

Tamar Schlick

Molecular Modeling and Simulation

An Interdisciplinary Guide

With 147 Full-Color Illustrations



Springer

Tamar Schlick
Department of Mathematics and Chemistry
Courant Institute of Mathematical Sciences
New York University
New York, NY 10012
USA
schlick@nyu.edu

Editors

J.E. Marsden
Control and Dynamical Systems
Mail Code 107-81
California Institute of Technology
Pasadena, CA 91125
USA
marsden@cds.caltech.edu

S. Wiggins
School of Mathematics
University of Bristol
Bristol, BS8 1TW
United Kingdom
s.wiggins@bristol.ac.uk

L. Sirovich
Division of Applied Mathematics
Brown University
Providence, RI 02912
USA
chico@camelot.mssm.edu

S.S. Antman
Department of Mathematics
and
Institute for Physical Science and Technology
University of Maryland
College Park, MD 20742-4015
USA
ssa@math.umd.edu

Cover illustration: © Wayne Thiebaud/Licensed by VAGA, New York, NY. Courtesy of the Allan Stone Gallery, NYC.

Mathematics Subject Classification (2000): 92–01, 92Exx, 92C40

Library of Congress Cataloging-in-Publication Data

Schlick, Tamar.

Molecular modeling and simulation: an interdisciplinary guide / Tamar Schlick.

p. cm. — (Interdisciplinary applied mathematics ; 21)

Includes bibliographical references and index.

ISBN 978-1-4757-5893-1

ISBN 978-0-387-22464-0 (eBook)

DOI 10.1007/978-0-387-22464-0

1. Biomolecules—Models. 2. Biomolecules—Models—Computer simulation. I. Title.
II. Interdisciplinary applied mathematics ; v. 21.

QD480 .S37 2002

572'.33'015118—dc21

2002016003

ISBN 978-1-4757-5893-1

Printed on acid-free paper.

© 2002 Springer Science+Business Media New York

Originally published by Springer-Verlag New York, Inc. in 2002

Softcover reprint of the hardcover 1st edition 2002

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher Springer Science+Business Media, LLC , except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use in this publication of trade names, trademarks, service marks and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

9 8 7 6 5 4 3 2

Typesetting: Pages created by author using a Springer T_EX macro package.

www.springer-ny.com

Contents

About the Cover	v
Book URLs	ix
Preface	xi
Prelude	xix
List of Figures	xxxix
List of Tables	xxxviii
Acronyms, Abbreviations, and Units	xli
1 Biomolecular Structure and Modeling: Historical Perspective	1
1.1 A Multidisciplinary Enterprise	2
1.1.1 Consilience	2
1.1.2 What is Molecular Modeling?	3
1.1.3 Need For Critical Assessment	4
1.1.4 Text Overview	6
1.2 Molecular Mechanics	7
1.2.1 Pioneers	7
1.2.2 Simulation Perspective	10
1.3 Experimental Progress	12
1.3.1 Protein Crystallography	12

1.3.2	DNA Structure	14
1.3.3	Crystallography	16
1.3.4	NMR Spectroscopy	18
1.4	Modern Era	19
1.4.1	Biotechnology	19
1.4.2	PCR and Beyond	20
1.5	Genome Sequencing	23
1.5.1	Sequencing Overview	23
1.5.2	Human Genome	27
2	Biomolecular Structure and Modeling: Problem and Application Perspective	33
2.1	Computational Challenges	33
2.1.1	Bioinformatics	33
2.1.2	Structure From Sequence	35
2.2	Protein Folding	37
2.2.1	Folding Views	37
2.2.2	Folding Challenges	39
2.2.3	Folding Simulations	40
2.2.4	Chaperones	42
2.2.5	Unstructured Proteins	42
2.3	Protein Misfolding	44
2.3.1	Prions	44
2.3.2	Infectious Proteins?	44
2.3.3	Hypotheses	45
2.3.4	Other Misfolding Processes	46
2.3.5	Function From Structure	47
2.4	Practical Applications	47
2.4.1	Drug Design	48
2.4.2	AIDS Drugs	49
2.4.3	Other Drugs	53
2.4.4	A Long Way To Go	54
2.4.5	Better Genes	54
2.4.6	Designer Foods	56
2.4.7	Designer Materials	59
2.4.8	Cosmeceuticals	59
3	Protein Structure Introduction	61
3.1	Machinery of Life	61
3.1.1	From Tissues to Hormones	61
3.1.2	Size and Function Variability	62
3.1.3	Chapter Overview	63
3.2	Amino Acid Building Blocks	66
3.2.1	Basic C ^α Unit	66
3.2.2	Essential and Nonessential Amino Acids	67

