

# Contents

<b>About the Cover</b>	<b>v</b>
<b>Book URLs</b>	<b>ix</b>
<b>Preface</b>	<b>xi</b>
<b>Prelude</b>	<b>xix</b>
<b>Table of Contents</b>	<b>xxi</b>
<b>List of Figures</b>	<b>xxxii</b>
<b>List of Tables</b>	<b>xxxviii</b>
<b>Acronyms, Abbreviations, and Units</b>	<b>xli</b>
<b>1 Biomolecular Structure and Modeling: Historical Perspective</b>	<b>1</b>
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 . . . . .	5
1.1.4 Text Overview . . . . .	6
1.2 Molecular Mechanics . . . . .	8
1.2.1 Pioneers . . . . .	8
1.2.2 Simulation Perspective . . . . .	11

1.3	Experimental Progress . . . . .	14
1.3.1	Protein Crystallography . . . . .	14
1.3.2	DNA Structure . . . . .	17
1.3.3	Crystallography . . . . .	18
1.3.4	NMR Spectroscopy . . . . .	20
1.4	Modern Era . . . . .	22
1.4.1	Biotechnology . . . . .	22
1.4.2	PCR and Beyond . . . . .	23
1.5	Genome Sequencing . . . . .	25
1.5.1	Sequencing Overview . . . . .	25
1.5.2	Human Genome . . . . .	30
<b>2</b>	<b>Biomolecular Structure and Modeling: Problem and Application Perspective</b>	<b>41</b>
2.1	Computational Challenges . . . . .	41
2.1.1	Bioinformatics . . . . .	41
2.1.2	Structure From Sequence . . . . .	46
2.2	Protein Folding . . . . .	46
2.2.1	Folding Views . . . . .	46
2.2.2	Folding Challenges . . . . .	48
2.2.3	Folding Simulations . . . . .	49
2.2.4	Chaperones . . . . .	50
2.2.5	Unstructured Proteins . . . . .	53
2.3	Protein Misfolding . . . . .	53
2.3.1	Prions . . . . .	53
2.3.2	Infectious Proteins? . . . . .	54
2.3.3	Hypotheses . . . . .	54
2.3.4	Other Misfolding Processes . . . . .	56
2.3.5	Function From Structure . . . . .	56
2.4	Practical Applications . . . . .	57
2.4.1	Drug Design . . . . .	58
2.4.2	AIDS Drugs . . . . .	60
2.4.3	Other Drugs and Future Prospects . . . . .	65
2.4.4	Gene Therapy – Better Genes . . . . .	67
2.4.5	Designed Compounds and Foods . . . . .	70
2.4.6	Nutrigenomics . . . . .	73
2.4.7	Designer Materials . . . . .	74
2.4.8	Cosmeceuticals . . . . .	74
<b>3</b>	<b>Protein Structure Introduction</b>	<b>77</b>
3.1	Machinery of Life . . . . .	77
3.1.1	From Tissues to Hormones . . . . .	77
3.1.2	Size and Function Variability . . . . .	78
3.1.3	Chapter Overview . . . . .	79
3.2	Amino Acid Building Blocks . . . . .	82

