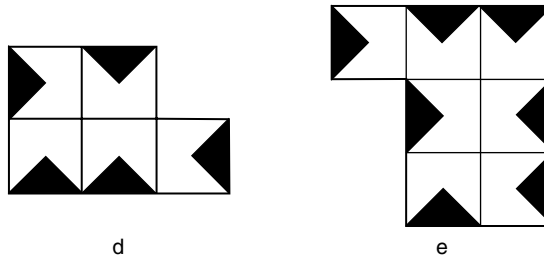

Figure 18 Two examples of micelles formed from the dynamics of amphiphile interactions. The dark faces represent the polar exterior of the molecules, and the light faces represent the lipophilic fragments of the amphiphiles.

polar, the micelle formation is retarded. Both of these influences produce models mimicking experiment.[63,64]

Another study on these variegated cells depicting an amphiphile revealed a temperature effect on the critical micelle concentration (cmc) that was minimal at about $P_B(W) = 0.25$. Experimentally, the minimal cmc value occurs at about 25 °C.[64] The onset of the cmc was also modeled and shown to be dependent on a modestly polar fragment of the amphiphile.

Membrane Permeability

An extension of the micelle and diffusion models was a simulation of the diffusion of a solute through a layer of lipophilic cells simulating a membrane separating two water compartments.[65] A layer five cells wide was positioned on a grid between two compartments, each filled with water. The membrane cells were endowed with a $P_B(WS)$ rule making them lipophilic. The membrane cells could move about within the layer according to their rule response, but they could not escape from the layer. The two water cell compartments on either side were assigned identical rules but were ''colored'' differently in order to monitor their origins after some movement into and through the membrane layer. This model included a section of a membrane with aqueous compartments on either side. In biological structures, this might be a living cell membrane or tissue in an aqueous environment. Of interest was how effectively this cellular automata model reflected experimental evidence about diffusion through a membrane. It is recognized that water and some small molecules pass through membranes by passive diffusion.[66] The cellular automata dynamics revealed that water molecules from both compartments pass into and through the membrane as expected.

To model the behavior of a solute in a membrane environment, 15 cells simulating solute molecules at a concentration of 0.01 were positioned randomly near the lower edge of the membrane surface. These cells were endowed with rules making them hydrophilic. As the dynamics proceeded, it was

observed that more water molecules from the upper compartment passed into and through the membrane than water from the lower compartment. Since the solute molecules were hydrophilic, the membrane was relatively impervious to their passage. The behavior of this model is consistent with experimental observations collectively referred to as an osmotic effect.

A further part of the study[65] was designed to examine the effects of various degrees of lipophilicity of the solute molecules. As the lipophilicity of the solute increased, it was observed that an increasing number of solute particles passed through the membrane from the lower aqueous compartment. There was no accumulation of solute molecules within the lipophilic membrane. At a level of lipophilicity midway on the scale, that is, about $P_B(WS) = 0.5$, there was an abrupt change in this behavior. Beyond this critical lipophilicity, the number of solute molecules passing through the membrane started dropping (Figure 19) until the number is nearly zero. At the critical lipophilicity, the accumulation of solute molecules in the membrane *increased* sharply (Figure 19). The passage of drug molecules of varying lipophilicity into the central nervous system or other body compartments is known to follow this general pattern. It has not been noted, however, that a change in diffusion may be so abrupt with so small a change in the lipophilicity as the cellular automata models predict. The modeling suggests that a very modest change in molecular structure changing the lipophilicity on either side of the critical $P_B(WS)$ value might produce a significant change in the degree of diffusion through a membrane. This concept may be of value in molecular design directed toward regulating membrane passage.

The observation that levels of lipophilicity above the critical $P_B(WS)$ value produce a dramatic increase in the concentration of the solute molecules in the membranes can have consequences in terms of the need for an accurate understanding of molecular changes that enhance membrane absorption. This

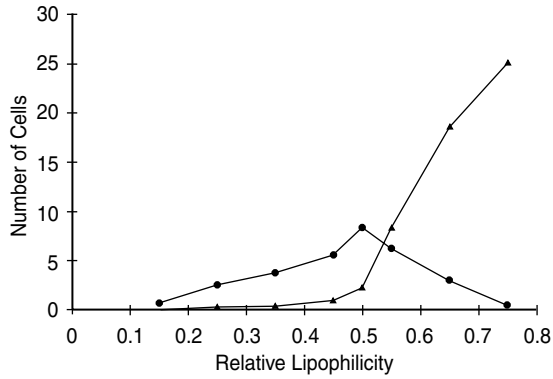

Figure 19 The relative concentrations of solutes (**a**) passing through the membrane, • and (**b**) lodging within the membrane, ▲.

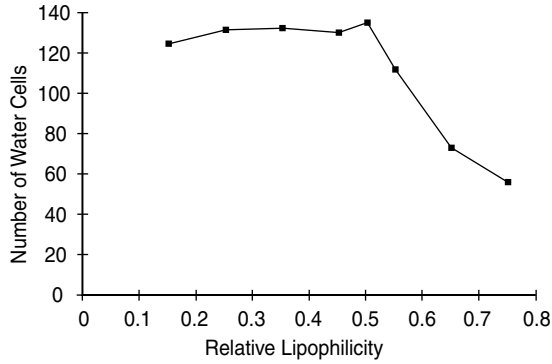

Figure 20 The flux of water through the membrane as a function of changing solute lipophilicity.

observation led to the speculation that the increasing concentration of solute molecules within the membrane should retard somewhat the flux of water through the membrane. Retardation was indeed found to result from increasing the P_B(WS) value of the solute molecules beyond the critical point (Figure 20). At the critical P_B(WS) value, the flux of water through the membrane went from a relatively constant amount to a lower level, decreasing with increasing P_B(WS) values. This behavior has apparently not been noted or studied, but it provokes interest to seek experimental confirmation.

Acid Dissociation

The success with the variegated cell in creating a model of an amphiphile,[62] described above, opened the possibility of using a similar type of cell to model an organic acid molecule. The dissociation of an acid and the influence of the environment on this process was investigated[67] using the general reaction scheme shown in Figure 21. A cell modeling an organic acid molecule was divided into two parts, one face representing the carboxyl group Y, and the other three faces, X, representing the nondissociating, nonpolar parts of the acid molecule. The rules assigned to each face corresponded

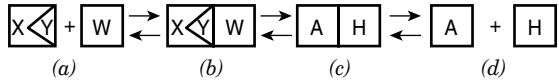

Figure 21 The scheme employed in the modeling of the dissociation of an organic acid. Scheme (a) denotes the neutral organic acid with the Y sector representing the carboxyl group and X the remainder of the molecule. Scheme (b) denotes the hydrated organic acid. Scheme (c) denotes the water–acid ion pair. Scheme (d) denotes the dissociated acid and a hydronium ion.

to their lipophilicity relative to water. The acid molecule moved through the water the same as any solute molecule in the previously described studies. The "strength" of the acid, that is, its propensity to dissociate, was governed by the probability rule, P_D, shown as the change from (**b**) to (**c**) in Figure 21. The anion, A, and the hydronium ion, H, moved about the system as separate ingredients. The hydronium ion was endowed with greater mobility because it is known to move rapidly from one oxygen to another within the hydrogen-bonded system of water. This mobility was accomplished by allowing any of the four possible neighboring water molecules to exchange positions with the hydronium cell H.

The purpose of the modeling was to examine the influence of the solution environment on the extent of dissociation of an organic acid. A series of studies was performed initially to establish the validity of the model in preparation for later work. An initial test of the model was to vary the P_D value and monitor the concentration of products. As expected, an increase in the P_D rule produced an increase in the calculated acid dissociation constant K_a. A second study examined the influence of acid concentration on the observed properties. As expected, the K_a was approximately constant over a modest concentration range. A third study considered the effect of water temperature on the acid dissociation. As the modeled water temperature was increased by increasing the $P_B(W)$ value, the value of the K_a decreased, in agreement with a common, but not universal, observation of the effect of temperature on acid dissociation. These three preliminary studies thus revealed emerging attributes consistent with experimental observations.

