

Contents

Preface	V
Foreword	IX

Part I General Introduction

1	Basic Principles	3
1.1	Systems Biology is Biology!	3
1.2	Systems Biology is Modeling	5
1.2.1	Properties of Models	6
1.2.1.1	Model Assignment is not Unique	6
1.2.1.2	System State	6
1.2.1.3	Steady States	7
1.2.1.4	Variables, Parameters, and Constants	7
1.2.1.5	Model Behavior	8
1.2.1.6	Process Classification	8
1.2.1.7	Purpose and Adequateness of Models	8
1.2.1.8	Advantages of Computational Modeling	8
1.2.1.9	Model Development	9
1.2.2	Typical Aspects of Biological Systems and Corresponding Models	10
1.2.2.1	Network Versus Elements	10
1.2.2.2	Modularity	10
1.2.2.3	Robustness and Sensitivity are Two Sides of the Same Coin	11
1.3	Systems Biology is Data Integration	11
1.4	Systems Biology is a Living Science	14
	References	15
2	Biology in a Nutshell	19
	Introduction	19
2.1	The Origin of Life	20
2.2	Molecular Biology of the Cell	23
2.2.1	Chemical Bonds and Forces Important in Biological Molecules	23
2.2.2	Functional Groups in Biological Molecules	26

2.2.3	Major Classes of Biological Molecules	27
2.2.3.1	Carbohydrates	27
2.2.3.2	Lipids	27
2.2.3.3	Proteins	31
2.2.3.4	Nucleic Acids	35
2.3	Structural Cell Biology	37
2.3.1	Structure and Function of Biological Membranes	38
2.3.2	Nucleus	40
2.3.3	Cytosol	41
2.3.4	Mitochondria	42
2.3.5	Endoplasmatic Reticulum and Golgi Complex	43
2.3.6	Other Organelles	44
2.4	Expression of Genes	45
2.4.1	Transcription	45
2.4.2	Processing of the mRNA	47
2.4.3	Translation	48
2.4.4	Protein Sorting and Posttranslational Modifications	50
2.4.5	Regulation of Gene Expression	51
2.5	Cell Cycle	52
2.5.1	Mitosis	54
2.5.2	Meiosis and Genetic Recombination	54
	References	55
3	Mathematics in a Nutshell	57
	Introduction	57
3.1	Linear Algebra	57
3.1.1	Linear Equations	57
3.1.1.1	The Gaussian Elimination Algorithm	59
3.1.1.2	Systematic Solution of Linear Systems	60
3.1.2	Matrices	62
3.1.2.1	Basic Notions	62
3.1.2.2	Linear Dependency	62
3.1.2.3	Basic Matrix Operations	62
3.1.2.4	Dimension and Rank	64
3.1.2.5	Eigenvalues and Eigenvectors of a Square Matrix	65
3.2	Ordinary Differential Equations	66
3.2.1	Notions	67
3.2.2	Linearization of Autonomous Systems	69
3.2.3	Solution of Linear ODE Systems	70
3.2.4	Stability of Steady States	71
3.2.4.1	Global Stability of Steady States	73
3.2.4.2	Limit Cycles	74
3.3	Difference Equations	75
3.4	Statistics	77
3.4.1	Basic Concepts of Probability Theory	78