3.2.3	Linking Amino Acids	69
3.2.4	The Amino Acid Repertoire	72
3.3	Sequence Variations in Proteins	74
3.3.1	Globular Proteins	74
3.3.2	Membrane and Fibrous Proteins	75
3.3.3	Emerging Patterns from Genome Databases	76
3.3.4	Sequence Similarity	77
3.4	Protein Conformation Framework	80
3.4.1	The Flexible ϕ and ψ and Rigid ω Dihedral Angles	80
3.4.2	Rotameric Structures	84
3.4.3	Ramachandran Plots	84
3.4.4	Conformational Hierarchy	86
4	Protein Structure Hierarchy	91
4.1	Structure Hierarchy	92
4.2	Helices	92
4.2.1	Classic α -Helix	92
4.2.2	3_{10} and π Helices	93
4.2.3	Left-Handed α -Helix	96
4.2.4	Collagen Helix	96
4.3	β -Sheets: A Common Secondary Structural Element	96
4.4	Turns and Loops	96
4.5	Supersecondary and Tertiary Structure	99
4.5.1	Complex 3D Networks	99
4.5.2	Classes in Protein Architecture	99
4.5.3	Classes are Further Divided into Folds	100
4.6	α -Class Folds	100
4.6.1	Bundles	100
4.6.2	Folded Leafs	101
4.6.3	Hairpin Arrays	101
4.7	β -Class Folds	102
4.7.1	Anti-Parallel β Domains	102
4.7.2	Parallel and Antiparallel Combinations	103
4.8	α/β and $\alpha+\beta$ -Class Folds	103
4.8.1	α/β Barrels	104
4.8.2	Open Twisted α/β Folds	104
4.8.3	Leucine-Rich α/β Folds	104
4.8.4	$\alpha+\beta$ Folds	104
4.9	Number of Folds	105
4.9.1	Finite Number?	105
4.9.2	Concerted Target Selection: Structural Genomics	105
4.10	Quaternary Structure	106
4.10.1	Viruses	106
4.10.2	From Ribosomes to Dynamic Networks	110
4.11	Structure Classification	111

5	Nucleic Acids Structure Minitutorial	113
5.1	DNA, Life's Blueprint	114
5.1.1	The Kindled Field of Molecular Biology	114
5.1.2	DNA Processes	116
5.1.3	Challenges in Nucleic Acid Structure	117
5.1.4	Chapter Overview	118
5.2	Basic Building Blocks	118
5.2.1	Nitrogenous Bases	119
5.2.2	Hydrogen Bonds	119
5.2.3	Nucleotides	119
5.2.4	Polynucleotides	121
5.2.5	Stabilizing Polynucleotide Interactions	122
5.2.6	Chain Notation	124
5.2.7	Atomic Labeling	125
5.2.8	Torsion Angle Labeling	125
5.3	Conformational Flexibility	126
5.3.1	The Furanose Ring	126
5.3.2	Backbone Torsional Flexibility	131
5.3.3	The Glycosyl Rotation	131
5.3.4	Sugar/Glycosyl Combinations	131
5.3.5	Basic Helical Descriptors	133
5.3.6	Base-Pair Parameters	135
5.4	Canonical DNA Forms	139
5.4.1	B-DNA	141
5.4.2	A-DNA	142
5.4.3	Z-DNA	145
5.4.4	Comparative Features	146
6	Topics in Nucleic Acids Structure	147
6.1	Introduction	148
6.2	DNA Sequence Effects	149
6.2.1	Local Deformations	149
6.2.2	Orientation Preferences in Dinucleotide Steps	150
6.2.3	Intrinsic DNA Bending in A-Tracts	153
6.2.4	Sequence Deformability Analysis Continues	156
6.3	DNA Hydration and Ion Interactions	157
6.3.1	Resolution Difficulties	159
6.3.2	Basic Patterns	159
6.4	DNA/Protein Interactions	163
6.5	Variations on a Theme	165
6.5.1	Hydrogen Bonding Patterns in Polynucleotides	165
6.5.2	Hybrid Helical/Nonhelical Forms	171
6.5.3	Overstretched and Understretched DNA	173
6.6	RNA Structure	175
6.6.1	RNA Chains Fold Upon Themselves	175