The next study on solution environment influences included the presence of another molecule in the solution. This cosolute was endowed with an attribute of nondissociation. Its lipophilicity was varied in a series of studies revealing that the dissociation of the acid decreased when the lipophilicity of the cosolute decreased. These results may present a new insight into this effect.

The interaction of two acids of different strength was also simulated using basically the same model. The observed dissociations revealed a strong and unequal influence of the two acids on each other. Both acids exhibit a suppression in their dissociations relative to their behavior in pure solution. The weaker acid is significantly more suppressed than the stronger acid. The decrease in dissociation of the two acids in a mixture cannot be readily calculated from the acid concentrations and their individual dissociation constants because of the complicating influences of ionic solvation effects on the water structure plus temperature factors.[68] Thus the cellular automata model may offer a distinct advantage in this regard.

Percolation

Percolation is a phenomenon associated with ingredients in a system reaching a critical state of association so that information may be transmitted

from one ingredient to another across or through the system without interruption. Some objects are distributed over a space in a random fashion. Because of the scarcity of these objects, little or no physical contact is encountered. No information is exchanged within the system. If enough additional objects are randomly added to the system, there arises a finite probability that some of these objects may be associated to form clusters. There is some exchange of information within the clusters, but they are isolated and so the information exchange is confined within each cluster. If enough objects are randomly added to the system, the possibility arises that some clusters may appear as a single cluster that spans the entire length or width of the system. This spanning cluster produces a conduit over which an uninterrupted flow of information is possible across the system. This flow of information takes place via a process called percolation. The minimum number of objects in the system necessary to have a finite probability of percolation occurring is called the percolation threshold or percolation point.

Percolation is widely observed in chemical systems. It is a process that can describe how small, branched molecules react to form polymers, ultimately leading to an extensive network connected by chemical bonds. Other applications of percolation theory include conductivity, diffusivity, and the critical behavior of sols and gels. In biological systems, the role of the connectivity of different elements is of great importance. Examples include self-assembly of tobacco mosaic virus, actin filaments, and flagella, lymphocyte patch and cap formation, precipitation and agglutination phenomena, and immune system function.

The onset of percolation and the conditions that produce this phase-transition phenomenon have been of considerable interest to many disciplines of science. The classical studies have focused on immobile ingredients in a system that increase in concentration by randomly adding to the collection of particles. It is possible to estimate the conditions leading to the onset of percolation under these circumstances. When the ingredients are in motion, this estimation is far more difficult. It is an obvious challenge that was tackled using cellular automata.

The studies[69] were designed to reveal several attributes of the dynamic system. The first of these is the configuration of the system as additional cells were added to the grid. Of particular interest were the relative proportions of the five possible configurations of the cells, described by the concentrations of each configuration, f_x, defined earlier in this review. The second attribute studied was the growth of the size of the largest cluster as the number of cells in the system increases. Of prime interest is the concentration of occupied cells at which the percolation phenomenon occurs. Also of interest were the occupied cell concentration when the onset of percolation begins and the occupied cell concentration where percolation is occurring, on average, 50% of the time. The study also looked at the information content of the system as a function of concentration change and at possible relationships among attributes of the system.

A series of runs were made starting with a concentration of 100 occupied cells in a grid of 3025 cells. Successive increases in the occupied cell concentration were introduced in increments of 100. The concentrations of the five configurations, f_0 through f_4, were recorded for each occupied cell concentration. The size of the largest cluster was also recorded at each occupied cell concentration. Finally, the occupied cell concentrations at which the onset of percolation takes place and at which there is a 50% probability of percolation were recorded.

The percolation point in these studies was calculated from the concentration producing a 50% probability of percolation. A comparison between these percolation onset concentrations and the concentrations corresponding to the maximum Shannon information content, I_{max}, revealed a close correspondence.[69] The I_{max} concentration appears to be very close to the percolation onset concentration for each of the three parameter sets. The largest cluster size at a concentration producing 50% percolation was about the same for all parameter sets. This size is a cluster size occupying close to 20% of the grid area in these studies. Stated another way, as the concentration increases for any parameter set study, there is a relatively common cluster size corresponding to a 50% probability that percolation will occur. This interpretation is an alternative definition of a percolation point when the system is dynamic. It appears from this modeling that a maximum of diversity among the f_x types is necessary to create conditions in which the percolation process may begin.

SOLUTION KINETIC MODELS

First-Order Kinetics

Many important natural processes ranging from nuclear decay to unimolecular chemical reactions are first order, or can be approximated as first order, which means that these processes depend only on the concentration to the first power of the transforming species itself. A cellular automaton model for such a system takes on an especially simple form, since rules for the movements of the ingredients are unnecessary and only transition rules for the interconverting species need to be specified. We have recently described such a general cellular automaton model for first-order kinetics and tested its ability to simulate a number of classic first-order phenomena.[70]

The prototype first-order transition is radioactive decay, $A \rightarrow B$, in which the concentration [A] of a species A decreases according to the rule that each A ingredient has a probability $P_t(A,B)$ per unit time (here, per iteration) of converting to some other form B. For small numbers of A ingredients, the actual decay curve observed for [A] is rather jagged and only roughly exponential, as a result of the irregular decays expected in this very finite, stochastic system, as illustrated in Figure 22. However, as the number of decaying ingredients is increased, the decay curve approaches the smooth exponential fall-off

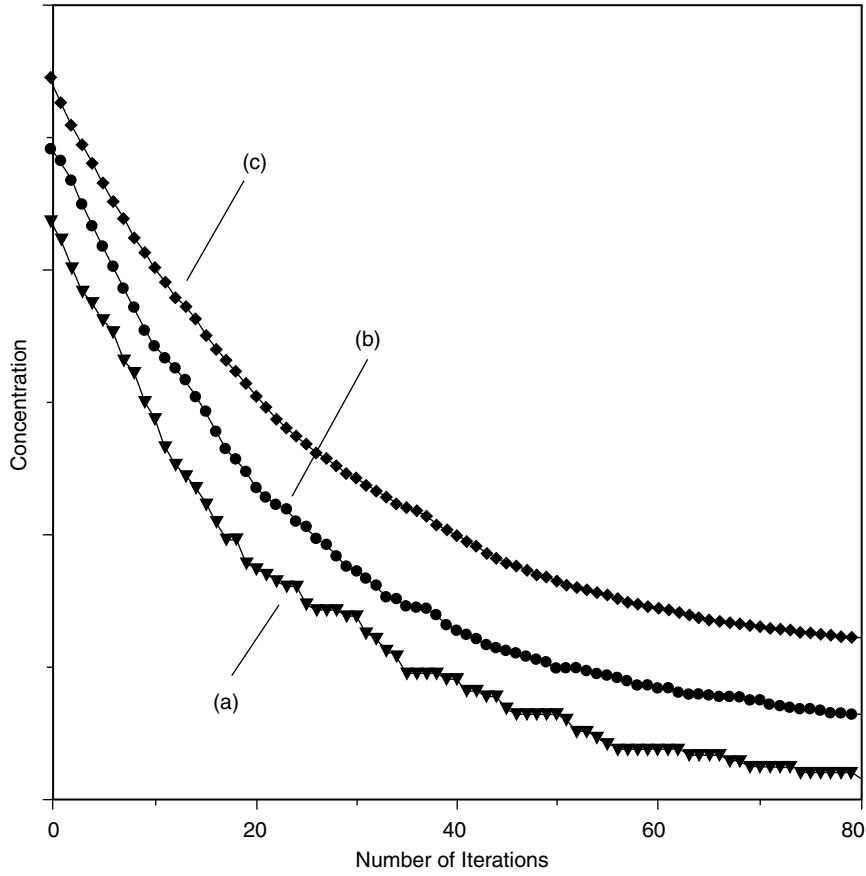

Figure 22 Exponential decays (offset) for first-order models with (**a**) 100 cells, (**b**) 400 cells, and (**c**) 2500 cells. The transition probability is 0.04 per/iteration.

expected for a deterministic system obeying the rate equation

$$d[\mathrm{A}]/dt = -k[\mathrm{A}] \qquad [6]$$

This smooth decay is also illustrated in Figure 22.