3.2.1	Basic C <sup>α</sup> Unit . . . . .	82
3.2.2	Essential and Nonessential Amino Acids . . . . .	83
3.2.3	Linking Amino Acids . . . . .	85
3.2.4	The Amino Acid Repertoire . . . . .	88
3.3	Sequence Variations in Proteins . . . . .	90
3.3.1	Globular Proteins . . . . .	90
3.3.2	Membrane and Fibrous Proteins . . . . .	91
3.3.3	Emerging Patterns from Genome Databases . . . . .	92
3.3.4	Sequence Similarity . . . . .	93
3.4	Protein Conformation Framework . . . . .	98
3.4.1	The Flexible $\phi$ and $\psi$ and Rigid $\omega$ Dihedral Angles . . . . .	98
3.4.2	Rotameric Structures . . . . .	100
3.4.3	Ramachandran Plots . . . . .	101
3.4.4	Conformational Hierarchy . . . . .	104
<b>4</b>	<b>Protein Structure Hierarchy</b> . . . . .	<b>107</b>
4.1	Structure Hierarchy . . . . .	108
4.2	Helices . . . . .	108
4.2.1	Classic $\alpha$ -Helix . . . . .	108
4.2.2	$3_{10}$ and $\pi$ Helices . . . . .	109
4.2.3	Left-Handed $\alpha$ -Helix . . . . .	112
4.2.4	Collagen Helix . . . . .	112
4.3	$\beta$ -Sheets: A Common Secondary Structural Element . . . . .	112
4.4	Turns and Loops . . . . .	112
4.5	Supersecondary and Tertiary Structure . . . . .	115
4.5.1	Complex 3D Networks . . . . .	115
4.5.2	Classes in Protein Architecture . . . . .	115
4.5.3	Classes are Further Divided into Folds . . . . .	116
4.6	$\alpha$ -Class Folds . . . . .	116
4.6.1	Bundles . . . . .	116
4.6.2	Folded Leafs . . . . .	117
4.6.3	Hairpin Arrays . . . . .	117
4.7	$\beta$ -Class Folds . . . . .	118
4.7.1	Anti-Parallel $\beta$ Domains . . . . .	118
4.7.2	Parallel and Antiparallel Combinations . . . . .	119
4.8	$\alpha/\beta$ and $\alpha+\beta$ -Class Folds . . . . .	119
4.8.1	$\alpha/\beta$ Barrels . . . . .	120
4.8.2	Open Twisted $\alpha/\beta$ Folds . . . . .	120
4.8.3	Leucine-Rich $\alpha/\beta$ Folds . . . . .	120
4.8.4	$\alpha+\beta$ Folds . . . . .	120
4.8.5	Other Folds . . . . .	120
4.9	Number of Folds . . . . .	121
4.9.1	Finite Number? . . . . .	121
4.10	Quaternary Structure . . . . .	121
4.10.1	Viruses . . . . .	123

4.10.2	From Ribosomes to Dynamic Networks . . . . .	126
4.11	Structure Classification . . . . .	128
<b>5</b>	<b>Nucleic Acids Structure Minitutorial</b>	<b>131</b>
5.1	DNA, Life's Blueprint . . . . .	132
5.1.1	The Kindled Field of Molecular Biology . . . . .	132
5.1.2	DNA Processes . . . . .	135
5.1.3	Challenges in Nucleic Acid Structure . . . . .	135
5.1.4	Chapter Overview . . . . .	136
5.2	Basic Building Blocks . . . . .	137
5.2.1	Nitrogenous Bases . . . . .	137
5.2.2	Hydrogen Bonds . . . . .	139
5.2.3	Nucleotides . . . . .	141
5.2.4	Polynucleotides . . . . .	141
5.2.5	Stabilizing Polynucleotide Interactions . . . . .	141
5.2.6	Chain Notation . . . . .	143
5.2.7	Atomic Labeling . . . . .	143
5.2.8	Torsion Angle Labeling . . . . .	144
5.3	Conformational Flexibility . . . . .	145
5.3.1	The Furanose Ring . . . . .	145
5.3.2	Backbone Torsional Flexibility . . . . .	148
5.3.3	The Glycosyl Rotation . . . . .	150
5.3.4	Sugar/Glycosyl Combinations . . . . .	151
5.3.5	Basic Helical Descriptors . . . . .	153
5.3.6	Base-Pair Parameters . . . . .	154
5.4	Canonical DNA Forms . . . . .	158
5.4.1	B-DNA . . . . .	160
5.4.2	A-DNA . . . . .	160
5.4.3	Z-DNA . . . . .	163
5.4.4	Comparative Features . . . . .	164
<b>6</b>	<b>Topics in Nucleic Acids Structure: DNA Interactions and Folding</b>	<b>167</b>
6.1	Introduction . . . . .	168
6.2	DNA Sequence Effects . . . . .	169
6.2.1	Local Deformations . . . . .	169
6.2.2	Orientation Preferences in Dinucleotide Steps . . . . .	170
6.2.3	Orientation Preferences in Dinucleotide Steps With Flanking Sequence Context: Tetranucleotide Studies . . . . .	173
6.2.4	Intrinsic DNA Bending in A-Tracts . . . . .	174
6.2.5	Sequence Deformability Analysis Continues . . . . .	178
6.3	DNA Hydration and Ion Interactions . . . . .	178
6.3.1	Resolution Difficulties . . . . .	180
6.3.2	Basic Patterns . . . . .	180
6.4	DNA/Protein Interactions . . . . .	184
6.5	Cellular Organization of DNA . . . . .	186

