

Contents

Preface *XVII*

Part One Introduction to Systems Biology *1*

1	Introduction	3
1.1	Biology in Time and Space	3
1.2	Models and Modeling	4
1.2.1	What is a Model?	5
1.2.2	Purpose and Adequateness of Models	5
1.2.3	Advantages of Computational Modeling	6
1.3	Basic Notions for Computational Models	7
1.3.1	Model Scope	7
1.3.2	Model Statements	8
1.3.3	System State	8
1.3.4	Variables, Parameters, and Constants	8
1.3.5	Model Behavior	9
1.3.6	Model Classification	9
1.3.7	Steady States	9
1.3.8	Model Assignment is not Unique	10
1.4	Data Integration	11
1.5	Standards	12
	References	12
2	Modeling of Biochemical Systems	13
2.1	Kinetic Modeling of Enzymatic Reactions	13
2.1.1	The Law of Mass Action	14
2.1.2	Reaction Kinetics and Thermodynamics	15
2.1.3	Michaelis–Menten Kinetics	18
2.1.3.1	How to Derive a Rate Equation	19
2.1.3.2	Parameter Estimation and Linearization of the Michaelis–Menten Equation	20
2.1.3.3	The Michaelis–Menten Equation for Reversible Reactions	22

2.1.4	Regulation of Enzyme Activity by Effectors	22
2.1.4.1	Substrate Inhibition	25
2.1.4.2	Binding of Ligands to Proteins	26
2.1.4.3	Positive Homotropic Cooperativity and the Hill Equation	27
2.1.4.4	The Monod–Wyman–Changeux Model for Sigmoid Kinetics	28
2.1.5	Generalized Mass Action Kinetics	29
2.1.6	Approximate Kinetic Formats	30
2.1.7	Convenience Kinetics	30
2.2	Structural Analysis of Biochemical Systems	31
2.2.1	Systems Equations	31
2.2.2	Information Encoded in the Stoichiometric Matrix N	34
2.2.3	Elementary Flux Modes and Extreme Pathways	36
2.2.3.1	Flux Cone	37
2.2.4	Conservation Relations: Null Space of N^T	39
2.3	Kinetic Models of Biochemical Systems	42
2.3.1	Describing Dynamics with ODEs	42
2.3.1.1	Notations	43
2.3.1.2	Linearization of Autonomous Systems	44
2.3.1.3	Solution of Linear ODE Systems	45
2.3.1.4	Stability of Steady States	46
2.3.1.5	Global Stability of Steady States	49
2.3.1.6	Limit Cycles	49
2.3.2	Metabolic Control Analysis	51
2.3.2.1	The Coefficients of Control Analysis	52
2.3.2.2	The Elasticity Coefficients	52
2.3.2.3	Control Coefficients	55
2.3.2.4	Response Coefficients	55
2.3.2.5	Matrix Representation of the Coefficients	55
2.3.2.6	The Theorems of Metabolic Control Theory	56
2.3.2.7	The Summation Theorems	56
2.3.2.8	The Connectivity Theorems	58
2.3.2.9	Derivation of Matrix Expressions for Control Coefficients	59
2.4	Tools and Data Formats for Modeling	63
2.4.1	Simulation Techniques	64
2.4.1.1	Petri Nets	64
2.4.1.2	Cellular Automata	65
2.4.2	Simulation Tools	65
2.4.2.1	CellDesigner	66
2.4.2.2	COPASI	67
2.4.2.3	PyBioS	68
2.4.3	Data Formats	70
2.4.3.1	Systems Biology Markup Language	70
2.4.3.2	BioPAX	73
2.4.3.3	Systems Biology Graphical Notation	73