When a reverse transition probability $P_t(\mathrm{B,A})$ for the transition $\mathrm{B} \rightarrow \mathrm{A}$ is included, the model simulates the first-order equilibrium:

$$\mathrm{A} \rightleftarrows \mathrm{B}$$

Here too, the finite size of the system causes notable fluctuations, in this case in the value of the equilibrium constant K, which fluctuates with time about the deterministic value

$$K = P_t(\mathrm{A,B})/P_t(\mathrm{B,A}) \qquad [7]$$

As an example, for 10 trials with 400 ingredients taking $P_t(A,B) = 0.05$ and $P_t(B,A) = 0.04$, we obtained $K = 1.27$ with a standard deviation of 0.13, compared to the deterministic value of 1.25. As a further test of the model, one can ask whether it is ergodic (see Sklar[71] for a discussion) in the sense that the average of K over time (i.e., for a single system observed for a long time after reaching equilibrium) is equal to the average K for identical systems taken at a particular time in a large number of trials. When this was tested for 1000 time steps (separated by 100 iterations) versus 1000 trials, the results were statistically identical, indicating that the first-order cellular automaton model is sensibly ergodic.[70]

The first-order model can also be used to examine sequences of transformations of the form A $\rightarrow$ B $\rightarrow$ $C\cdots$. For the simple example A $\rightarrow$ B $\rightarrow$ C, the concentration of the initial reactant A falls exponentially, that of the intermediate species B rises then falls, and that of C builds up as it is fed from B. These time-dependent changes are illustrated in Figure 23 using one set of transformation probabilities.

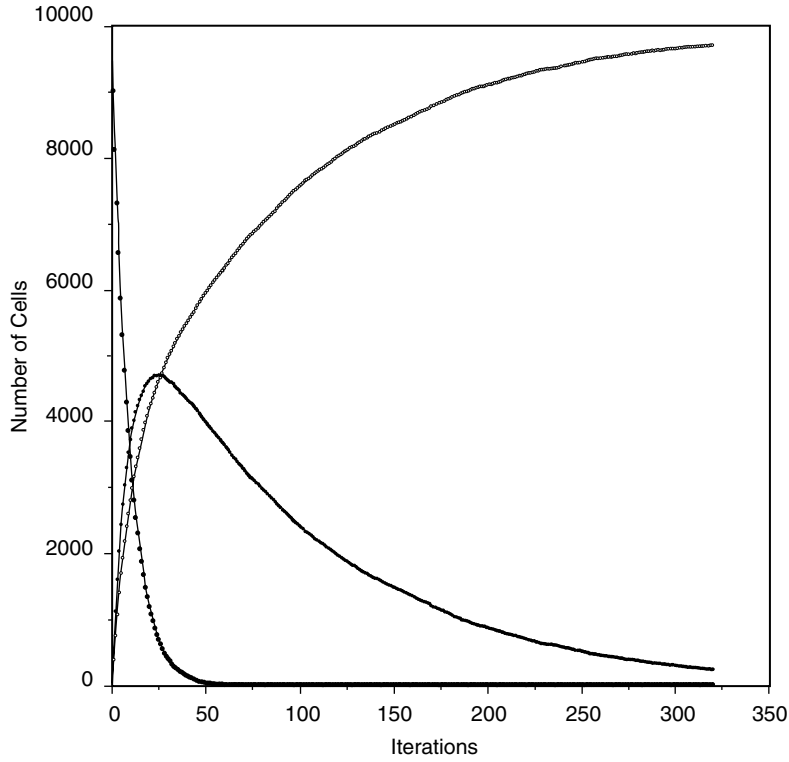

Figure 23 Pattern of cell counts for a 100×100 cell simulation of two consecutive first-order reactions $A \rightarrow B \rightarrow C$, showing the rise and decay of the intermediate species B.

The series model can be extended to longer series and to the inclusion of reversibility to illustrate a variety of fundamental kinetic phenomena in an especially simple and straightforward manner. Depending on the relative rates employed, one can demonstrate the classic kinetic phenomena of a rate-limiting step and preequilibrium,[72] and one can examine the conditions needed for the validity of the steady-state approximation commonly used in chemical kinetics.[70]

Kinetic and Thermodynamic Reaction Control

Parallel competing reactions

$$A \rightarrow B \quad [8]$$

$$A \rightarrow C \quad [9]$$

can be simulated.[70] An especially interesting example is present when the reactions are reversible:

$$A \rightleftarrows B \quad [10]$$

$$A \rightleftarrows C \quad [11]$$

With properly chosen transformation probabilities, cellular automata can be used to examine the conditions governing thermodynamic and kinetic control of reactions.[73] Knowledge of such conditions is an important element in organic synthesis[74,75] and is an important consideration for many industrial reactions.[76,77] Figure 24 shows a reaction being modeled with the parameters $P_t(A,B) = 0.01$, $P_t(A,C) = 0.001$, $P_t(B,A) = 0.02$, and $P_t(C,A) = 0.0005$, and a set of 10,000 ingredients. Starting with reactants A, the kinetically favored product B is produced in excess in the initial stages of the reaction, whereas at later times the thermodynamically favored product C gains dominance. The number of interacting species (10,000) is rather large in this illustration, and the results of the cellular automata model are in good agreement with those found in a deterministic, numerical solution for the same conditions.[78] For example, the cellular automata model yields final, equilibrium concentrations for species B and C of $[B] = 0.1439 \pm 0.0038$ and $[C] = 0.5695 \pm 0.0048$ compared to the reported deterministic values of 0.14 and 0.571, respectively.

Excited-State Kinetics

Another important application of the first-order model is the examination of the ground- and excited-state kinetics of atoms and molecules.[79] These systems are characterized by competing first-order transitions representing both radiative and nonradiative processes. The radiative processes normally

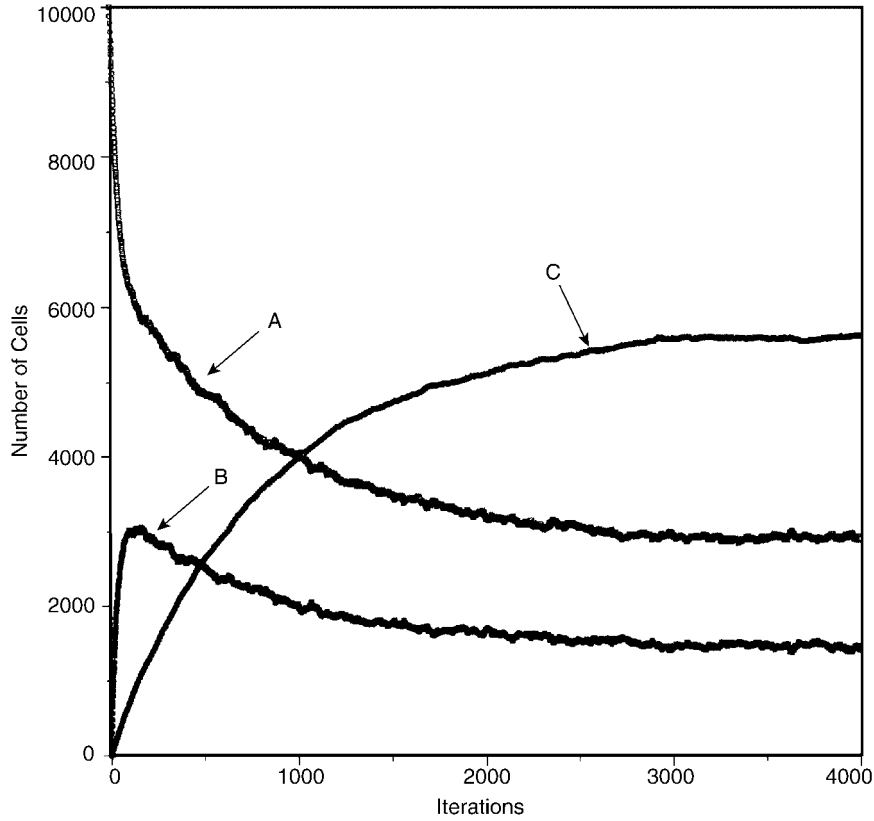

Figure 24 Illustration of kinetic and thermodynamic reaction control: B is the kinetically favored product (higher probability of formation from A), and C is the thermodynamically favored product (greater equilibrium constant with A).

