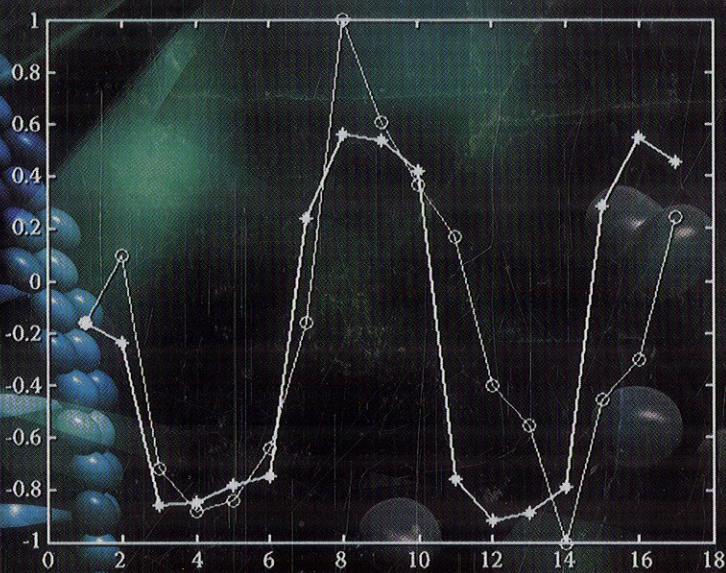


Systems Biology and Bioinformatics

A Computational Approach



Kayvan Najarian
Christopher N. Eichelberger
Siamak Najarian
Shahriar Gharibzadeh

Contents

Preface.....	XV
Acknowledgment	XXI
Authors.....	xxiii

Chapter 1

Cell Biology	1
1.1 Introduction and Overview	1
1.2 An Introduction to Cell Structure	1
1.2.1 Plasma Membrane	1
1.2.1.1 Membrane Phospholipids.....	1
1.2.1.2 Proteins	1
1.2.1.3 Fluidity	2
1.2.1.4 Stability	2
1.2.1.5 Capacitive Property.....	2
1.2.1.6 Surface for Protein Attachment	2
1.2.1.7 Control of Passage of Different Substances.....	2
1.2.2 Cytoplasm.....	3
1.2.2.1 Mitochondria.....	3
1.2.2.2 Endoplasmic Reticula	3
1.2.2.3 Golgi Apparatus.....	3
1.2.2.4 Lysosomes.....	3
1.2.2.5 Ribosomes.....	3
1.2.3 Nucleus	3
1.3 Proteins as Tools of the Cell	4
1.3.1 Structure of Proteins.....	4
1.3.1.1 Higher Levels of Protein Organization.....	4
1.3.2 Functions of Different Proteins	4
1.3.2.1 Structural Proteins	4
1.3.2.2 Transport Proteins.....	5
1.3.2.3 Receptors	5
1.3.2.4 Immunological Proteins.....	5
1.3.2.5 Blood Carrier Proteins.....	6
1.3.2.6 Enzymes.....	6
1.4 Genes: Directors of the Cell	6
1.4.1 Introduction to Inheritance	6
1.4.2 The Structure of Genes.....	6
1.4.2.1 Chromosomes and Chromatin	7
1.4.2.2 Genetic Code.....	8
1.4.3 DNA Replication	8
1.4.3.1 Initiation.....	8

1.4.3.2	Elongation	9
1.4.3.3	Termination	10
1.4.4	Transcription	10
1.4.4.1	Three Kinds of RNA	11
1.4.4.2	tRNA Anticodons	11
1.4.5	Translation	12
1.5	Cell Cycle	12
1.5.1	Cell Cycle: An Overview	12
1.5.2	Mitosis	12
1.5.2.1	Mitotic Apparatus	12
1.5.2.2	Four Stages of Mitosis	13
1.6	Summary	13
1.7	Problems	13

Chapter 2

Bioassays	15	
2.1	Introduction and Overview	15
2.2	Southern Blotting	15
2.3	Fluorescent In Situ Hybridization	15
2.4	Sanger (Dideoxy) Method	16
2.5	Polymerase Chain Reaction	16
2.6	Analyzing Protein Structure and Function	17
2.6.1	X-ray Crystallography	17
2.6.2	NMR Spectroscopy	17
2.6.3	Affinity Chromatography	17
2.6.4	Immunoprecipitation	17
2.6.5	Two-Hybrid Systems	18
2.6.6	Phage Display Method	18
2.6.7	DNA Footprint Analysis	18
2.7	Studying Gene Expression and Function—DNA Microarrays	18
2.7.1	Spotted Microarrays	19
2.7.2	Oligonucleotide Microarrays	19
2.7.3	SNP Microarrays	19
2.8	Summary	19
2.9	Problems	20