2.4.3.4	Standards for Systems Biology	74
2.4.4	Data Resources	75
2.4.4.1	Pathway Databases	76
2.4.4.2	Databases of Kinetic Data	77
2.4.4.3	Model Databases	77
	References	79
3	Specific Biochemical Systems	83
3.1	Metabolic Systems	83
3.1.1	Basic Elements of Metabolic Modeling	84
3.1.2	Toy Model of Upper Glycolysis	85
3.1.3	Threonine Synthesis Pathway Model	88
3.2	Signaling Pathways	91
3.2.1	Introduction	92
3.2.2	Function and Structure of Intra- and Intercellular Communication	92
3.2.3	Receptor–Ligand Interactions	93
3.2.4	Structural Components of Signaling Pathways	96
3.2.4.1	G proteins	96
3.2.4.2	Small G proteins	99
3.2.4.3	Phosphorelay Systems	100
3.2.4.4	MAP Kinase Cascades	102
3.2.4.5	Jak/Stat Pathways	106
3.2.5	Signaling – Dynamic and Regulatory Features	106
3.2.5.1	Quantitative Measures for Properties of Signaling Pathways	107
3.2.5.2	Crosstalk in Signaling Pathways	109
3.3	The Cell Cycle	111
3.3.1	Steps in the Cycle	114
3.3.2	Minimal Cascade Model of a Mitotic Oscillator	115
3.3.3	Models of Budding Yeast Cell Cycle	117
3.3.4	Modeling Nucleo/Cytoplasmatic Compartmentalization	119
3.4	Spatial Models	121
3.4.1	Types of Spatial Models	122
3.4.1.1	Compartment Models and Partial Differential Equations	122
3.4.1.2	Stochastic Models	123
3.4.1.3	Cellular Automata	123
3.4.2	Compartment Models	123
3.4.3	Reaction–Diffusion Systems	125
3.4.3.1	The Diffusion Equation	125
3.4.3.2	Solutions of the Diffusion Equation	126
3.4.3.3	Reaction–Diffusion Equation	127
3.4.4	Pattern Formation in Tissue Development	128
3.4.5	Spontaneous Pattern Formation	130
3.5	Apoptosis	132
3.5.1	Molecular Biology of Apoptosis	132

3.5.2	Modeling of Apoptosis	135
	References	142
4	Model Fitting	147
4.1	Data for Small Metabolic and Signaling Systems	147
4.1.1	Databases for Kinetic Modeling	148
4.1.2	Measuring Promoter Activities Using GFP Reporter Genes	150
4.2	Parameter Estimation	152
4.2.1	Regression	153
4.2.2	Estimators	153
4.2.2.1	Method of Least Squares and Maximum-Likelihood Estimation	155
4.2.3	Identifiability	155
4.2.4	Bootstrapping	157
4.2.5	Crossvalidation	158
4.2.6	Bayesian Parameter Estimation	159
4.2.7	Local and Global Optimization	160
4.2.7.1	Local Optimization	161
4.2.7.2	Global Optimization	161
4.2.7.3	Sampling Methods	162
4.2.7.4	Genetic Algorithms	163
4.3	Reduction and Coupling of Models	164
4.3.1	Model Simplification	164
4.3.2	Tacit Model Assumptions	166
4.3.3	Reduction of Fast Processes	167
4.3.3.1	Response Time	167
4.3.3.2	Time-Scale Separation	167
4.3.4	Global Model Reduction	170
4.3.4.1	Linearized Biochemical Models	171
4.3.4.2	Linear Relaxation Modes	171
4.3.5	Coupled Systems and Emergent Behavior	172
4.3.6	Modeling of Coupled Systems	174
4.3.6.1	Bottom-Up and Top-Down Modeling	174
4.3.6.2	Modeling the System Boundary	175
4.3.6.3	Coupling of Submodels	175
4.3.6.4	Model Merging	175
4.4	Model Selection	176
4.4.1	What is a Good Model?	177
4.4.2	Statistical Tests and Model Selection	178
4.4.3	Maximum-Likelihood Estimation and χ^2 -Test	180
4.4.4	Overfitting	181
4.4.5	Likelihood Ratio Test	182
4.4.6	Selection Criteria	183
4.4.7	Bayesian Model Selection	184
4.4.8	Cycle of Experiments and Modeling	186