6.5.1	Compaction of Genomic DNA . . . . .	186
6.5.2	Coiling of the DNA Helix Itself . . . . .	189
6.5.3	Chromosomal Packaging of Coiled DNA . . . . .	189
6.6	Mathematical Characterization of DNA Supercoiling . . . . .	198
6.6.1	DNA Topology and Geometry . . . . .	198
6.7	Computational Treatments of DNA Supercoiling . . . . .	200
6.7.1	DNA as a Flexible Polymer . . . . .	200
6.7.2	Elasticity Theory Framework . . . . .	201
6.7.3	Simulations of DNA Supercoiling . . . . .	203
<b>7</b>	<b>Topics in Nucleic Acids Structure: Noncanonical Helices and RNA Structure</b>	<b>211</b>
7.1	Introduction . . . . .	211
7.2	Variations on a Theme . . . . .	212
7.2.1	Hydrogen Bonding Patterns in Polynucleotides . . . . .	212
7.2.2	Hybrid Helical/Nonhelical Forms . . . . .	216
7.2.3	Overstretched and Understretched DNA . . . . .	220
7.3	RNA Structure and Function . . . . .	222
7.3.1	DNA's Cousin Shines . . . . .	222
7.3.2	RNA Chains Fold Upon Themselves . . . . .	222
7.3.3	RNA's Diversity . . . . .	225
7.3.4	Non-Coding and Micro-RNAs . . . . .	227
7.3.5	RNA at Atomic Resolution . . . . .	229
7.4	Current Challenges in RNA Modeling . . . . .	231
7.4.1	RNA Folding . . . . .	231
7.4.2	RNA Motifs . . . . .	232
7.4.3	RNA Structure Prediction . . . . .	232
7.5	Application of Graph Theory to Studies of RNA Structure and Function . . . . .	236
7.5.1	Graph Theory . . . . .	236
7.5.2	RNA-As-Graphs (RAG) Resource . . . . .	237
<b>8</b>	<b>Theoretical and Computational Approaches to Biomolecular Structure</b>	<b>243</b>
8.1	Merging of Theory and Experiment . . . . .	244
8.1.1	Exciting Times for Computationalists! . . . . .	244
8.1.2	The Future of Biocomputations . . . . .	246
8.1.3	Chapter Overview . . . . .	246
8.2	QM Foundations . . . . .	247
8.2.1	The Schrödinger Wave Equation . . . . .	247
8.2.2	The Born-Oppenheimer Approximation . . . . .	248
8.2.3	Ab Initio . . . . .	248
8.2.4	Semi-Empirical QM . . . . .	250
8.2.5	Recent Advances in Quantum Mechanics . . . . .	250
8.2.6	From Quantum to Molecular Mechanics . . . . .	253

8.3	Molecular Mechanics Principles . . . . .	257
8.3.1	The Thermodynamic Hypothesis . . . . .	258
8.3.2	Additivity . . . . .	259
8.3.3	Transferability . . . . .	261
8.4	Molecular Mechanics Formulation . . . . .	264
8.4.1	Configuration Space . . . . .	265
8.4.2	Functional Form . . . . .	266
8.4.3	Some Current Limitations . . . . .	268
<b>9</b>	<b>Force Fields</b>	<b>271</b>
9.1	Formulation of the Model and Energy . . . . .	273
9.2	Normal Modes . . . . .	273
9.2.1	Characteristic Motions . . . . .	273
9.2.2	Spectra of Biomolecules . . . . .	275
9.2.3	Spectra As Force Constant Sources . . . . .	276
9.2.4	In-Plane and Out-of-Plane Bending . . . . .	277
9.3	Bond Length Potentials . . . . .	278
9.3.1	Harmonic Term . . . . .	279
9.3.2	Morse Term . . . . .	280
9.3.3	Cubic and Quartic Terms . . . . .	282
9.4	Bond Angle Potentials . . . . .	283
9.4.1	Harmonic and Trigonometric Terms . . . . .	283
9.4.2	Cross Bond Stretch / Angle Bend Terms . . . . .	285
9.5	Torsional Potentials . . . . .	287
9.5.1	Origin of Rotational Barriers . . . . .	287
9.5.2	Fourier Terms . . . . .	288
9.5.3	Torsional Parameter Assignment . . . . .	289
9.5.4	Improper Torsion . . . . .	293
9.5.5	Cross Dihedral/Bond Angle and Improper/Improper Dihedral Terms . . . . .	294
9.6	van der Waals Potential . . . . .	295
9.6.1	Rapidly Decaying Potential . . . . .	295
9.6.2	Parameter Fitting From Experiment . . . . .	295
9.6.3	Two Parameter Calculation Protocols . . . . .	296
9.7	Coulomb Potential . . . . .	297
9.7.1	Coulomb's Law: Slowly Decaying Potential . . . . .	297
9.7.2	Dielectric Function . . . . .	299
9.7.3	Partial Charges . . . . .	301
9.8	Parameterization . . . . .	302
9.8.1	A Package Deal . . . . .	302
9.8.2	Force Field Comparisons . . . . .	302
9.8.3	Force Field Performance . . . . .	303
<b>10</b>	<b>Nonbonded Computations</b>	<b>305</b>
10.1	Computational Bottleneck . . . . .	307