include light absorption from the ground state as well as fluorescence and phosphorescence originating from the excited states. The nonradiative processes include spin-allowed internal conversions and spin-forbidden intersystem crossings. The "species" in this case are the different atomic or molecular electronic states involved, represented in the visualizations by different colors. Cellular automata models can be constructed to represent both pulse conditions, which are normally used, for example, to determine excited-state lifetimes, and steady-state conditions, which are customarily employed to determine spectra and the quantum yields of the processes taking place. Pulse conditions are simulated by starting all the ingredients in a particular excited state or in a distribution of excited states characteristic of a starting distribution created by some process of interest; the states are allowed to decay according to the transition probabilities appropriate to that system, and

the time evolutions of their populations are followed. Under pulse conditions, the decays eventually lead to a condition in which all of the ingredients are in the ground state (in the absence of additional trapping states). For steady-state conditions, additional transitions from the ground to the excited states are introduced with appropriate transition probabilities.

One illustration of the excited-state cellular automata model is the dynamics of the excited-state transitions of oxygen atoms.[80] The oxygen atom has a ^3P ground state and ^1S and ^1D excited states. Emissions from the latter two excited states play an important role in the dramatic light displays—the aurora borealis or "northern lights"—seen under certain conditions in the northern polar skies and similar emissions have been detected in the atmospheres of Mars and Venus. The excited states are believed to be produced mainly by dissociative recombinations of ionized oxygen molecules and electrons generated in the atmosphere by ultraviolet (UV) bombardment during the daylight hours.[81,82]

$$O_2^+ + e^- \rightarrow O^* + O^{**} \qquad [12]$$

In this equation, the species O* and O** are unspecified atomic oxygen states, ^3P, ^1S, or ^1D. The most prominent feature in the atmospheric displays is normally the green, spin-allowed $^1S \rightarrow {}^1D$ transition appearing at 5577 Å.

Using transition probabilities taken from the compilation of Okabe,[83] we simulated the dynamics associated with these atomic transitions under both pulse and steady-state conditions. For the pulse simulations, two starting conditions were examined: the first in which all ingredients started in the upper ^1S excited state, and the second in which the ingredients started in a distribution believed characteristic of that produced by the dissociative recombination process in Eq. [12].[82] The simulations yield excited-state lifetimes and luminescence quantum yields consistent with the experimental observations for these properties. An interesting feature arising from these studies is the possibility that a $^1D/^3P$ population inversion could occur under certain atmospheric conditions, as a result of the exceedingly long lifetime of the ^1D state, estimated to be about 200 s.[82]

Second-Order Kinetics

Several groups have developed cellular automata models for particular reaction–diffusion systems. In particular, the Belousov–Zhabotinsky oscillating reaction has been examined in a number of studies.[84–86] Attention has also been directed at the $A + B \rightarrow C$ reaction, using both lattice-gas models [87–90] and a generalized Margolus diffusion approach.[91] We developed a simple, direct cellular automaton model[92] for hard-sphere bimolecular chemical reactions of the form

$$A + B \rightarrow C + D \qquad [13]$$

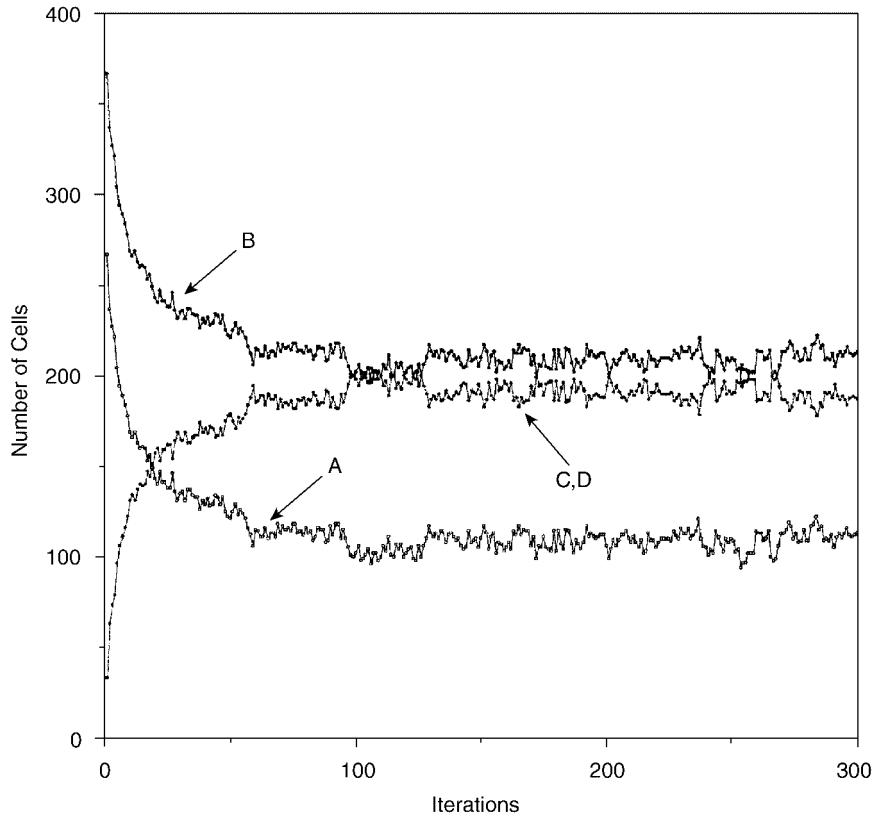

Figure 25 Irreversible second-order reaction $A + B \rightarrow C + D$, with $P_r(A,B) = 0.1$ and initial conditions $[A]_0 = 100$ cells, $[B]_0 = 200$ cells.

As before, the different species are assigned different colors in the visualization. In this model, the reactant and product species diffuse about the grid in random walks. When the species A and B encounter each other (come to adjacent cells) on the grid, the probability that these species transform to C and D is determined by an assigned reaction probability $P_r(AB)$. The simulations take place on a toroidal space such that ingredients leaving the grid on one side appear at the opposite edge. Initially, the ingredients are placed randomly on the grid.

The production of species C over time for starting counts of 100 A and 200 B ingredients on a $100 \times 100 = 10{,}000$ cell grid is shown in Figure 25. The expected second-order rate law

$$d[C]/dt = k[A][B] \qquad [14]$$

is found to be obeyed in the simulations, subject to fluctuations expected for a system containing a finite number of reacting ingredients. When the results from a number of trials are combined to mimic, in effect, the results for a much larger system, the fluctuations become relatively small and the results approach the deterministic forms.

When a back reaction $C + D \rightarrow A + B$ with probability $P_r(C,D)$ is included in the automaton rules, the equilibrium

$$A + B \rightleftarrows C + D \qquad [15]$$

can be simulated. Once the system has stabilized from its initial nonequilibrium concentrations, fluctuations about the equilibrium concentrations occur over time, and the relative size of these fluctuations decreases as the number of ingredients increases. Figure 26 shows the approach to equilibrium that occurs for an initial system containing 300 A cells and 400 B cells on a 100×100 cell grid, with reaction probabilities $P_r(A,B) = 1.0$ and $P_r(C,D) = 0.5$.

