

Preface

As I update parts of this textbook seven years after the original edition, I find the progress in the field to be overwhelming, almost unfitting to justify maintaining the same book. In fact, the sports analogy “Bigger, faster, stronger” seems most appropriate to the field of biomolecular modeling. Indeed, as modeling and simulation are used to probe more biological and chemical processes — with improved force fields and algorithms and faster computational platforms — new discoveries are being made that help interpret as well as extend experimental data. To experimentalists and theoreticians alike, modeling remains a valuable, albeit challenging, tool for probing numerous conformational, dynamic, and thermodynamic questions. We can certainly anticipate more exciting developments in biomolecular modeling as the first decade of this new century has ended and another began. At the same time, we should be reminded by the wisdom of the great French mathematician and scientist Pierre Simon de Laplace, who I quote more than once in this text, who also said: “*Ce que nous connaissons est peu de chose; ce que nous ignorons est immense*”. (What we know is little; what we do not know is immense).

Besides small additions and revisions made throughout the text and displayed materials to reflect the latest literature and field developments, some chapters have undergone more extensive revisions for this second edition. These include Chapters 1 and 2 that provide a historical perspective and an overview of current applications to biomolecular systems; Chapter 4, which reflects modified protein classification with new protein examples and sequence statistics; the chapter Topics in Nucleic Acids (now expanded into two chapters, 6 and 7), which includes recent developments in RNA structure and function; the force field chapters 8–10, which contain new sections on enhanced sampling methods; Chapter 15, which

includes an update on pharmacogenomics developments; and Appendices B and C which list key papers in the field and reference books, respectively.

As in the original book, the focus is on a broad and critical introduction to the field rather than a comprehensive view, though some algorithmic topics are presented in more depth. There are many books now since the first edition was written that provide more details on various aspects of biomolecular modeling and simulation (see Appendix C).

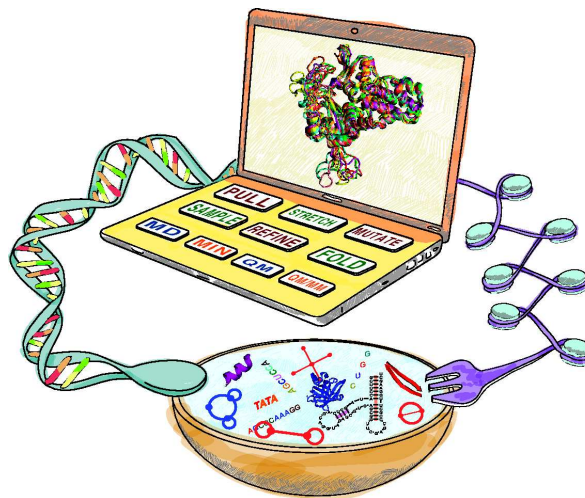
I would like to thank my many lab members and colleagues who have contributed to this effort, by providing scientific and technical information, making figures, and/or reading various versions of this text, including Lisa Chase, Rosana Collepardo, Ron Dror, Shereef Elmetwaly, Meredith Foley, Joachim Frank, Hin Hark Gan, Joe Izzo, Namhee Kim, Itzhak Krinsky, Christian Laing, Pierre L'Ecuyer, Connie Lee, Rubisco Li, Michael Overton, Vijay Pande, Ogi Perisic, Giulio Quarta, Klaus Schulten, Rick Solway, James Van Arsdale, Arieh Warshel, Michael Watters, Ada Yonath, and Yingkai Zhang.

As before, I invite readers to share their comments and thoughts with me directly via email; I enjoy reading them all.

Tamar Schlick

New York, New York

May 14, 2010



Preface to the 2002 Edition

Science is a way of looking, reverencing. And the purpose of all science, like living, which amounts to the same thing, is not the accumulation of gnostic power, the fixing of formulas for the name of God, the stockpiling of brutal efficiency, accomplishing the sadistic myth of progress. The purpose of science is to revive and cultivate a perpetual state of wonder. For nothing deserves wonder so much as our capacity to experience it.

Roald Hoffman and Shira Leibowitz Schmidt, in *Old Wine, New Flasks: Reflections on Science and Jewish Tradition* (W.H. Freeman, 1997).