Chapter 3

Review of Some Computational Methods	21	
3.1	Introduction and Overview	21
3.2	Introduction to Probability and Stochastic Processes Theories	21
3.2.1	Metrics for Stochastic Processes	22
3.2.1.1	Probability Density Function	22
3.2.1.2	Mean and Variance	23
3.2.1.3	Expectation Function	23

3.2.3	Gaussian Distribution	24
3.2.4	Correlation Function	24
3.2.5	Stationary and Ergodic Stochastic Processes	26
3.3	Bayesian Theory	28
3.3.1	From Binary Networks to Bayesian Networks	28
3.4	Test of Hypothesis	29
3.4.1	t Test	30
3.5	Expectation Maximization Method	31
3.6	Maximum Likelihood Theory	31
3.6.1	Kullback-Leibler Distance	32
3.6.2	Log-Likelihood Function	32
3.6.3	Akaike's Information Criterion	33
3.6.4	Bayesian Information Criterion	33
3.7	System Identification Theory	34
3.7.1	Autoregressive Exogenous Model	34
3.7.2	Least Square Estimation	37
3.8	Summary	38
3.9	Problems	39

Chapter 4

	Computational Structural Biology: Protein Structure Prediction	41
4.1	Introduction and Overview	41
4.2	Protein Structure Prediction Methods	42
4.2.1	Ab Initio Methods: Molecular Dynamics	42
4.2.2	Comparative Methods	42
4.2.2.1	Homology-Based Methods	43
4.2.2.2	Threading Methods	43
4.2.3	Hybrid/Informed Methods	43
4.2.3.1	Artificial Neural Networks	43
4.2.3.2	Hidden Markov Models	44
4.2.3.3	Support Vector Machines	45
4.2.3.4	Clustering Methods	45
4.2.3.5	Blended Methods	45
4.3	Data Resources	46
4.3.1	Protein Sequence and Raw Structure Databases	46
4.3.2	Protein Structure Classification Databases	47
4.3.2.1	The SCOP Database	47
4.3.2.2	Class, Architecture, Topology, and Homologous Superfamily	47
4.3.2.3	Molecular Modeling Database	48
4.3.2.4	Fold Classification Based on Structure-Structure Alignment of Proteins (FSSP/DALI) Database	48
4.3.2.5	Spatial Arrangement of Backbone Fragments Database	49
4.3.2.6	InterPro	49

4.4	Summary	50
4.5	Problems	50
	References.....	50

Chapter 5

	Computational Structural Biology: Protein Sequence Analysis	51
5.1	Introduction	51
5.2	Pairwise Sequence Matching.....	52
5.2.1	Pairwise Background.....	52
5.2.2	Sequence Matching Methods.....	52
5.2.2.1	Dot Plot Visualization Method	53
5.2.2.2	Dynamic Programming	53
5.2.2.3	Word Methods.....	54
5.2.2.4	Bayesian Methods	55
5.2.3	Statistical Significance.....	55
5.3	Multiple Sequence Alignment	55
5.3.1	Dynamic Programming	56
5.3.2	Progressive Methods.....	56
5.3.3	Iterative Strategies	56
5.4	Summary	56
5.5	Problems	57

Chapter 6

	Genomics and Proteomics	59
6.1	Introduction and Overview	59
6.2	Genomics	59
6.2.1	Genetic Sequence Data.....	59
6.2.2	Genetic Annotations	60
6.2.3	Genomic Analysis.....	60
6.2.4	Metagenomics.....	61
6.3	Proteomics	61
6.3.1	Protein Sequence Data.....	62
6.3.2	Protein Annotations	62
6.3.3	Proteomic Analysis.....	63
6.4	Summary	63
6.5	Problems	64

Chapter 7

	Methods for Identification of Differentially Expressed Genes or Proteins	65
7.1	Introduction and Overview	65
7.2	Why <i>t</i> Test Is Not Enough?	65
7.3	Mixture Model Method	66

7.3.1	Basic Formulations	66
7.3.2	More Advanced Versions.....	68
7.4	Genes Involved in Leukemia: A Case Study	70
7.5	Summary	72
7.6	Problems	72
	References.....	72

Chapter 8

	Binary and Bayesian Networks as Static Models of Regulatory Pathways	73
8.1	Introduction and Overview	73
8.2	Binary Regulator Pathways	73
8.3	Bayesian Networks: Algorithm.....	73
8.4	Applications and Practical Considerations	74
8.5	Summary	76
8.6	Problems	76

Chapter 9

	Metabolic Control Theory for Static Modeling of Metabolic Pathways.....	77
9.1	Introduction and Overview	77
9.2	Basic Ideas	77
9.3	Main Concepts in Metabolic Control Theory	79
9.3.1	Step 1	79
9.3.2	Step 2	81
9.3.3	Step 3	82
9.4	Mason Method	83
9.5	Metabolic Control Model for Galactose Regulation Pathway: A Case Study	86
9.6	Summary	92
9.7	Problems	93
	References.....	93