4.4.9	Models are Growing in Complexity	186
	References	189
5	Analysis of High-Throughput Data	193
5.1	High-Throughput Experiments	193
5.1.1	DNA Array Platforms	193
5.1.2	Platform Comparison	196
5.1.3	Next Generation Sequencing	196
5.1.4	Image Analysis and Data Quality Control	198
5.1.4.1	Grid Finding	198
5.1.4.2	Spot Quantification	200
5.1.4.3	Signal Validity	200
5.1.5	Preprocessing	202
5.1.5.1	Global Measures	203
5.1.5.2	Linear Models	203
5.1.5.3	Nonlinear and Spatial Effects	204
5.1.5.4	Other Approaches	204
5.2	Analysis of Gene Expression Data	205
5.2.1	Planning and Designing Experiments for Case-Control Studies	205
5.2.2	Tests for Differential Expression	206
5.2.2.1	DNA Arrays	206
5.2.2.2	Next Generation Sequencing	209
5.2.3	Multiple Testing	209
5.2.4	ROC Curve Analysis	211
5.2.5	Clustering Algorithms	213
5.2.5.1	Hierarchical Clustering	215
5.2.5.2	Self-Organizing Maps (SOMs)	218
5.2.5.3	K-Means	218
5.2.6	Cluster Validation	220
5.2.7	Overrepresentation and Enrichment Analyses	223
5.2.8	Classification Methods	226
5.2.8.1	Support Vector Machines	227
5.2.8.2	Other Approaches	229
	References	232
6	Gene Expression Models	235
6.1	Mechanisms of Gene Expression Regulation	235
6.1.1	Transcription-Factor Initiated Gene Regulation	235
6.1.2	General Promoter Structure	237
6.1.3	Prediction and Analysis of Promoter Elements	239
6.1.3.1	Sequence-Based Analysis	239
6.1.3.2	Approaches that Incorporate Additional Information	241
6.1.4	Posttranscriptional Regulation Through microRNAs	243
6.1.4.1	Identification of microRNAs in the Genome Sequence	245
6.1.4.2	MicroRNA Target Prediction	246

6.1.4.3	Experimental Implications – RNA Interference	246
6.2	Gene Regulation Functions	248
6.2.1	The Lac Operon in <i>Escherichia coli</i>	249
6.2.2	Gene Regulation Functions Derived from Equilibrium Binding	250
6.2.3	Occupation Probability Derived from Statistical Thermodynamics	251
6.2.4	Gene Regulation Function of the Lac Operon	253
6.2.5	Transcriptional Regulation in Larger Networks	254
6.2.6	Network Component Analysis	254
6.3	Dynamic Models of Gene Regulation	256
6.3.1	One Gene Regulatory Network: Different Approaches	256
6.3.2	Representation of a Gene Regulatory Network as Graph	256
6.3.3	Bayesian Networks	258
6.3.4	Boolean Networks	259
6.3.5	Description with Ordinary Differential Equations	262
6.3.6	Gene Expression Modeling with Stochastic Processes	264
	References	267
7	Stochastic Systems and Variability	271
7.1	Stochastic Modeling of Biochemical Reactions	271
7.1.1	Chemical Random Process for Molecule Numbers	272
7.1.2	The Chemical Master Equation	273
7.1.3	Stochastic Simulation	275
7.1.3.1	Direct Method	275
7.1.3.2	Explicit τ -Leaping Method	276
7.1.3.3	Stochastic Simulation and Spatial Models	276
7.1.4	The Chemical Langevin Equation	276
7.1.5	Deterministic and Stochastic Modeling Frameworks	278
7.1.6	Temporal Fluctuations	279
7.2	Fluctuations in Gene Expression	281
7.2.1	Stochastic Model of Transcription and Translation	283
7.2.1.1	Macroscopic Kinetic Model	283
7.2.1.2	Microscopic Stochastic Model	284
7.2.1.3	Fluctuations and Protein Bursts	285
7.2.2	Measuring the Intrinsic and Extrinsic Variability	286
7.2.3	Temporal Fluctuations in a Gene Cascade	288
7.2.3.1	Linear Model of Two Genes	288
7.2.3.2	Measuring the Time Correlations in Protein Levels	290
7.2.4	Biological Functions of Noise	291
7.2.4.1	Random Switching	291
7.2.4.2	Exploration Strategies	291
7.3	Variability and Uncertainty	292
7.3.1	Models with Uncertain Constant Parameters	292
7.3.2	Computing the Distribution of Output Variables	293
7.3.2.1	Monte Carlo Simulation	293