3.4.1.1	Random Variables, Densities, and Distribution Functions	81
3.4.1.2	Transforming Probability Densities	84
3.4.1.3	Product Experiments and Independence	84
3.4.1.4	Limit Theorems	85
3.4.2	Descriptive Statistics	86
3.4.2.1	Statistics for Sample Location	86
3.4.2.2	Statistics for Sample Variability	87
3.4.2.3	Density Estimation	88
3.4.2.4	Correlation of Samples	89
3.4.3	Testing Statistical Hypotheses	91
3.4.3.1	Statistical Framework	91
3.4.3.2	Two-sample Location Tests	93
3.4.4	Linear Models	96
3.4.4.1	ANOVA	96
3.4.4.2	Multiple Linear Regression	98
3.5	Graph and Network Theory	99
3.5.1	Introduction	100
3.5.2	Regulatory Networks	101
3.5.2.1	Linear Networks	101
3.5.2.2	Boolean Networks	101
3.5.2.3	Bayesian Networks	102
3.6	Stochastic Processes	103
3.6.1	Gillespie's Direct Method	105
3.6.2	Other Algorithms	105
3.6.3	Stochastic and Macroscopic Rate Constants	106
3.6.3.1	First-order Reaction	106
3.6.3.2	Second-order Reaction	107
	References	107
4	Experimental Techniques in a Nutshell	109
	Introduction	109
4.1	Elementary Techniques	109
4.1.1	Restriction Enzymes and Gel Electrophoresis	109
4.1.2	Cloning Vectors and DNA Libraries	113
4.1.3	1D and 2D Protein Gels	117
4.1.4	Hybridization and Blotting Techniques	119
4.1.4.1	Southern Blotting	120
4.1.4.2	Northern Blotting	121
4.1.4.3	Western Blotting	121
4.1.4.4	<i>In situ</i> Hybridization	121
4.1.5	Further Protein Separation Techniques	122
4.1.5.1	Centrifugation	122
4.1.5.2	Column Chromatography	123
4.2	Advanced Techniques	124
4.2.1	PCR	124

4.2.2	DNA and Protein Chips	126
4.2.2.1	DNA Chips	126
4.2.2.2	Protein Chips	127
4.2.3	Yeast Two-hybrid System	128
4.2.4	Mass Spectrometry	129
4.2.5	Transgenic Animals	130
4.2.6	RNA Interference	131
	References	133

Part II Standard Models and Approaches in Systems Biology

5	Metabolism	137
	Introduction	137
5.1	Enzyme Kinetics and Thermodynamics	140
5.1.1	The Law of Mass Action	141
5.1.2	Reaction Kinetics and Thermodynamics	142
5.1.3	Michaelis-Menten Kinetics	144
5.1.3.1	How to Derive a Rate Equation	146
5.1.3.2	Parameter Estimation and Linearization of the Michaelis-Menten Equation	147
5.1.3.3	The Michaelis-Menten Equation for Reversible Reactions	148
5.1.4	Regulation of Enzyme Activity by Protein Interaction	149
5.1.5	Inhibition by Irreversible Binding of Inhibitor to Enzyme	152
5.1.6	Substrate Inhibition	152
5.1.7	Inhibition by Binding of Inhibitor to Substrate	153
5.1.8	Binding of Ligands to Proteins	153
5.1.9	Positive Homotropic Cooperativity and the Hill Equation	154
5.1.10	The Monod-Wyman-Changeux Rate Expression for Enzymes with Sigmoid Kinetics	155
5.2	Metabolic Networks	157
5.2.1	Systems Equations	158
5.2.2	Information Contained in the Stoichiometric Matrix N	159
5.2.3	Elementary Flux Modes and Extreme Pathways	162
5.2.4	Flux Balance Analysis	164
5.2.5	Conservation Relations: Null Space of N^T	165
5.2.6	Compartments and Transport across Membranes	168
5.2.7	Characteristic Times	168
5.2.8	Approximations Based on Timescale Separation	171
5.2.8.1	The Quasi-steady-state Approximation	171
5.2.8.2	Quasi-equilibrium Approximation	172
5.3	Metabolic Control Analysis	174
5.3.1	The Coefficients of Control Analysis	175
5.3.1.1	The Elasticity Coefficients	177
5.3.1.2	Control Coefficients	178