6.6.2	RNA's Diversity	176
6.6.3	RNA at Atomic Resolution	176
6.6.4	Emerging Themes in RNA Structure and Folding	179
6.7	Cellular Organization of DNA	181
6.7.1	Compaction of Genomic DNA	181
6.7.2	Coiling of the DNA Helix Itself	182
6.7.3	Chromosomal Packaging of Coiled DNA	183
6.8	Mathematical Characterization of DNA Supercoiling	186
6.8.1	DNA Topology and Geometry	186
6.9	Computational Treatments of DNA Supercoiling	189
6.9.1	DNA as a Flexible Polymer	190
6.9.2	Elasticity Theory Framework	191
6.9.3	Simulations of DNA Supercoiling	192
7	Theoretical and Computational Approaches to Biomolecular Structure	199
7.1	Merging of Theory and Experiment	200
7.1.1	Exciting Times for Computationalists!	200
7.1.2	The Future of Biocomputations	202
7.1.3	Chapter Overview	202
7.2	QM Foundations	202
7.2.1	The Schrödinger Wave Equation	203
7.2.2	The Born-Oppenheimer Approximation	203
7.2.3	Ab Initio	204
7.2.4	Semi-Empirical QM	205
7.2.5	Recent Advances in Quantum Mechanics	205
7.2.6	From Quantum to Molecular Mechanics	207
7.3	Molecular Mechanics Principles	211
7.3.1	The Thermodynamic Hypothesis	211
7.3.2	Additivity	212
7.3.3	Transferability	214
7.4	Molecular Mechanics Formulation	217
7.4.1	Configuration Space	218
7.4.2	Functional Form	219
7.4.3	Some Current Limitations	222
8	Force Fields	225
8.1	Formulation of the Model and Energy	227
8.2	Normal Modes	227
8.2.1	Characteristic Motions	227
8.2.2	Spectra of Biomolecules	229
8.2.3	Spectra As Force Constant Sources	230
8.2.4	In-Plane and Out-of-Plane Bending	231
8.3	Bond Length Potentials	232
8.3.1	Harmonic Term	233

8.3.2	Morse Term	234
8.3.3	Cubic and Quartic Terms	236
8.4	Bond Angle Potentials	237
8.4.1	Harmonic and Trigonometric Terms	237
8.4.2	Cross Bond Stretch / Angle Bend Terms	239
8.5	Torsional Potentials	241
8.5.1	Origin of Rotational Barriers	241
8.5.2	Fourier Terms	242
8.5.3	Torsional Parameter Assignment	243
8.5.4	Improper Torsion	247
8.5.5	Cross Dihedral/Bond Angle and Improper/Improper Dihedral Terms	248
8.6	van der Waals Potential	249
8.6.1	Rapidly Decaying Potential	249
8.6.2	Parameter Fitting From Experiment	249
8.6.3	Two Parameter Calculation Protocols	250
8.7	Coulomb Potential	251
8.7.1	Coulomb's Law: Slowly Decaying Potential	251
8.7.2	Dielectric Function	253
8.7.3	Partial Charges	254
8.8	Parameterization	255
8.8.1	A Package Deal	255
8.8.2	Force Field Performance	256
9	Nonbonded Computations	259
9.1	Computational Bottleneck	261
9.2	Reducing Computational Cost	262
9.2.1	Simple Cutoff Schemes	262
9.2.2	Ewald and Multipole Schemes	263
9.3	Spherical Cutoff Techniques	264
9.3.1	Technique Categories	264
9.3.2	Guidelines for Cutoff Functions	265
9.3.3	General Cutoff Formulations	266
9.3.4	Potential Switch	268
9.3.5	Force Switch	269
9.3.6	Shift Functions	270
9.4	Ewald Method	271
9.4.1	Periodic Boundary Conditions	271
9.4.2	Ewald Sum and Crystallography	274
9.4.3	Morphing A Conditionally Convergent Sum	276
9.4.4	Finite-Dielectric Correction	280
9.4.5	Ewald Sum Complexity	280
9.4.6	Resulting Ewald Summation	281
9.4.7	Practical Implementation	283
9.5	Multipole Method	284