7.3.2.2	Approximation for Narrow Parameter Distributions	294
7.3.2.3	Temporal Parameter Fluctuations	295
7.3.3	Uncertainty Analysis of Biochemical Models	295
7.3.3.1	Sampling of Reaction Elasticities	297
7.3.4	Distributions for Kinetic Parameters	298
7.3.4.1	Principle of Minimal Information	298
7.3.4.2	Thermodynamic Constraints on Parameters	299
7.3.4.3	Obtaining Parameter Distributions from Experimental Data	299
7.4	Robustness	300
7.4.1	Robustness Properties in Biochemical Systems	301
7.4.1.1	Biological Robustness Properties	301
7.4.1.2	Mathematical Robustness Criteria	301
7.4.1.3	Precise Robustness in a Bacterial Two-Component System	301
7.4.2	Structural Robustness in Large Networks	303
7.4.2.1	Backup Genes	303
7.4.2.2	Backup Pathways	304
7.4.3	Quantitative Robustness by Feedback	304
7.4.3.1	Negative Feedback	304
7.4.3.2	Integral Feedback	306
7.4.4	Scaling Laws, Invariance, and Dimensional Analysis	306
7.4.5	Summation Laws and Homogeneous Functions	308
7.4.5.1	Summation Theorems	308
7.4.5.2	Conservation Laws for Sensitivity	308
7.4.5.3	Compensation of Correlated Fluctuations	309
7.4.6	Robustness and Evolvability	309
7.4.7	Robustness and Modeling	310
	References	312
8	Network Structures, Dynamics, and Function	315
8.1	Structure of Biochemical Networks	315
8.1.1	Mathematical Graphs	317
8.1.2	Random Graphs	318
8.1.2.1	Erdős–Rényi Random Graphs	318
8.1.2.2	Geometric Random Graphs	319
8.1.2.3	Random Graphs with Predefined Degree Sequence	319
8.1.3	Scale-Free Networks	319
8.1.4	Clustering and Local Structure	321
8.1.4.1	Clustering Coefficient	321
8.1.4.2	Small-World Networks	321
8.1.5	Network Motifs	322
8.1.6	Structure of Metabolic Networks	323
8.1.7	The Network Picture	324
8.2	Network Motifs	325
8.2.1	Transcription Networks and Network Motifs	326
8.2.2	Single Regulation Arrows and Their Steady-State Response	328