5.3.1.3	Response Coefficients	180
5.3.1.4	Matrix Representation of the Coefficients	180
5.3.2	The Theorems of Metabolic Control Theory	181
5.3.2.1	The Summation Theorems	181
5.3.2.2	The Connectivity Theorems	183
5.3.2.3	Derivation of Matrix Expressions for Control Coefficients	185
5.3.3	Extensions of Metabolic Control Analysis	191
5.3.3.1	Control Analysis for Variables other than Fluxes and Concentrations	191
5.3.3.2	Time-dependent Control Coefficients	193
5.3.3.3	Spatially Heterogeneous and Time-varying Cellular Reaction Networks	194
	Suggested Further Reading	195
	References	196
6	Signal Transduction	201
	Introduction	201
6.1	Function and Structure of Intra- and Intercellular Communication	202
6.2	Receptor-Ligand Interactions	203
6.3	Structural Components of Signaling Pathways	206
6.3.1	G Proteins	206
6.3.2	Ras Proteins	208
6.3.3	Phosphorelay Systems	209
6.3.4	MAP Kinase Cascades	211
6.3.5	Jak-Stat Pathways	216
6.4	Signaling: Dynamic and Regulatory Features	217
6.4.1	Simple Motifs	217
6.4.2	Adaptation Motif	219
6.4.3	Negative Feedback	220
6.4.4	Quantitative Measures for Properties of Signaling Pathways	220
	References	223
7	Selected Biological Processes	225
	Introduction	225
7.1	Biological Oscillations	225
7.1.1	Glycolytic Oscillations: The Higgins-Sel'kov Oscillator	226
7.1.2	Other Modes of Behavior	229
7.1.3	Coupling of Oscillators	230
7.1.4	Sustained Oscillations in Signaling Cascades	233
7.2	Cell Cycle	234
7.2.1	Steps in the Cycle	235
7.2.2	Minimal Cascade Model of a Mitotic Oscillator	236
7.2.3	Models of Budding Yeast Cell Cycle	238
7.3	Aging	240
7.3.1	Evolution of the Aging Process	241

7.3.2	Accumulation of Defective Mitochondria	245
7.3.2.1	Synthesis Rates	247
7.3.2.2	Radical Levels	248
7.3.2.3	Dilution of Membrane Damage	248
7.3.2.4	The Equations	249
7.3.2.5	Choice of Parameters and Simulation Results	251
	References	254
8	Modeling of Gene Expression	257
	Introduction	257
8.1	Modules of Gene Expression	258
8.2	Promoter Identification	259
8.2.1	General Promoter Structure	260
8.2.2	Sequence-based Prediction of Promoter Elements	262
8.2.3	Approaches that Incorporate Additional Information	264
8.3	Modeling Specific Processes in Eukaryotic Gene Expression	265
8.3.1	One Example, Different Approaches	266
8.3.1.1	Description with Ordinary Differential Equations	266
8.3.1.2	Representation of Gene Network as Directed and Undirected Graphs	269
8.3.1.3	Bayesian Networks	270
8.3.1.4	Boolean Networks	270
8.3.1.5	Gene Expression Modeling with Stochastic Equations	273
8.3.2	Time Delay in Gene Regulation	274
8.3.3	Modeling the Elongation of a Peptide Chain	276
8.4	Modeling the Regulation of Operons in <i>E. coli</i>	278
8.4.1	Mechanism of the Lac Operon in <i>E. coli</i>	278
8.4.2	The Model According to Griffith	280
8.4.3	The Model According to Nicolis and Prigogine	282
	References	286
9	Analysis of Gene Expression Data	289
	Introduction	289
9.1	Data Capture	289
9.1.1	DNA Array Platforms	289
9.1.2	Image Analysis and Data Quality Control	291
9.1.2.1	Grid Finding	291
9.1.2.2	Quantification of Signal Intensities	292
9.1.2.3	Signal Validity	293
9.1.3	Pre-processing	296
9.1.3.1	Global Measures	296
9.1.3.2	Linear Model Approaches	297
9.1.3.3	Nonlinear and Spatial Effects	297
9.1.3.4	Other Approaches	298
9.2	Fold-change Analysis	299