10.2	Reducing Computational Cost . . . . .	308
10.2.1	Simple Cutoff Schemes . . . . .	308
10.2.2	Ewald and Multipole Schemes . . . . .	309
10.3	Spherical Cutoff Techniques . . . . .	310
10.3.1	Technique Categories . . . . .	310
10.3.2	Guidelines for Cutoff Functions . . . . .	311
10.3.3	General Cutoff Formulations . . . . .	313
10.3.4	Potential Switch . . . . .	314
10.3.5	Force Switch . . . . .	314
10.3.6	Shift Functions . . . . .	316
10.4	Ewald Method . . . . .	317
10.4.1	Periodic Boundary Conditions . . . . .	317
10.4.2	Ewald Sum and Crystallography . . . . .	320
10.4.3	Morphing A Conditionally Convergent Sum . . . . .	322
10.4.4	Finite-Dielectric Correction . . . . .	326
10.4.5	Ewald Sum Complexity . . . . .	326
10.4.6	Resulting Ewald Summation . . . . .	327
10.4.7	Practical Implementation . . . . .	329
10.5	Multipole Method . . . . .	331
10.5.1	Basic Hierarchical Strategy . . . . .	331
10.5.2	Historical Perspective . . . . .	335
10.5.3	Expansion in Spherical Coordinates . . . . .	337
10.5.4	Biomolecular Implementations . . . . .	338
10.5.5	Other Variants . . . . .	339
10.6	Continuum Solvation . . . . .	339
10.6.1	Need for Simplification! . . . . .	339
10.6.2	Potential of Mean Force . . . . .	340
10.6.3	Stochastic Dynamics . . . . .	341
10.6.4	Continuum Electrostatics . . . . .	344
<b>11</b>	<b>Multivariate Minimization in Computational Chemistry</b>	<b>351</b>
11.1	Optimization Applications . . . . .	353
11.1.1	Algorithmic Understanding Needed . . . . .	353
11.1.2	Chapter Overview . . . . .	353
11.2	Fundamentals . . . . .	354
11.2.1	Problem Formulation . . . . .	354
11.2.2	Independent Variables . . . . .	355
11.2.3	Function Characteristics . . . . .	355
11.2.4	Local and Global Minima . . . . .	357
11.2.5	Derivatives . . . . .	359
11.2.6	Hessian Matrix . . . . .	361
11.3	Basic Algorithms . . . . .	362
11.3.1	Greedy Descent . . . . .	362
11.3.2	Line Searches . . . . .	365
11.3.3	Trust Region Methods . . . . .	367