8.2.3	Adaptation Motif	329
8.2.4	Negative Feedback	330
8.2.5	Feed-Forward Loops	331
8.2.6	Dynamic Model of the Feed-Forward Loop	332
8.2.7	Dynamics and Function of Network Motifs	333
8.3	Modularity	335
8.3.1	Modularity as a Fact or as an Assumption	336
8.3.2	Aspects of Modularity: Structure, Function, Dynamics, Regulation, and Genetics	337
8.3.3	Structural Modules in Cellular Networks	337
8.3.4	Modular Response Analysis	338
8.3.5	Functional Modules Detected by Epistasis	339
8.3.6	Evolution of Modularity and Complexity	341
8.3.6.1	Tinkering and Engineering	341
8.3.6.2	Analogy in Evolution	342
8.3.6.3	Modularity, Robustness, and Evolvability	342
	References	343
9	Optimality and Evolution	349
9.1	Optimality and Constraint-Based Models	349
9.1.1	Optimization by Evolution	350
9.1.2	Optimality Studies in Systems Biology	350
9.1.2.1	The Fitness Function	351
9.1.2.2	Optimality and Compromise	351
9.1.2.3	Cost-Benefit Calculations	351
9.1.2.4	Inequality Constraints	352
9.1.2.5	Local Optima	353
9.1.3	Constraint-Based Flux Optimization	353
9.1.3.1	Flux-Balance Analysis	353
9.1.3.2	Geometric Interpretation of Flux-Balance Analysis	354
9.1.4	Thermodynamic Constraints	355
9.1.5	Applications and Tests of Flux-Optimization Paradigm	356
9.2	Optimal Enzyme Concentrations	357
9.2.1	Optimization of Catalytic Properties of Single Enzymes	358
9.2.2	Optimal Distribution of Enzyme Concentrations in a Metabolic Pathway	360
9.2.3	Temporal Transcription Programs	363
9.3	Evolutionary Game Theory	367
9.3.1	Game Theory	369
9.3.1.1	Hawk–Dove Game and Prisoner’s Dilemma	369
9.3.1.2	Best Choices and Nash Equilibrium	370
9.3.2	Evolutionary Game Theory	371
9.3.3	Replicator Equation for Population Dynamics	371
9.3.3.1	The Replicator Equation	372
9.3.3.2	Outcomes of Frequency-Dependent Selection	372

9.3.4	Evolutionary Stable Strategies	373
9.3.5	Dynamical Behavior in the Rock-Scissors-Paper Game	374
9.3.6	Evolution of Cooperative Behavior	375
9.3.6.1	Kin Selection	376
9.3.6.2	Other Scenarios for Evolution of Cooperation	376
9.3.7	Yield and Efficiency in Metabolism	377
9.3.7.1	Trade-off Between Fast and Efficient Energy Metabolism	377
9.3.7.2	Multicellularity Enables Cells to Profit from Respiration	377
	References	379
10	Cell Biology	383
10.1	Introduction	383
10.2	The Origin of Life	384
10.3	Molecular Biology of the Cell	387
10.3.1	Chemical Bonds and Forces Important in Biological Molecules	387
10.3.2	Functional Groups in Biological Molecules	390
10.3.3	Major Classes of Biological Molecules	391
10.3.3.1	Carbohydrates	392
10.3.3.2	Lipids	392
10.3.3.3	Proteins	396
10.3.3.4	Nucleic Acids	400
10.4	Structural Cell Biology	402
10.4.1	Structure and Function of Biological Membranes	403
10.4.2	Nucleus	406
10.4.3	Cytosol	406
10.4.4	Mitochondria	407
10.4.5	Endoplasmic Reticulum and Golgi Complex	408
10.4.6	Other Organelles	409
10.5	Expression of Genes	410
10.5.1	Transcription	412
10.5.2	Processing of the mRNA	412
10.5.3	Translation	413
10.5.4	Protein Sorting and Posttranslational Modifications	415
10.5.5	Regulation of Gene Expression	416
	References	417
11	Experimental Techniques in Molecular Biology	419
11.1	Introduction	420
11.2	Restriction Enzymes and Gel Electrophoresis	420
11.3	Cloning Vectors and DNA Libraries	422
11.4	1D and 2D Protein Gels	425
11.5	Hybridization and Blotting Techniques	427
11.5.1	Southern Blotting	428
11.5.2	Northern Blotting	429
11.5.3	Western Blotting	429