Chapter 10

	System Identification and Control Theory for Dynamic Modeling of Biological Pathways.....	95
10.1	Introduction and Overview	95
10.2	Methodology	95
10.2.1	Nonlinear Factor Analysis.....	97
10.2.2	ARX Model.....	98
10.2.3	Linearized Model: Extracting Pathway Links from Nonlinear Model	101
10.3	Modeling of Cell Cycle: A Case Study.....	102
10.3.1	Stability Analysis of System	103

10.4	Summary	106
10.5	Problems	106
	References	107

Chapter 11

	Gene Silencing for Systems Biology	109
11.1	Introduction and Overview	109
11.2	A Brief Review of RNA Interference, Small Interfering RNA, and Gene Silencing Mechanisms	109
	11.2.1 Pretranscriptional Gene Silencing	109
	11.2.2 Posttranscriptional Gene Silencing	110
	11.2.3 RNA Interference	110
11.3	Pathway Perturbation Using Gene Silencing	111
	11.3.1 Overexpression	111
	11.3.2 Gene Silencing	112
	11.3.3 Model Improvement and Validation	112
11.4	Summary	113
11.5	Problems	113
	References	113

Chapter 12

	Simulation and Systems Biology	115
12.1	Introduction and Overview	115
12.2	What Is Simulation?	115
	12.2.1 Power of Expression	116
	12.2.2 Economy of Form	116
12.3	Challenges to Effective Simulation	117
	12.3.1 Identifying Meaningful Themes	117
	12.3.2 Detail Selection and Rejection	118
	12.3.3 Validation	119
	12.3.3.1 Measuring Quality Once	119
	12.3.3.2 Handling Stochastic Results	119
12.4	Case Studies	120
	12.4.1 Simulated Protein Folding: Rosetta@Home	120
	12.4.1.1 Generally How It Works	121
	12.4.1.2 Details to Use or Discard	121
	12.4.1.3 Computing Power	122
	12.4.1.4 Future Goals	123
	12.4.2 Simulated Protein Computing	124
	12.4.2.1 Generally How It Works	124
	12.4.2.1 Details to Use or Discard	125
	12.4.2.3 Computing Power	128
	12.4.2.4 Future Goals	128

12.5 Summary	129
12.6 Problems	129
References and Notes.....	129

Chapter 13

Software, Databases, and Other Resources for Systems Biology.....	131
13.1 Introduction and Overview	131
13.2 Databases	132
13.2.1 Nucleotide Databases	132
13.2.1.1 GenBank	132
13.2.1.2 DNA Data Bank of Japan	132
13.2.1.3 European Molecular Biology Laboratory.....	132
13.2.1.4 International Sequence Database.....	132
13.2.2 Protein Database	133
13.2.2.1 Swiss-Prot.....	133
13.2.2.2 Protein Information Resource.....	133
13.2.2.3 Translated EMBL	133
13.2.2.4 Universal Protein Resource	133
13.3 Scoring Matrix	134
13.3.1 Point Accepted Mutation.....	134
13.3.2 Blocks Substitution Matrix	134
13.4 Analysis Software	135
13.4.1 FASTA.....	135
13.4.1.1 FASTA Algorithm.....	135
13.4.1.2 FASTA Variants	137
13.4.2 BLAST	138
13.4.2.1 BLAST Algorithm.....	138
13.4.2.2 BLAST Variants.....	140
13.5 Bioinformatics in MATLAB	140
13.5.1 Sequence Analysis	140
13.5.2 Sequence Alignment	141
13.6 Summary	141
13.7 Problems	141
References.....	141

Chapter 14

Future Directions	143
14.1 Introduction and Overview	143
14.2 Single-Cell Microarray and Systems Biology	143
14.2.1 Microarrays to Identify Single Cells.....	143
14.2.2 Microarray Analysis of a Single Cell.....	145
14.3 High-Throughput Protein Assays and Systems Biology.....	146
14.4 Integration of Molecular Data with Higher-Level Datasets.....	147

14.4.1	Higher-Level Data Sets and Publishers.....	148
14.4.2	Tools to Navigate Metadata.....	149
14.4.3	Role of Higher-Level Data in Systems Biology	149
14.5	Identifying Genes Controlling Macrolevel Changes	150
14.5.1	Generate Candidate Markers	151
14.5.2	Correlate Markers with Phenotype	151
14.5.3	Infer Genes or Gene Expression Patterns	152
14.6	Molecular-Level Image Systems and Systems Biology.....	153
14.6.1	Molecular-Level Imaging.....	153
14.6.2	Visual Analytics.....	154
14.7	Summary	155
14.8	Problems	155
	References and Notes.....	156
	Index	159