Challenges in Teaching Molecular Modeling

This textbook evolved from a graduate course termed *Molecular Modeling* introduced in the fall of 1996 at New York University. The primary goal of the course is to stimulate excitement for molecular modeling research — much in the spirit of Hoffman and Leibowitz Schmidt above — while providing grounding in the discipline. Such knowledge is valuable for research dealing with many practical problems in both the academic and industrial sectors, from developing treatments for AIDS (via inhibitors to the protease enzyme of the human immunodeficiency virus, HIV-1) to designing potatoes that yield spot-free potato chips (via transgenic potatoes with altered carbohydrate metabolism). In the course of writing this text, the notes have expanded to function also as an introduction to the field for scientists in other disciplines by providing a global perspective into problems and approaches, rather than a comprehensive survey.

As a textbook, my intention is to provide a framework for teachers rather than a rigid guide, with material to be supplemented or substituted as appropriate for the audience. As a reference book, scientists who are interested in learning about biomolecular modeling may view the book as a broad introduction to an exciting new field with a host of challenging, interdisciplinary problems.

The intended audience for the course is beginning graduate students in medical schools and in all scientific departments: biology, chemistry, physics, mathematics, computer science, and others. This interdisciplinary audience presents a special challenge: it requires a broad presentation of the field but also good coverage of specialized topics to keep experts interested. Ideally, a good grounding in basic biochemistry, chemical physics, statistical and quantum mechanics, scientific computing (i.e., numerical methods), and programming techniques is desired. The rarity of such a background required me to offer tutorials in both biological and mathematical areas.

The introductory chapters on biomolecular structure are included in this book (after much thought) and are likely to be of interest to physical and mathematical

scientists. Chapters 3 and 4 on proteins, together with Chapters 5 and 6 on nucleic acids, are thus highly abbreviated versions of what can be found in numerous texts specializing in these subjects. The selections in these tutorials also reflect some of my group's areas of interest. Because many introductory and up-to-date texts exist for protein structure, only the basics in protein structure are provided, while a somewhat more expanded treatment is devoted to nucleic acids.

Similarly, the introductory material on mathematical subjects such as basic optimization theory (Chapter 11) and random number generators (Chapter 12) is likely to be of use more to readers in the biological / chemical disciplines. General readers, as well as course instructors, can skip around this book as appropriate and fill in necessary gaps through other texts (e.g., in protein structure or programming techniques).

Text Limitations

By construction, this book is very broad in scope and thus no subjects are covered in great depth. References to the literature are only representative. The material presented is necessarily selective, unbalanced in parts, and reflects some of my areas of interest and expertise. This text should thus be viewed as an attempt to introduce the discipline of molecular modeling to students and to scientists from disparate fields, and should be taken together with other related texts, such as those listed in Appendix C, and the representative references cited.

The book format is somewhat unusual for a textbook in that it is nonlinear in parts. For example, protein folding is introduced early (before protein basics are discussed) to illustrate challenging problems in the field and to interest more advanced readers; the introduction to molecular dynamics incorporates illustrations that require more advanced techniques for analysis; some specialized topics are also included throughout. For this reason, I recommend that students re-read certain parts of the book (e.g., first two chapters) after covering others (e.g., the biomolecular tutorial chapters). Still, I hope most of all to grab the reader's attention with exciting and current topics.

Given the many caveats of introducing and teaching such a broad and interdisciplinary subject as molecular modeling, the book aims to introduce selected biomolecular modeling and simulation techniques, as well as the wide range of biomolecular problems being tackled with these methods. Throughout these presentations, the central goal is to develop in students a good understanding of the inherent approximations and errors in the field so that they can adequately assess modeling results. Diligent students should emerge with basic knowledge in modeling and simulation techniques, an appreciation of the fundamental problems — such as force field approximations, nonbonded evaluation protocols, size and timestep limitations in simulations — and a healthy critical eye for research. A historical perspective and a discussion of future challenges are also offered.

Dazzling Modeling Advances Demand Perspective

The topics I chose for this course are based on my own unorthodox introduction to the field of modeling. As an applied mathematician, I became interested in the field during my graduate work, hearing from Professor Suse Broyde — whose path I crossed thanks to Courant Professor Michael Overton — about the fascinating problem of modeling carcinogen/DNA adducts.

The goal was to understand some structural effects induced by certain compounds on the DNA (deduced by energy minimization); such alterations can render DNA more sensitive to replication errors, which in turn can eventually lead to mutagenesis and carcinogenesis. I had to roam through many references to obtain a grasp of some of the underlying concepts involving force fields and simulation protocols, so many of which seemed so approximate and not fully physically grounded. By now, however, I have learned to appreciate the practical procedures and compromises that computational chemists have formulated out of sheer necessity to obtain answers and insights into important biological processes that cannot be tackled by instrumentation. In fact, approximations and simplifications are not only tolerated when dealing with biomolecules; they often lead to insights that cannot easily be obtained from more detailed representations. Furthermore, it is often the neglect of certain factors that teaches us their importance, sometimes in subtle ways.