11.3.4	Convergence Criteria . . . . .	368
11.4	Newton's Method . . . . .	370
11.4.1	Newton in One Dimension . . . . .	371
11.4.2	Newton's Method for Minimization . . . . .	374
11.4.3	Multivariate Newton . . . . .	376
11.5	Large-Scale methods . . . . .	376
11.5.1	Quasi-Newton (QN) . . . . .	377
11.5.2	Conjugate Gradient (CG) . . . . .	379
11.5.3	Truncated-Newton (TN) . . . . .	381
11.5.4	Simple Example . . . . .	383
11.6	Software . . . . .	384
11.6.1	Popular Newton and CG . . . . .	384
11.6.2	CHARMM's ABNR . . . . .	384
11.6.3	CHARMM's TN . . . . .	386
11.6.4	Comparative Performance on Molecular Systems . . . . .	387
11.7	Recommendations . . . . .	387
11.8	Future Outlook . . . . .	391
<b>12</b>	<b>Monte Carlo Techniques</b>	<b>393</b>
12.1	Monte Carlo Popularity . . . . .	394
12.1.1	A Winning Combination . . . . .	394
12.1.2	From Needles to Bombs . . . . .	395
12.1.3	Chapter Overview . . . . .	395
12.1.4	Importance of Error Bars . . . . .	396
12.2	Random Number Generators . . . . .	396
12.2.1	What is <i>Random</i> ? . . . . .	396
12.2.2	Properties of Generators? . . . . .	397
12.2.3	Linear Congruential Generators . . . . .	400
12.2.4	Other Generators . . . . .	404
12.2.5	Artifacts . . . . .	408
12.2.6	Recommendations . . . . .	410
12.3	Gaussian Random Variates . . . . .	411
12.3.1	Manipulation of Uniform Random Variables . . . . .	411
12.3.2	Normal Variates in Molecular Simulations . . . . .	411
12.3.3	Odeh/Evans . . . . .	412
12.3.4	Box/Muller/Marsaglia . . . . .	414
12.4	Monte Carlo Means . . . . .	414
12.4.1	Expected Values . . . . .	414
12.4.2	Error Bars . . . . .	416
12.4.3	Batch Means . . . . .	419
12.5	Monte Carlo Sampling . . . . .	420
12.5.1	Probability Density Function . . . . .	420
12.5.2	Equilibria or Dynamics . . . . .	420
12.5.3	Ensembles . . . . .	421
12.5.4	Importance Sampling . . . . .	422



12.6	MC Applications . . . . .	427
12.6.1	General attractiveness . . . . .	427
12.6.2	Biased MC . . . . .	428
12.6.3	MC and MD . . . . .	429
12.6.4	Parallel Tempering and Other MC Variants . . . . .	430
<b>13</b>	<b>Molecular Dynamics: Basics</b>	<b>433</b>
13.1	Introduction . . . . .	434
13.1.1	Why Molecular Dynamics? . . . . .	434
13.1.2	Background . . . . .	435
13.1.3	Outline of MD Chapters . . . . .	438
13.2	Laplace's Vision . . . . .	439
13.2.1	The Dream Becomes Reality . . . . .	439
13.2.2	Deterministic Mechanics . . . . .	440
13.2.3	Neglect of Electronic Motion . . . . .	440
13.2.4	Critical Frequencies . . . . .	441
13.2.5	Electron/Nuclear Treatment . . . . .	442
13.3	Basics . . . . .	444
13.3.1	Following Motion . . . . .	444
13.3.2	Trajectory Quality . . . . .	444
13.3.3	Initial System Settings . . . . .	446
13.3.4	Trajectory Sensitivity . . . . .	448
13.3.5	Simulation Protocol . . . . .	450
13.3.6	High-Speed Implementations . . . . .	451
13.3.7	Analysis and Visualization . . . . .	453
13.3.8	Reliable Numerical Integration . . . . .	454
13.3.9	Computational Complexity . . . . .	454
13.4	Verlet Algorithm . . . . .	457
13.4.1	Position and Velocity Propagation . . . . .	458
13.4.2	Leapfrog, Velocity Verlet, and Position Verlet . . . . .	459
13.5	Constrained Dynamics . . . . .	462
13.6	Various MD Ensembles . . . . .	464
13.6.1	Ensemble Types . . . . .	464
13.6.2	Simple Algorithms . . . . .	464
13.6.3	Extended System Methods . . . . .	467
<b>14</b>	<b>Molecular Dynamics: Further Topics</b>	<b>471</b>
14.1	Introduction . . . . .	472
14.2	Symplectic Integrators . . . . .	473
14.2.1	Symplectic Transformation . . . . .	474
14.2.2	Harmonic Oscillator Example . . . . .	475
14.2.3	Linear Stability . . . . .	475
14.2.4	Timestep-Dependent Rotation in Phase Space . . . . .	476
14.2.5	Resonance Condition for Periodic Motion . . . . .	477
14.2.6	Resonance Artifacts . . . . .	479