9.5.1	Basic Hierarchical Strategy	285
9.5.2	Historical Perspective	289
9.5.3	Expansion in Spherical Coordinates	291
9.5.4	Biomolecular Implementations	292
9.5.5	Other Variants	293
9.6	Continuum Solvation	293
9.6.1	Need for Simplification!	293
9.6.2	Potential of Mean Force	294
9.6.3	Stochastic Dynamics	295
9.6.4	Continuum Electrostatics	298
10	Multivariate Minimization in Computational Chemistry	305
10.1	Optimization Applications	306
10.1.1	Algorithmic Understanding Needed	307
10.1.2	Chapter Overview	307
10.2	Fundamentals	308
10.2.1	Problem Formulation	308
10.2.2	Independent Variables	308
10.2.3	Function Characteristics	309
10.2.4	Local and Global Minima	310
10.2.5	Derivatives	312
10.2.6	Hessian Matrix	313
10.3	Basic Algorithms	317
10.3.1	Greedy Descent	317
10.3.2	Line Searches	318
10.3.3	Trust Region Methods	321
10.3.4	Convergence Criteria	322
10.4	Newton's Method	323
10.4.1	Newton in One Dimension	324
10.4.2	Newton's Method for Minimization	327
10.4.3	Multivariate Newton	329
10.5	Large-Scale methods	329
10.5.1	Quasi-Newton (QN)	330
10.5.2	Conjugate Gradient (CG)	332
10.5.3	Truncated-Newton (TN)	334
10.5.4	Simple Example	336
10.6	Software	338
10.6.1	Popular Newton and CG	338
10.6.2	CHARMM's ABNR	338
10.6.3	CHARMM's TN	338
10.6.4	Comparative Performance on Molecular Systems	339
10.7	Recommendations	339
10.8	Future Outlook	342

11 Monte Carlo Techniques	345
11.1 Monte Carlo Popularity	346
11.1.1 A Winning Combination	346
11.1.2 From Needles to Bombs	347
11.1.3 Chapter Overview	347
11.1.4 Importance of Error Bars	348
11.2 Random Number Generators	348
11.2.1 What is <i>Random</i> ?	348
11.2.2 Properties of Generators?	349
11.2.3 Linear Congruential Generators	352
11.2.4 Other Generators	356
11.2.5 Artifacts	360
11.2.6 Recommendations	362
11.3 Gaussian Random Variates	363
11.3.1 Manipulation of Uniform Random Variables	363
11.3.2 Normal Variates in Molecular Simulations	363
11.3.3 Odeh/Evans	364
11.3.4 Box/Muller/Marsaglia	366
11.4 Monte Carlo Means	366
11.4.1 Expected Values	366
11.4.2 Error Bars	368
11.4.3 Batch Means	370
11.5 Monte Carlo Sampling	371
11.5.1 Probability Density Function	371
11.5.2 Equilibria or Dynamics	371
11.5.3 Ensembles	372
11.5.4 Importance Sampling	373
11.6 Hybrid MC	377
11.6.1 MC and MD	377
11.6.2 Basic Idea	378
11.6.3 Variants and Other Hybrid Approaches	379
12 Molecular Dynamics: Basics	383
12.1 Introduction	384
12.1.1 Why Molecular Dynamics?	384
12.1.2 Background	385
12.1.3 Outline of MD Chapters	388
12.2 Laplace's Vision	389
12.2.1 The Dream	389
12.2.2 Deterministic Mechanics	389
12.2.3 Neglect of Electronic Motion	389
12.2.4 Critical Frequencies	390
12.2.5 Electron/Nuclear Treatment	391
12.3 Basics	392
12.3.1 Following Motion	392