For example, when Suse Broyde and I viewed in the mid 1980s her intriguing carcinogen/modified DNA models, we used a large Evans and Sutherland computer while wearing special stereoviewers; the hard-copy drawings were ball and stick models, though the dimensionality projected out nicely in black and white. (Today, we still use stereo glasses, but current hardware stereo capabilities are much better, and marvelous molecular renderings are available). At that time, only small pieces of DNA could be modeled, and the surrounding salt and solvent environment was approximated. Still, structural and functional insights arose from those earlier works, many of which were validated later by more comprehensive computation, as well as laboratory experiments.

Book Overview

The book provides an overview of three broad topics: (a) biomolecular structure and modeling: current problems and state of computations (Chapters 1–6); (b) molecular mechanics: force field origin, composition, and evaluation techniques (Chapters 8–10); and (c) simulation techniques: conformational sampling by geometry optimization, Monte Carlo, and molecular dynamics approaches (Chapters 11–14). Chapter 15 on the similarity and diversity problems in chemical design introduces some of the challenges in the growing field related to combinatorial chemistry (Chapter 15).

Specifically, Chapters 1 and 2 give a historical perspective of biomolecular modeling, outlining progress in experimental techniques, the current computational challenges, and the practical applications of this enterprise — to convey the

immense interest in, and support of, the discipline. *Since these chapters discuss rapidly changing subjects (e.g., genome projects, disease treatments), they will be updated as possible on the text website.* General readers may find these chapters useful as an introduction to biomolecular modeling and its applications.

Chapters 3 and 4 review the basic elements in protein structure, and Chapter 5 similarly presents the basic building blocks and conformational flexibility in nucleic acids. Chapter 6 presents additional topics in nucleic acids, such as DNA sequence effects, DNA/protein interactions, departures from the canonical DNA helix forms, RNA structure, and DNA supercoiling.

The second part of the book begins in Chapter 8 with a view of the discipline of molecular mechanics as an offspring of quantum mechanics and discusses the basic premises of molecular mechanics formulations. A detailed presentation of the force field terms — origin, variation, and parameterization — is given in Chapter 9. Chapter 10 is then devoted to the computation of the nonbonded energy terms, including cutoff techniques, Ewald and multipole schemes, and continuum solvation alternatives.

The third part of the book, simulation algorithms,¹ begins with a description of optimization methods for multivariate functions in Chapter 11, emphasizing the tradeoff between algorithm complexity and performance. Basic issues of Monte Carlo techniques, appropriate to a motivated novice, are detailed in Chapter 12, such as pseudorandom number generators, Gaussian random variates, Monte Carlo sampling, and the Metropolis algorithm. Chapters 13 and 14 describe the algorithmic challenges in biomolecular dynamics simulations and present various categories of integration techniques, from the popular Verlet algorithm to multiple-timestep techniques and Brownian dynamics protocols. Chapter 15 outlines the challenges in similarity and diversity sampling in the field of chemical design, related to the new field of combinatorial chemistry.

The book appendices complement the material in the main text through homework assignments, reading lists, and other information useful for teaching molecular modeling.

Instructors may find the sample course syllabus in Appendix A helpful. Important also to teaching is an introduction to the original literature; a representative reading list of articles used for the course is collected in Appendix B. An annotated general reference list is given in Appendix C.

Selected biophysics applications are highlighted through the homework assignments (Appendix D). Humor in the assignments stimulates creativity in many students. These homeworks are a central component of learning molecular modeling, as they provide hands-on experience, extend upon subjects covered in the chapters, and expose the students to a wide range of current topics in biomolec-

¹The word *algorithm* is named after the ninth-century Persian (Iranian in present-day terminology) mathematician al-Khwarizmi (nicknamed after his home town of Khwarizm, now Khiva in the Uzbek Republic), who stressed the importance of methodical procedures for solving problems in his algebra textbook. The term has evolved to mean the systematic process of solving problems by machine execution.

ular structure. Advanced students may use these homework assignments to learn about molecular modeling through independent research.

Many homework assignments involve a molecular modeling software package. I selected the *Insight* program in conjunction with our Silicon Graphics computer laboratory, but other suitable modeling programs can be used. Students also learn other basic research tools (such as programming and literature searches) through the homeworks.