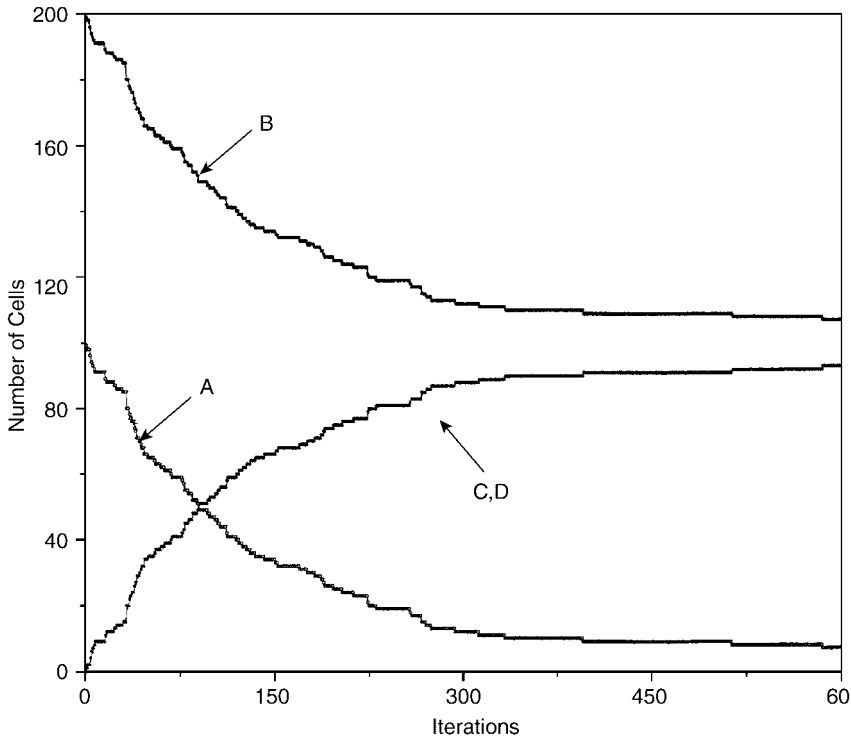

Figure 26 Reversible second-order reaction $A + B \rightleftharpoons C + D$, with $P_r(A,B) = 1.0$, and $P_r(C,D) = 0.5$. Initial conditions are $[A]_0 = 300$ cells, $[B]_0 = 400$ cells, $[C]_0 = [D]_0 = 0$.

A variation of this theme occurs in a *pseudo-first-order reaction*. This type of reaction involves an actual second-order reaction, $A + B \rightarrow C + D$, in which one of the reactants, say B, is present in sufficient excess that its variation becomes effectively unnoticeable. Simulations with, for instance, 50 A ingredients and 1000 B ingredients, can bear out the expectation that the reaction kinetics are such that the rate of production of C and D appears to depend only the concentration of A.

This second-order model is a first step toward the examination of more complex models that will include, for example, the influences of intermolecular forces (departing from the hard-sphere ideal), solvent, temperature, and other features. Variations such as the inclusion of unstable intermediate species and additional reactions can also be simulated, so that, potentially, quite complex systems can be studied.

Enzyme Reactions

Many aqueous solution cellular automata models discussed earlier were created for systems in which there have been no changes in the states of any cells that model ingredients. Of great interest are the reactions catalyzed by enzymes, the engines of biochemical function. Some studies relating to this have been reported,[89,90] but more attention to this area of modeling would be of value. A recent study on the kinetics of an enzyme reaction[93] considered the Michaelis–Menten model shown in Eq. [16].

$$E + S \rightleftarrows ES \rightarrow E + P \qquad [16]$$

Rules selected for this reaction included the joining and breaking probabilities as described earlier. In addition, it was necessary to include a probability of conversion, P_c, of an enzyme–substrate pair, ES, to an enzyme–product pair, EP, that was programmed to be an irreversible event. The cells designated as enzymes, E, were not permitted to move, and their random distribution in the grid limited them to a separation of at least 10 cells. Once a substrate molecule joined with an enzyme, no other pairings were possible. The lipophilicity of the substrate and product molecules were varied using the P_B(WS) and P_B(WP) rules.

The dynamics were run for several concentrations of substrate and variations in the P_c values. Initial velocities of the reaction were recorded. The Michaelis–Menten model was observed and characteristic Lineweaver–Burk plots were found from the model. Systematic variation of the lipophilicity of substrates and products showed that a lower affinity between a substrate and water leads to more of the $S \rightarrow P$ reaction at a common point along the reaction progress curve. This influence is greater than that of the affinity between the substrate and the enzyme. The study created a model in which the more lipophilic substrates are more reactive. The water–substrate affinity appears

to influence primarily the concentration of the ES complex at the observed point along the reaction progress curve. A low affinity between water and substrate favors a high ES concentration at this point. A hydrophilic substrate appears to be more entrapped in the water continuum, hence to be less available to the enzyme. It was also observed that an accumulation of product molecules around the enzymes coincides with a decline in the reaction rate.

An Anticipatory Model

An anticipatory system is one in which information is sent from an element in the early stages of a sequence of events to a nonacting element that *will* arise in that sequence. The information sent ahead *prepares* the future element to function based on information arising in advance of the normal sequence of events. The unmanifested element anticipates its ultimate manifestation in the sequence. An anticipatory system has been modeled by Kier and Cheng[94] using the dynamic characteristics of cellular automata. A concentration of an intermediate product influences the creation of a supplemental enzyme that enhances the competence of an enzyme downstream. This anticipation of a future event creates a condition in which the concentration of a later substrate is suppressed, a property characteristic of the system.

The anticipatory model employed in this study had the following stepwise reactions.

$$\mathrm{A} + e_1 \rightarrow \mathrm{A}e_1 \rightarrow \mathrm{B}e_1 \rightarrow \mathrm{B} + e_1 \qquad [17]$$

Equation [17] is the conversion of A to B, assuming an irreversible first-order reaction catalyzed by the enzyme e_1. The rules governing the initial encounter, $P_{\mathrm{B}}(\mathrm{A}e_1)$ and $J(\mathrm{A}e_1)$, are set at the beginning of each run. The next step in the reaction is modeled as shown in Eqs. [18].

$$\mathrm{B} + e_2 \rightarrow \mathrm{B}e_2 \rightarrow \mathrm{C}e_2 \rightarrow \mathrm{C} + e_2 \qquad [18a]$$

$$\mathrm{B}e_2 \rightarrow \mathrm{C}e_{2,4} \rightarrow \mathrm{C} + e_{2,4} \qquad [18b]$$

If the system is not an anticipatory one, the conversion of $\mathrm{B}e_2$ to $\mathrm{C}e_2$ would follow only one route. If the system is anticipatory, as shown, there are two paths for the conversion of $\mathrm{B}e_2$. The formation of $e_{2,4}$ represents an enzyme that may function upon substrates B or D (Eqs. [19] and [20]).

$$\mathrm{C} + e_3 \rightarrow \mathrm{C}e_3 \rightarrow \mathrm{D}e_3 \rightarrow \mathrm{D} + e_3 \qquad [19]$$

$$\mathrm{D} + e_4/e_{2,4} \rightarrow \mathrm{D}e_4/e_{2,4} \rightarrow \mathrm{E}e_4/e_{2,4} \rightarrow \mathrm{E} + e_4/e_{2,4} \qquad [20]$$

Thus the available enzyme to convert B to C is the same for an anticipatory or nonanticipatory system. In the case of an anticipatory system, there is formed, in advance of the creation of substrate D an enzyme, $e_{2,4}$, that will enhance the

reaction of D and e_4 to form product E. Substrate B serves as a predictor of the concentration of D, reducing its accumulation, relative to that in a nonanticipatory system. Substrate D is confronted with an enhanced competence of its specific enzyme, modeled by e_4 and $e_{2,4}$ facilitating the conversion of D to E. The decrease in the concentration of D is the property of the anticipatory system that is created by the feed-forward influence of $e_{2,4}$.

In each study, the dynamics revealed concentrations over time that are influenced by the presence or absence of a feed-forward or preadaption state in the system. The concentration of A steadily diminishes as successive concentrations of B, C, and D rise and fall at the same levels. The concentration of E rises at the end of the run, eventually becoming the only ingredient in the system. The concentration of D is ca. 0.25 in a nonanticipatory model. In contrast, with an anticipatory or feed-forward step in the system there is created an additional amount of enzyme specific for substrate D (enzyme, $e_{2,4}$) that is available at a future time to catalyze the conversion of D to E. This anticipatory attribute creates a property of the system in which the concentration of ingredient D is not allowed to accumulate to its normal level. In contrast, the concentration of D in an anticipatory model is approximately 0.13, about half the D concentrations for the nonanticipatory models. The concentration of B therefore serves as a predictor of the concentration of D at a later time.