9.2.1	Planning and Designing Experiments	299
9.2.2	Tests for Differential Expression	301
9.2.3	Multiple Testing	303
9.2.4	ROC Curve Analysis	306
9.2.5	Validation Methods	307
9.3	Clustering Algorithms	307
9.3.1	Hierarchical Clustering	311
9.3.2	Self-organizing Maps (SOMs)	314
9.3.3	K-means	315
9.4	Validation of Gene Expression Data	316
9.4.1	Cluster Validation	316
9.4.2	Principal Component Analysis	318
9.4.3	Functional Categorization	321
9.5	Classification Methods	322
9.5.1	Basic Concepts	322
9.5.2	Support Vector Machines	323
9.5.3	Other Approaches	326
9.5.4	Cross-validation	327
9.5.4.1	The Holdout Method	327
9.5.4.2	k-fold Cross-validation	327
9.5.4.3	Leave-one-out Cross-validation	327
9.6	Reverse Engineering Genetic Networks	328
9.6.1	Reconstructing Boolean Networks	328
9.6.2	Other Approaches	330
9.6.3	Network Motifs	331
	References	333
10	Evolution and Self-organization	337
	Introduction	337
10.1	Quasispecies and Hypercycles	338
10.1.1	Selection Equations for Biological Macromolecules	339
10.1.1.1	Self-replication Without Interaction	340
10.1.1.2	Selection at Constant Total Concentration of Self-replicating Molecules	340
10.1.1.3	Self-replication with Mutations: The Quasispecies Model	342
10.1.1.4	The Genetic Algorithm	343
10.1.1.5	Assessment of Sequence Length for Stable Passing-on of Sequence Information	344
10.1.1.6	Coexistence of Self-replicating Sequences: Complementary Replication of RNA	345
10.1.2	The Hypercycle	346
10.2	Other Mathematical Models of Evolution	349
10.2.1	Spin-glass Model of Evolution	349
10.2.2	Neutral Theory of Molecular Evolution	351
10.2.3	Boolean Network Models	352

10.3	Prediction of Biological Systems from Optimality Principles	355
10.3.1	Optimization of Catalytic Properties of Single Enzymes	356
10.3.2	Optimal Distribution of Enzyme Concentrations in a Metabolic Pathway	359
10.3.3	Temporal Transcription Programs	362
	References	364
11	Data Integration	367
	Introduction	367
11.1	Database Networks	368
11.1.1	Basic Concepts of Database Integration	369
11.1.2	SRS	370
11.1.3	EnsMart	371
11.1.4	DiscoveryLink	372
11.1.5	Data Exchange	372
11.2	Information Measurement in Heterogeneous Data	374
11.2.1	Information and Entropy	374
11.2.2	Mutual Information	376
11.2.3	Information Correlation: Example	379
11.3	Biclustering	381
11.3.1	The Problem	381
11.3.2	Algorithmic Example	382
11.3.3	Biclustering and Data Integration	384
	References	385
12	What's Next?	387
12.1	Systems Biology: The Core of Biological Research and Medical Practice of the Future?	387
12.2	Experimental Planning in the Systems Biology Phase of Biological Research	388
12.3	Publication in the Era of Systems Biology	389
12.4	Systems Biology and Text Mining	389
12.5	Systems Biology in Medicine	390
12.6	Systems Biology in Drug Development	390
12.7	Systems Biology in Food Production and Biotechnology	391
12.8	Systems Biology in Ecology	391
12.9	Systems Biology and Nanotechnology	391
12.10	Guiding the Design of New Organisms	392
12.11	Computational Limitations	393
12.12	Potential Dangers	394
	References	394

Part III Computer-based Information Retrieval and Examination

13	Databases and Tools on the Internet	399
	Introduction	399
13.1	Gene Ontology	399
13.2	KEGG	403
13.3	BRENDA	404
13.4	Databases of the National Center for Biotechnology	405
13.5	Databases of the European Bioinformatics Institute	406
13.5.1	EMBL Nucleotide Sequence Database	407
13.5.2	Ensembl	407
13.5.3	InterPro	408
13.6	Swiss-Prot, TrEMBL, and UniProt	408
13.7	Reactome	409
13.8	PDB	410
13.9	TRANSFAC and EPD	413
13.9.1	TRANSFAC	413
13.9.2	EPD	414
13.10	Genome Matrix	415
	References	417
14	Modeling Tools	419
	Introduction	419
14.1	Modeling and Visualization	419
14.1.1	Mathematica and Matlab	419
14.1.1.1	Mathematica Example	421
14.1.1.2	Matlab Example	422
14.1.2	Gepasi	422
14.1.3	E-Cell	424
14.1.4	PyBioS	426
14.1.5	Systems Biology Workbench	428
14.1.5.1	JDesigner	429
14.1.5.2	CellDesigner	431
14.1.6	Petri Nets	433
14.1.7	STOCKS 2	435
14.1.8	Genetic Programming	438
14.2	Model Exchange Languages, Data Formats	440
14.2.1	Introduction to XML	440
14.2.2	Systems Biology Markup Language	442
14.2.3	MathML	447
	References	448