Our memorable “force field debate” (see homework 7 in Appendix D) even brought the AMBER team to class in white lab coats, each accented with a name tag corresponding to one of AMBER’s original authors. The late Peter Kollman would have been pleased. Harold Scheraga would have been no less impressed by the long list of ECEPP successes prepared by his loyal troopers. Martin Karplus would not have been disappointed by the strong proponents of the CHARMM approach. I only hope to have as much spunk and talent in my future molecular modeling classes.

Extensive use of web resources is encouraged, while keeping in mind the caveat of lack of general quality control. I was amazed to find some of my students’ discoveries regarding interesting molecular modeling topics mentioned in the classroom, especially in the context of the term project, which requires them to find outstanding examples of the successes and/or failures of molecular modeling.

Interested readers might also want to glance at additional course information as part of my group’s home page, monod.biomath.nyu.edu/. Supplementary text information (such as program codes and figure files) can also be obtained.

To future teachers of molecular modeling who plan to design similar assignments and material, I share with you my following experience regarding student reactions to this discipline: what excited students the most about the subject matter and led to enthusiasm and excellent feedback in the classroom were the rapid pace at which the field is developing, its exciting discoveries, and the medical and technological breakthroughs made possible by important findings in the field.

In more practical terms, a mathematics graduate student, Brynja Kohler, expressed this enthusiasm succinctly in the introduction to her term project:

As I was doing research for this assignment, I found that one interesting article led to another. Communication via e-mail with some researchers around the world about their current investigations made me eagerly anticipate new results. The more I learned the more easy it became to put off writing a final draft because my curiosity would lead me on yet another line of inquiry. However, alas, there comes a time when even the greatest procrastinator must face the music, and evaluate what it is that we know and not linger upon what we hope to find out.

Future teachers are thus likely to have an enjoyable experience with any good group of students.

Acknowledgments

I am indebted to Jing Huang for her devoted assistance with the manuscript preparation, file backups, data collection, and figure design. I also thank Wei Xu and Mulin Ding for important technical assistance. I am grateful to my other devoted current and former group members who helped read book segments, collect data, prepare the figures found throughout this book, and run to libraries throughout New York City often: Karunesh Arora, Danny Barash, Paul Batcho, Dan Beard, Mulin Ding, Hin Hark Gan, Jennifer Isbell, Joyce Noah, Xiaoliang Qian, Sonia Rivera, Adrian Sandu, Dan Strahs, Dexuan Xie, Linjing Yang, and Qing Zhang. Credits for each book figure and table are listed on the text's website.

I thank my colleagues Ruben Abagyan, Helen Berman, Dave Case, Jonathan Goodman, Andrej Sali, and Harold Scheraga, who gave excellent guest lectures in the course; and my course assistants Karunesh Arora, Margaret Mandziuk, Qing Zhang, and Zhongwei Zhu for their patient, dedicated assistance to the students with their homework and queries.

I am also very appreciative of the following colleagues for sharing reprints, information, and unpublished data and/or for their willingness to comment on segments of the book: Lou Allinger, Nathan Baker, Mike Beer, Helen Berman, Suse Broyde, John Board, Dave Beveridge, Ken Breslauer, Steve Burley, Dave Case, Philippe Derreumaux, Ron Elber, Eugene Fluder, Leslie Greengard, Steve Harvey, Jan Hermans, the late Peter Kollman, Robert Krasny, Michael Levitt, Xiang-Jun Lu, Pierre L'Ecuyer, Neocles Leontis, the late Shneior Lifson, Kenny Lipkowitz, Jerry Manning, Andy McCammon, Mihaly Mezei, Jorge Nocedal, Wilma Olson, Michael Overton, Vijay Pande, Dinshaw Patel, Harold Scheraga, Shulamith Schlick, Klaus Schulten, Suresh Singh, Bob Skeel, A. R. Srinivasan, Emad Tajkhorshid, Yuri Ushkaryov, Wilfred van Gunsteren, Arie Warshel, Eric Westhof, Weitao Yang, and Darren York. Of special note are the extremely thorough critiques which I received from Lou Allinger, Steve Harvey, Jerry Manning, Robert Krasny, Wilma Olson, and Bob Skeel; their extensive comments and suggestions led to enlightening discussions and helped me see the field from many perspectives. I thank my colleague and friend Suse Broyde for introducing me to the field and for reading nearly every page of this book's draft.

To my family — parents Haim and Shula, sisters Yael and Daphne, aunt Cecilia, and especially Rick and Duboni — I am grateful for tolerating my long months on this project.

Finally, I thank my excellent students for making the course enjoyable and inspiring.

Tamar Schlick
New York, New York
June 10, 2002