Chromatographic Separation

Models of chromatographic separation based on cellular automata have been reported.[95] Solvent cells were randomly distributed over the grid at the initiation of each run. These 5230 cells, designated W, constituted 69% of the cells in the grid before the introduction of any other ingredients. The stationary phase, designated B, was simulated by the presence of cells distributed randomly over the grid, replacing 600 W cells. The B cells were immobile and were positioned at least three cells from another B cell. The solute cells, usually simulating a mixture of two different compounds, were represented by 10 cells each. These solutes replace a corresponding number of W cells and were positioned initially on the first row of the column.

The movements of the solutes and mobile phase in the grid were governed by rules denoting the joining and breaking of like or unlike cells. The gravity term was applied uniformly to all ingredients except the stationary-phase cells. The position of each solute cell was recorded at a given row in the "chromatographic" column after a certain number of iterations. Each run was repeated 100 times to achieve an average position on the column for 1000 cells of a certain solute. The position of the peak maximum was determined by summing the number of cells found in contiguous groups of 10 rows on the column. These averages were then plotted against the iteration time.

The gravity parameter for each ingredient in the simulation defines the flow rate. The polarity of the solvent, W, is encoded in the relative

self-association experienced, which in turn is governed by the rules, $P_B(W)$ and $J(W)$. High values of $P_B(W)$ and low values of $J(W)$ simulate a weak self-association, corresponding to a relatively nonpolar solvent. The migration of the solutes was found to be faster when the solvent was nonpolar. Another study in the same paper[95] modeled the influence of the relative affinities of solutes for the stationary phase B. This affinity is encoded in the parameters $P_B(SB)$ and $J(SB)$. High values of $P_B(SB)$ and low values of $J(SB)$ denote a weak affinity. These parameters characterize the structural differences among solutes that give rise to different migratory rates and separations in chromatography. These studies revealed that solutes with a greater affinity for the stationary phase migrated at a slower rate.

The effect of solvating the stationary phase was significant because this property reflects the variability of this ingredient in the selection of suitable solvents for experimental work. The variation in this property was encoded into the $P_B(WB)$ and $J(WB)$ parameters. A low value for the $P_B(WB)$ and a high value for the $J(WB)$ parameters denotes a strong affinity of the solvent, W, for the stationary phase B. The study revealed a modest influence of the parameters on migration through the "column" but a greater influence on the relative resolution among the two solutes in each study. The greater the relative solvation of the stationary phase, the poorer the resolution.

CONCLUSIONS

The examples given here illustrate some of the capabilities that cellular automata models have for enhancing our understanding of complex, dynamic systems in chemistry.[96] They also give a hint of some of the potential richness that these models may hold for the future. The limitations of the cellular automata models are quite clear: they do not make use of the customary tools of the trade—the molecular structures, the force fields, the energetic calculations, and their associated equations—that are so firmly embedded in our training and experience as computational chemists. Yet the absence of these tools also contributes to the strength of the method. In the absence of these tools, the modeler is forced to design appropriate "rules" that encapsulate the nature of things, thereby making the modeler an active participant in the process of understanding the complex systems under investigation. The heuristic rules, moreover, yield models that have an attractive simplicity, are computationally undemanding, and are readily visualizable. They also enable, in many cases, a parsing of the influences in a complex system—say, changing the temperature of one component and not of another—that can lead both to a better understanding of how these factors influence behavior and to the discovery of new insights and possibilities.

The cellular automata models described here contained some novel techniques for this paradigm. The use of a gravity rule to set the stage for the

demixing of immiscible liquids is one example. Another is the use of a variegated cell with independent states and trajectory rules for each of the four faces. These innovations have greatly enhanced the potential for cellular automata models to explore new areas of dynamic systems.

We close with two quotations. The first describes in a general way both the limitations and the promise of modeling work:[97] "Because models of this sort may provide an unjustified sense of verisimilitude, it is important to recognize them for what they are: imitations of reality that represent at best individual realizations of complex processes in which stochasticity, contingency, and non-linearity underlie a diversity of possible outcomes. Individual simulations cannot be taken as more than representative of this diversity, but repeated simulations can provide statistical ensembles that contain robust kernels of truth." The second quote positions the models between theory and experiment:[98] "It is worthwhile to emphasize that the computer based models offer a half-way house between theory and experiment. They have the rigor of mathematical models without the generality, while allowing the selection and repeatability of good, critical experiments without the enforced connection to reality."

APPENDIX

The computations were done on personal computers using a suite of programs called DING HAO. The programs are written in C.

Computing the Moving Probability

The moving probability is defined as follows:

Let A be a molecule and let S be the set of directions in which the j cell [see Figure 5(**a**)] of A is empty. Let n be the number of empty j neighbors [see Figure 5(**a**)] of A. We describe the computed direction breaking probability at d direction, written as $dirP_B(d)$, as: $dirP_B(d) = 1$ if the j cell at A's d direction is empty, $dirP_B(d) = P_B(AB)$ otherwise, where B is the molecule at the j cell in the d direction of A.

Let Q be the product of $dirP_B(d)$ for all d in S.

Let $J(d) = 1$ if the k cell and the j cell [see Figure 5(**c**)] at A's d direction are empty, $J(d) = J(AB) + y$ if only the j cell at A's d direction is empty, and the k cell at A's d direction is occupied by the molecule B, where $y = absG(A)$ if $d =$ south, $y = 0$ otherwise. Then $y = G_D(AB)$ if the j cell at A's d direction is occupied by the molecule B. Furthermore, in this case, n should be incremented by 1.

Let $temp = (n/Q) - 1$, if $Q > 0$.

Let $MP(d)$ be the probability of moving in a direction d and define it as follows:

$MP(d) = f/[1 + temp/J(d)]$ if $Q > 0$, where f is the free movement probability of A, $MP(d) = 0$ otherwise.

Let MP be the sum of $MP(d)$ for all four directions.

Normalize $MP(d)$ if it is necessary. That is, change $MP(d)$ to $MP(d)/MP$ for all directions d if $MP > 1$.

We choose this method to compute the moving probability because:

1. It allows a simple computation.
2. It allows the influence of k cells, but the effect is limited so that the influence of one direction does not overshadow the influence from other directions.
3. When $J(\mathrm{AB}) = 1$, and all adjacent j cells to the occupant are empty, the probability of moving in any direction is 0.25 of its free movement probability. This reduced joining parameter agrees with the intuitively reasonable assumption that any occupant should not be biased on any direction (unless gravity is considered).

The directional moving probabilities are computed for various configurations and rules, giving the results in Figures 7–10.

REFERENCES

1. J. I. Steinfeld, J. S. Francisco, and W. L. Hase, *Chemical Kinetics and Dynamics,* Prentice Hall, Englewood Cliffs, NJ, 1989.
2. A. Papoulis, *Probability, Random Variables, and Stochastic Processes,* McGraw-Hill, New York, 1991.
3. D. T. Gillespie, *J. Phys. Chem.,* **81,** 2340 (1977). Exact Stochastic Simulation of Coupled Chemical Reactions.
4. D. T. Gillespie, *J. Chem. Phys.,* **72,** 5363 (1980). Approximating the Master Equation by Fokker–Planck-type Equations for Single-Variable Chemical Systems.
5. D. T. Gillespie, *J. Comput. Phys.,* **22,** 403 (1976). A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions.
6. J. Shi and J. R. Barker, *Int. J. Chem. Kinetics,* **22,** 187 (1990). Incubation in Cyclohexane Decomposition at High Temperatures.
7. L. Vereecken, G. Huyberechts, and J. Peeters, *J. Chem. Phys.,* **106,** 6564 (1997). Stochastic Simulation of Chemically Activated Unimolecular Reactions.
8. J. M. Haile, *Molecular Dynamics Simulations: Elementary Methods,* Wiley, New York, 1992, pp. 11–18.
9. M. P. Allen and D. J. Tildesley, *Computer Simulations of Liquids,* Oxford University Press, New York, 1987.
10. W. L. Jorgensen, in *Encyclopedia of Computational Chemistry,* P. v. R. Schleyer, Ed., Wiley, New York, 1998, pp. 1754–1763. Monte Carlo Simulations for Liquids.