11.5.4	<i>In Situ</i> Hybridization	430
11.6	Further Protein Separation Techniques	430
11.6.1	Centrifugation	430
11.6.2	Column Chromatography	431
11.6.3	Polymerase Chain Reaction	432
11.7	DNA and Protein Chips	433
11.7.1	DNA Chips	433
11.7.2	Protein Chips	434
11.8	Yeast Two-Hybrid System	434
11.9	Mass Spectrometry	435
11.10	Transgenic Animals	436
11.11	RNA Interference	437
11.12	ChIP on Chip and ChIP-PET	439
11.13	Surface Plasmon Resonance	441
11.14	Population Heterogeneity and Single Entity Experiments	442
	References	444

12 Mathematics 449

12.1	Linear Modeling	449
12.1.1	Linear Equations	449
12.1.1.1	The Gaussian Elimination Algorithm	451
12.1.1.2	Systematic Solution of Linear Systems	452
12.1.2	Matrices	454
12.1.2.1	Basic Notions	454
12.1.2.2	Linear Dependency	454
12.1.2.3	Basic Matrix Operations	454
12.1.2.4	Dimension and Rank	456
12.1.2.5	Eigenvalues and Eigenvectors of a Square Matrix	457
12.2	Ordinary Differential Equations	458
12.2.1	Notions Regarding Differential Equations	459
12.2.2	Linearization of Autonomous Systems	461
12.2.3	Solution of Linear ODE Systems	462
12.2.4	Stability of Steady States	463
12.2.4.1	Global Stability of Steady States	465
12.2.5	Limit Cycles	466
12.3	Difference Equations	467
12.4	Graph and Network Theory	469
12.4.1	Linear Networks	471
12.4.2	Boolean Networks	471
12.4.3	Bayesian Networks	473
	References	474

13 Statistics 475

13.1	Basic Concepts of Probability Theory	475
13.1.1	Random Variables, Densities, and Distribution Functions	478

13.1.2	Transforming Probability Densities	481
13.1.3	Product Experiments and Independence	482
13.1.4	Limit Theorems	483
13.2	Descriptive Statistics	483
13.2.1	Statistics for Sample Location	484
13.2.2	Statistics for Sample Variability	485
13.2.3	Density Estimation	486
13.2.4	Correlation of Samples	487
13.3	Testing Statistical Hypotheses	488
13.3.1	Statistical Framework	489
13.3.2	Two Sample Location Tests	491
13.4	Linear Models	493
13.4.1	ANOVA	493
13.4.2	Multiple Linear Regression	495
13.5	Principal Component Analysis	496
	References	499
14	Stochastic Processes	501
14.1	Basic Notions for Random Processes	501
14.1.1	Reduced and Conditional Distributions	503
14.2	Markov Processes	505
14.2.1	Markov Chains	506
14.3	Jump Processes in Continuous Time: The Master Equation	507
14.4	Continuous Random Processes	508
14.4.1	Langevin Equations	508
14.4.2	The Fokker–Planck Equation	509
	References	510
15	Control of Linear Systems	511
15.1	Linear Dynamical Systems	511
15.2	System Response	512
15.2.1	Random Fluctuations and Spectral Density	514
15.3	The Gramian Matrices	515
16	Databases	517
16.1	Databases of the National Center for Biotechnology	517
16.2	Databases of the European Bioinformatics Institute	518
16.2.1	EMBL Nucleotide Sequence Database	519
16.2.2	Ensembl	519
16.2.3	InterPro	519
16.3	Swiss-Prot, TrEMBL, and UniProt	520
16.4	Protein Databank	520
16.5	BioNumbers	521
16.6	Gene Ontology	521
16.7	Pathway Databases	524

16.7.1 ConsensusPathDB 524
References 525

17 Modeling Tools 527

17.1 Introduction 527
17.2 Mathematica and Matlab 528
17.2.1 Mathematica Example 530
17.2.2 Matlab Example 531
17.3 Dizzy 532
17.4 Systems Biology Workbench 534
17.5 Tools Compendium 536
References 551

Index 553