12.3.2	Trajectory Quality	393
12.3.3	Initial System Settings	394
12.3.4	Trajectory Sensitivity	396
12.3.5	Simulation Protocol	399
12.3.6	High-Speed Implementations	400
12.3.7	Analysis and Visualization	402
12.3.8	Reliable Numerical Integration	402
12.3.9	Computational Complexity	403
12.4	Verlet Algorithm	406
12.4.1	Position and Velocity Propagation	406
12.4.2	Leapfrog, Velocity Verlet, and Position Verlet	408
12.5	Constrained Dynamics	410
12.6	Various MD Ensembles	412
12.6.1	Ensemble Types	412
12.6.2	Simple Algorithms	413
12.6.3	Extended System Methods	416
13	Molecular Dynamics: Further Topics	419
13.1	Introduction	420
13.2	Symplectic Integrators	421
13.2.1	Symplectic Transformation	422
13.2.2	Harmonic Oscillator Example	422
13.2.3	Linear Stability	423
13.2.4	Timestep-Dependent Rotation in Phase Space	424
13.2.5	Resonance Condition for Periodic Motion	426
13.2.6	Resonance Artifacts	427
13.3	Multiple-Timestep (MTS) Methods	428
13.3.1	Basic Idea	428
13.3.2	Extrapolation	429
13.3.3	Impulses	430
13.3.4	Resonances in Impulse Splitting	431
13.3.5	Resonance Artifacts in MTS	431
13.3.6	Resonance Consequences	434
13.4	Langevin Dynamics	435
13.4.1	Uses	435
13.4.2	Heat Bath	435
13.4.3	Effect of γ	435
13.4.4	Generalized Verlet for Langevin Dynamics	437
13.4.5	LN Method	438
13.5	Brownian Dynamics (BD)	442
13.5.1	Brownian Motion	442
13.5.2	Brownian Framework	444
13.5.3	General Propagation Framework	446
13.5.4	Hydrodynamics	447
13.5.5	BD Propagation	450

13.6	Implicit Integration	452
13.6.1	Implicit vs. Explicit Euler	452
13.6.2	Intrinsic Damping	454
13.6.3	Computational Time	454
13.6.4	Resonance Artifacts	454
13.7	Future Outlook	459
13.7.1	Integration Ingenuity	459
13.7.2	Current Challenges	459
14	Similarity and Diversity in Chemical Design	463
14.1	Introduction to Drug Design	464
14.1.1	Chemical Libraries	464
14.1.2	Early Days	465
14.1.3	Rational Drug Design	467
14.1.4	Automated Technology	469
14.1.5	Chapter Overview	469
14.2	Database Problems	470
14.2.1	Database Analysis	470
14.2.2	Similarity and Diversity Sampling	471
14.2.3	Bioactivity	473
14.3	General Problem Definitions	475
14.3.1	The Dataset	475
14.3.2	The Compound Descriptors	475
14.3.3	Biological Activity	478
14.3.4	The Target Function	479
14.3.5	Scaling Descriptors	479
14.3.6	The Similarity and Diversity Problems	480
14.4	Data Compression and Cluster Analysis	482
14.4.1	PCA compression	483
14.4.2	SVD compression	485
14.4.3	PCA and SVD	487
14.4.4	Projection Application	488
14.4.5	Example	489
14.5	Future Perspectives	492
	Epilogue	497
	Appendix A. Molecular Modeling Sample Syllabus	499
	Appendix B. Article Reading List	501
	Appendix C. Supplementary Course Texts	505
	Appendix D. Homework Assignments	511
	Index	621