11. 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 Simulation. T. P. Lybrand, in *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1990, Vol. 1, pp. 295–320. Computer Simulation of Biomolecular Systems Using Molecular Dynamics and Free Energy Perturbation Methods.
12. S. M. Ulam, *Proc. Int. Congr. Math.* (held in 1950), **2,** 264 (1952). Random Processes and Transformations.
13. S. M. Ulam, *Adventures of a Mathematician,* Charles Scribner's Sons, New York, 1976.
14. J. von Neumann, *Theory of Self-Replicating Automata,* A. Burks, Ed., University of Illinois Press, Urbana, 1966.
15. K. Zuse, *Int. J. Theoret. Phys.,* **21,** 589 (1982). The Computing Universe.
16. T. Toffoli and N. Margolus, *Cellular Automata Machines,* MIT Press, Cambridge, MA, 1987.
17. M. Schroeder, *Fractals, Chaos, Power Laws,* W. H. Freeman, New York, 1991, p. 371.
18. J. Signorini, in *Cellular Automata and Modeling of Complex Physical Systems,* P. Manneville, N. Boccara, G. Y. Vishniac, and R. Bidaux, Eds., Springer-Verlag, New York, 1990, pp. 57–70. Complex Computing with Cellular Automata.
19. G. Y. Vichniac, *Physica D,* **10,** 96 (1984). Simulating Physics with Cellular Automata.
20. T. Toffoli, *Physica D,* **10,** 117 (1984). Cellular Automata as an Alternative (Rather than an Approximation of) Differential Equations in Modeling Physics.
21. S. Wolfram, *Physica D,* **10,** vii (1984) Preface.
22. S. Kauffman, *Physica D,* **10,** 145 (1984). Emergent Properties in Random Complex Automata.
23. S. Wolfram, Ed., *Theory and Applications of Cellular Automata,* World Scientific, Singapore, 1986.
24. R. J. Gaylord and K. Nishidate, *Modeling Nature: Cellular Automata Simulations,* Springer-Verlag, New York, 1996.
25. H. Gutowitz, *Cellular Automata: Theory and Experiment,* MIT Press, Cambridge, MA, 1991.
26. R. Kapral and K. Showalter, *Chemical Waves and Patterns,* Kluwer, Dordrecht, 1995.
27. G. B. Ermentrout and L. Edelstein-Keshet, *J. Theor. Biol.,* **160,** 97 (1993). Cellular Automata Approaches to Biological Modeling.
28. B. Chopard, M. Droz, and B. Chopard, *Cellular Automata Modeling of Physical Systems,* Cambridge University Press, Cambridge, UK, 1998.
29. L. B. Kier and B. Testa, *Adv. Drug Res.,* **26,** 1 (1995). Complexity and Emergence in Drug Research.
30. A. Geiger, F. H. Stillinger, and A. Rahman, *J. Phys. Chem.,* **70,** 4186 (1979). Aspects of the Percolation Process for Hydrogen Bonded Networks in Water.
31. I. Ohmine and H. Tanaka, *Chem. Rev.,* **93,** 2545 (1993). Fluctuation, Relaxations and Hydration in Liquid Water, Hydrogen Bond Rearrangement Dynamics.
32. M. Mezei and D. L. Beveridge, *J. Chem. Phys.,* **74,** 622 (1981) Theoretical Studies of Hydrogen Bonding in Liquid Water and Aqueous Solutions.
33. P. L. M. Plummer, *J. Mol. Struct.,* **23,** 47 (1991). Molecular Dynamics Simulations and Quantum Mechanical Studies of the Hydrogen Bond in Water Cluster Systems.
34. P. J. Rossky and M. Karplus, *J. Am. Chem. Soc.,* **101,** 1913 (1979). Solvation. A Molecular Dynamics Study of a Dipeptide in Water.
35. S. A. Rice and M. G. Skeats, *J. Phys. Chem.,* **85,** 1108 (1981). A Random Network Model for Water.
36. L. B. Kier and C.-K. Cheng, *J. Chem. Inf. Comput. Sci.,* **34,** 647 (1994). A Cellular Automata Model of Water.

37. R. L. Blumberg, G. Shlifer, and H. E. Stanley, *J. Phys. A: Math. Gen.,* **13,** L 147 (1980). Monte Carlo Tests of the Universality in a Correlated-Site Percolation Problem.
38. M. H. Brodsky and P. H. Leary, *Bull. Am. Phys. Soc.,* **25,** 260 (1980). The Temperature Dependence of the Refractive Index of Hydrogenated Amorphous Silicon and Implications for Electroreflectance.
39. G. E. Walrafen, in *Structure of Water, Hydrogen-Bonded Systems* A. K. Covington and P. Jones, Eds., Taylor and Francis, London, UK, 1968, pp. 173–175. Structure of Water.
40. G. H. Haggis, J. B. Hasted, and T. J. Buchanan, *J. Phys. Chem.*, **54,** 1452 (1952). The Dielectric Properties of Water in Solutions.
41. W. A. P. Luck, *Acta Chim. Hung. Acad. Sci.,* **121,** 119 (1986). Role of Hydrogen Bonding in the Structure of Liquids.
42. L. B. Kier, C.-K. Cheng, and B. Testa, *Pharm. Res.,* **12,** 1521 (1995). A Cellular Automata Model of Dissolution.
43. L. B. Kier and C.-K. Cheng, *J. Math. Chem.*, **21,** 71 (1997). A Cellular Automata Model of the Soluble State.
44. L. B. Kier, C.-K. Cheng, and B. Testa, *Pharm. Res.,* **12,** 615 (1995). A Cellular Automata Model of the Hydrophobic Effect.
45. W. L. Jorgensen, J. Gao, and C. Ravimohan, *J. Phys. Chem.,* **89,** 3470 (1985). Monte Carlo Simulations of Alkanes in Water: Hydration Numbers and Hydrophobic Effect.
46. W. Blokzijl and J. B. F. N. Engberts, *Angew. Chem. Int. Ed. Engl.,* **32,** 1545 (1993). Hydrophobic Effects, Opinions and Facts.
47. G. Hummer, S. Garde, A. E. Garcia, A. Pohorille, and L. R. Pratt, *Proc. Natl. Acad. Sci. USA,* **93,** 8951 (1996). An Information Theory Model of Hydrophobic Interaction.
48. B. J. Berne, *Proc. Natl. Acad. Sci. USA,* **93,** 8800 (1996). Inferring the Hydrophobic Interaction from the Properties of Neat Water.
49. L. B. Kier, C.-K. Cheng, and B. Testa, *J. Pharm. Sci.,* **86,** 774 (1997). A Cellular Automata Model of Diffusion in Aqueous Systems.
50. C. M. Gary-Bobo and H. W. Weber, *J. Phys. Chem.,* **73,** 1155 (1969). Diffusion of Alcohols and Amides in Water from 4 to 37°.
51. S. B. Horowitz and I. R. Fenichel, *J. Phys. Chem.,* **68,** 3378 (1964). Solute Diffusional Specificity in Hydrogen Bonding Systems.
52. C.-K. Cheng and L. B. Kier, *J. Chem. Inf. Comput. Sci.,* **35,** 1054 (1995). A Cellular Automata Model of Oil–Water Partitioning.
53. L. B. Kier and C.-K. Cheng, in *Lipophilicity in Drug Action and Toxicity,* V. Pliska, B. Testa, and H. van de Waterbeemd, Eds., VCH Publishers, Weinheim, Germany, 1996, pp. 187–192. A Cellular Automata Model of Partitioning Between Liquid Phase.
54. I. Benjamin, *J. Phys. Chem.,* **97,** 1432 (1992). Theoretical Studies of Water/1,2-Dichloroethane Interface: Structure, Dynamics and Conformational Equilibria at the Liquid–Liquid Interface.
55. I. L. Carpenter and W. J. Hebre, *J. Chem. Phys.*, **94,** 531 (1990). A Molecular Dynamics Study of the Hexane/Water Interface.
56. J. Gao and W. L. Jorgensen, *J. Phys. Chem.,* **92,** 5813 (1988). Theoretical Examination of Hexanol/Water Interfaces.
57. P. Luise, *J. Chem. Phys.*, **86,** 4177 (1987). Monte Carlo Simulation of Liquid–Liquid Benzene/Water Interface.
58. M. Meyer, M. Mareschal, and M. Hayoun, *J. Chem. Phys.,* **89,** 1067 (1988). Computer Modeling of a Liquid Interface.
59. B. Smit, P. A. J. Hilbers, K. Esselink, L. A. M. Rubert, N. M. van Os, and A. Schlijper, *J. Phys. Chem.,* **59,** 6361 (1991). Structure of a Water/Oil Interface in the Presence of Micelles: A Computer Simulation Study.