14.3	Multiple-Timestep (MTS) Methods . . . . .	480
14.3.1	Basic Idea . . . . .	480
14.3.2	Extrapolation . . . . .	481
14.3.3	Impulses . . . . .	482
14.3.4	Resonances in Impulse Splitting . . . . .	483
14.3.5	Resonance Artifacts in MTS . . . . .	483
14.3.6	Resonance Consequences . . . . .	486
14.4	Langevin Dynamics . . . . .	487
14.4.1	Uses . . . . .	487
14.4.2	Heat Bath . . . . .	487
14.4.3	Effect of $\gamma$ . . . . .	488
14.4.4	Generalized Verlet for Langevin Dynamics . . . . .	490
14.4.5	LN Method . . . . .	490
14.5	Brownian Dynamics (BD) . . . . .	495
14.5.1	Brownian Motion . . . . .	495
14.5.2	Brownian Framework . . . . .	497
14.5.3	General Propagation Framework . . . . .	498
14.5.4	Hydrodynamics . . . . .	499
14.5.5	BD Propagation . . . . .	502
14.6	Implicit Integration . . . . .	505
14.6.1	Implicit vs. Explicit Euler . . . . .	505
14.6.2	Intrinsic Damping . . . . .	507
14.6.3	Computational Time . . . . .	507
14.6.4	Resonance Artifacts . . . . .	507
14.7	Enhanced Sampling Methods . . . . .	512
14.7.1	Overview . . . . .	512
14.7.2	Harmonic-Analysis Based Techniques . . . . .	512
14.7.3	Other Coordinate Transformations . . . . .	515
14.7.4	Coarse Graining Models . . . . .	515
14.7.5	Biasing Approaches . . . . .	517
14.7.6	Variations in MD Algorithm and Protocol . . . . .	518
14.7.7	Other Rigorous Approaches for Deducing Mechanisms, Free Energies, and Reaction Rates . . . . .	519
14.8	Future Outlook . . . . .	522
14.8.1	Integration Ingenuity . . . . .	522
14.8.2	Current Challenges . . . . .	523
<b>15</b>	<b>Similarity and Diversity in Chemical Design</b> . . . . .	<b>527</b>
15.1	Introduction to Drug Design . . . . .	528
15.1.1	Chemical Libraries . . . . .	528
15.1.2	Early Days . . . . .	529
15.1.3	Rational Drug Design . . . . .	531
15.1.4	Automated Technology . . . . .	532
15.1.5	Chapter Overview . . . . .	534
15.2	Database Problems . . . . .	534

15.2.1	Database Analysis . . . . .	535
15.2.2	Similarity and Diversity Sampling . . . . .	536
15.2.3	Bioactivity . . . . .	537
15.3	General Problem Definitions . . . . .	541
15.3.1	The Dataset . . . . .	541
15.3.2	The Compound Descriptors . . . . .	541
15.3.3	Biological Activity . . . . .	543
15.3.4	The Target Function . . . . .	544
15.3.5	Scaling Descriptors . . . . .	545
15.3.6	The Similarity and Diversity Problems . . . . .	546
15.4	Data Compression and Cluster Analysis . . . . .	548
15.4.1	PCA compression . . . . .	548
15.4.2	SVD compression . . . . .	550
15.4.3	PCA and SVD . . . . .	553
15.4.4	Projection Application . . . . .	553
15.4.5	Example . . . . .	555
15.5	Future Perspectives . . . . .	558
	<b>Epilogue</b>	<b>563</b>
	<b>Appendices</b>	<b>563</b>
	<b>A Molecular Modeling Sample Syllabus</b>	<b>565</b>
	<b>B Article Reading List</b>	<b>567</b>
	<b>C Supplementary Course Texts</b>	<b>573</b>
	<b>D Homework Assignments</b>	<b>581</b>
	<b>Bibliography</b>	<b>633</b>
	<b>References</b>	<b>633</b>
	<b>Index</b>	<b>719</b>