60. M. N. Jones and D. Chapman, *Micelles, Monolayers, Biomembranes,* Wiley, New York, 1994.
61. P. Mukergee and K. J. Mysels, *Critical Micelle Concentration in Aqueous Surfactant Systems,* National Bureau of Standards, Washington, DC, 1971.
62. L. B. Kier, C.-K. Cheng, and B. Testa, *Pharm. Res.,* **13,** 1419 (1996). A Cellular Automata Model of Micelle Formation.
63. L. Espada, M. N. Jones, and G. Pilcher, *J. Chem. Thermo.,* **12,** 1 (1970). Enthalpy of Micellization of *d*-Dodecyltrimethyl Ammonium Bromide.
64. T. Ingram and M. N. Jones, *J. Chem. Soc.,* **65,** 297 (1969). Membrane Potential Studies on Surfactant Solution.
65. L. B. Kier and C.-K. Cheng, *J. Theor. Biol.,* **186,** 75 (1997). A Cellular Automata Model of Membrane Permeability.
66. A. Walter and J. Gutknect, *J. Membr. Biol.,* **90,** 207 (1986). The Movement of Molecules Across Lipid Membranes: A Molecular Theory.
67. L. B. Kier, C.-K. Cheng, M. Tute, and P. G. Seybold, *J. Chem. Inf. Comput. Sci.,* **38,** 271 (1998). A Cellular Automata Model of Acid Dissociation.
68. G. B. Barlin and D. D. Perrin *Quart. Rev. Chem. Soc.,* **20,** 75 (1966). Prediction of the Strength of Organic Acids.
69. L. B. Kier, C.-K. Cheng, and B. Testa, *J. Chem. Inf. Comput. Sci.,* **39,** 32 (1999). A Cellular Automata Model of the Percolation Process.
70. P. Seybold, L. B. Kier, and C.-K. Cheng, *J. Chem. Inf. Comput. Sci.,* **37** 386 (1997). Simulation of First-Order Chemical Kinetics Using Cellular Automata.
71. L. Sklar, *Physics and Chance,* Cambridge University Press, New York, 1993.
72. P. Atkins, *Physical Chemistry,* 6th ed., Freeman, New York, 1998, pp. 778–783.
73. A. Neuforth, P. G. Seybold, L. B. Kier, and C.-K. Cheng, *Int. J. Chem. Kinetics,* **32,** 529 (2000). Cellular Automata Models of Kinetically and Thermodynamically Controlled Reactions.
74. J. D. Roberts and M. C. Caserio, *Basic Principles of Organic Chemistry,* 2nd ed., Benjamin, New York, 1977, pp. 374–376.
75. K.-C. Lin, *J. Chem. Educ.,* **65,** 857 (1988). Understanding Product Optimization.
76. G. N. Vriens, *Industr. Eng. Chem.,* **46,** 669 (1954). Kinetics of Coupled Reversible Reactions.
77. A. P. Gelbein, *CHEMTECH,* Sept., 1998, p. 1. Thinking Out of the Box.
78. M. E. Brown, K. J. Buchanan, and A. Goosen, *J. Chem. Educ.,* **62,** 575 (1985). Thermodynamically and Kinetically Controlled Products.
79. P. G. Seybold, L. B. Kier, and C.-K. Cheng, *J. Phys. Chem.,* **102,** 886 (1998). Stochastic Cellular Automata Models of Molecular Excited-State Dynamics.
80. P. G. Seybold, L. B. Kier, and C.-K. Cheng, *Int. J. Quantum Chem.,* **75,** 751 (1999). Aurora Borealis: Stochastic Cellular Automata. Simulations of the Excited-State Dynamics of Oxygen Atoms.
81. S. L. Guberman, *Science,* **278,** 1276 (1997). Mechanism for the Green Glow of the Upper Atmosphere.
82. D. Kella, L. Vejby-Christiansen, P. J. Johnson, H. B. Peterson, and L. H. Andersen, *Science,* **276,** 1530 (1997). The Source of Green Light Emission Determined from a Heavy-Atom Storage Ring Experiment.
83. H. Okabe, *Photochemistry of Small Molecules,* Wiley, New York, 1978, p. 370.
84. R. Kapral and K. Showalter, Eds., *Chemical Waves and Patterns,* Kluwer, Dordrecht, The Netherlands, 1995.
85. R. Kapral and X.-G. Wu, J. *Phys. Chem.,* **100,** 18976 (1996). Stochastic Description of Temporal and Spatial Dynamics of Far-From Equilibrium Reactions.

86. V. K. Vanag, *J. Phys. Chem.*, **100,** 11336 (1996). Probability Cellular Automaton-Aided Modeling of the Stirring Effect in the Autocatalytic Step of the Belousov–Zhabotinsky Reaction.

87. B. Chopard and M. Droz, *Cellular Automata Modeling of Physical Systems,* Cambridge University Press, Cambridge, UK, 1998.

88. B. Chopard and M. Droz, *Europhysics Lett.,* **15,** 459 (1991). Microscopic Study of the Properties of the Reaction Front in an $A + B \rightarrow C$ Reaction-Diffusion System.

89. S. Cornell, M. Droz, and B. Chopard, *Phys. Rev. A,* **44,** 4826 (1991). Role of Fluctuations for Inhomogeneous Reaction-Diffusion Phenomena.

90. S. Cornell, M. Droz, and B. Chopard, *Physica A,* **188,** 322 (1992). Some Properties of the Diffusion-Limited Reaction $nA + mB \rightarrow C$ with Homogeneous and Inhomogeneous Initial Conditions.

91. H. S. Berryman and D. R. Franceschetti, *Phys. Lett.* A, **136,** 348 (1989). Simulation of Diffusion Controlled Reaction Kinetics Using Cellular Automata.

92. J. P. Moore, P. G. Seybold, L. B. Kier, and C.-K. Cheng, to be published. Second Order Reactions Simulated by Cellular Automata.

93. L. B. Kier, C.-K. Cheng, and B. Testa, *J. Mol. Graphics,* **14,** 227 (1996). A Cellular Automata Model of Enzyme Kinetics.

94. L. B. Kier and C.-K. Cheng, *J. Mol. Graphics Modelling,* **18,** 29 (2000). A Cellular Automata Model of an Anticipatory System.

95. L. B. Kier, C.-K. Cheng, and T. Karnes, *J. Biol. Chromatogr.,* **14,** 530 (2000). A Cellular Automata Model of Chromatographic Separation.

96. L. B. Kier, C.-K. Cheng, and P. G. Seybold, *SAR QSAR Environ. Res.,* **11,** 79 (2000). Cellular Automata Models of Chemical Systems.

97. S. A. Levin, B. Grenfell, A. Hastings, and A. S. Perelson, *Science,* **275,** 335 (1997). Mathematical and Computational Challenges in Popululation Biology and Ecosystem Science.

98. J. H. Holland. *Emergence,* Perseus Books, Reading, MA, 1998.