
E. Klipp, R. Herwig, A. Kowald, C. Wierling, H. Lehrach

Systems Biology in Practice

Concepts, Implementation and Application



WILEY-VCH Verlag GmbH & Co. KGaA

Dr. Edda Klipp
Dr. Ralf Herwig
Dr. Axel Kowald
Christoph Wierling
Prof. Dr. Hans Lehrach
MPI für Molekulare Genetik
Ihnestraße 73
14195 Berlin
Germany

■ All books published by Wiley-VCH are carefully produced. Nevertheless, authors, editors, and publisher do not warrant the information contained in these books, including this book, to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

Library of Congress Card No. applied for

British Library Cataloguing-in-Publication

Data: A catalogue record for this book is available from the British Library.

Die Deutsche Bibliothek –

CIP Cataloguing-in-Publication Data:

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>

© 2005 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

All rights reserved (including those of translation into other languages). No part of this book may be reproduced in any form – by photoprinting, microfilm, or any other means – nor transmitted or translated into machine language without written permission from the publishers. Registered names, trademarks, etc. used in this book, even when not specifically marked as such, are not to be considered unprotected by law.

Printed in the Federal Republic of Germany
Printed on acid-free paper

Composition ProSatz Unger, Weinheim
Printing betz-druck GmbH, Darmstadt
Bookbinding Litges & Dopf Buchbinderei GmbH, Heppenheim

ISBN-13: 978-3-527-31078-4

ISBN-10: 3-527-31078-9

Preface

Systems biology is the coordinated study of biological systems by (1) investigating the components of cellular networks and their interactions, (2) applying experimental high-throughput and whole-genome techniques, and (3) integrating computational methods with experimental efforts. In this book we attempt to give a survey of this rapidly developing field. The systematic approach to biology is not new, but it has recently gained new attraction due to emerging experimental and computational methods. This book is intended as an introduction for students of biology, biophysics, and bioinformatics and for advanced researchers approaching systems biology from a different discipline.

We see the origin and the methodological foundations for systems biology (1) in the accumulation of detailed biological knowledge with the prospect of utilization in biotechnology and health care, (2) in the emergence of new experimental techniques in genomics and proteomics, (3) in the tradition of mathematical modeling of biological processes, (4) in the developing computer power as a prerequisite for databases and for the calculation of large systems, and (5) in the Internet as *the* medium for quick and comprehensive exchange of information.

Recently, researchers working in different fields of biology have expressed the need for systematic approaches. They have frequently demanded the establishment of computer models of biochemical and signaling networks in order to arrive at testable quantitative predictions despite the complexity of these networks. For example, Hartwell and colleagues (1999) argue that “[t]he best test of our understanding of cells will be to make quantitative predictions about their behavior and test them. This will require detailed simulations of the biochemical processes taking place within [cells]. ... We need to develop simplifying, higher-level models and find general principles that will allow us to grasp and manipulate the functions of [biochemical networks].” Fraser and Harland (2000) state, “As the sophistication of the data collection improves, so does the challenge of fully harvesting the fruits of these efforts. The results to date show a dizzying array of signaling systems acting within and between cells. ... In such settings, intuition can be inadequate, often giving incomplete or incorrect predictions. ... In the face of such complexity, computational tools must be employed as a tool for understanding.” Noble laureate Nurse (2000) writes, “Perhaps a proper understanding of the complex regulatory networks making up cellular systems like the cell cycle will require a ... shift from

common sense thinking. We might need to move into a strange more abstract world, more readily analyzable in terms of mathematics." And Kitano (2002 a) emphasizes that "computational biology, through pragmatic modeling and theoretical exploration, provides a powerful foundation from which to address critical scientific questions head-on."

The requirement to merge experimental techniques and theoretical concepts in the investigation of biological objects has been acknowledged, for example, by Kitano (2002 a): "To understand complex biological systems requires the integration of experimental and computational research – in other words a systems biology approach." Levchenko (2003) recommends "the systems biology approach, relying on computational modeling coupled with various experimental techniques and methodologies, ... combining the dynamical view of rapidly evolving responses and the structural view arising from high-throughput analyses of the interacting species." Ideker and colleagues (2001) state, "Systems biology studies biological systems by systematically perturbing them (biologically, genetically, or chemically); monitoring the gene, protein, and informational pathway responses; integrating these data; and ultimately, formulating mathematical models that describe the structure of the system and its response to individual perturbations."

Aebersold and colleagues (2000) see the fundamental experimental contribution in large-scale facilities for genome-wide analyses, including DNA sequencing, gene expression measurements, and proteomics, while Hood (2003) explains his path to systems biology in the following way: "Our view and how we practice biology have been profoundly changed by the Human Genome Project."

Importantly, it has been discovered that cellular regulation is organized into complex networks and that the various interactions of network elements in time and space must be studied. Kitano (2002 b) stresses that "[t]o understand biology at the system level, we must examine the structure and dynamics of cellular and organismal function, rather than the characteristics of isolated parts of a cell or organism. Properties of systems, such as robustness, emerge as central issues, and understanding these properties may have an impact on the future of medicine." Kholodenko and colleagues want to "untangle the wires" and "trace the functional interactions in signaling and gene networks." Levchenko (2003) sees advantages in understanding signaling: "A new view of signaling networks as systems consisting of multiple complex elements interacting in a multifarious fashion is emerging, a view that conflicts with the single-gene or protein-centric approach common in biological research. The postgenomic era has brought about a different, network-centric methodology of analysis, suddenly forcing researchers toward the opposite extreme of complexity, where the networks being explored are, to a certain extent, intractable and uninterpretable."

There are many fields of application besides the understanding of cellular regulation. With respect to modeling of the heart as whole organ, Noble (2002) discusses that "[s]uccessful physiological analysis requires an understanding of the functional interactions between the key components of cells, organs, and systems, as well as how these interactions change in disease states. This information resides neither in the genome nor even in the individual proteins that genes code for. It lies at the level of protein interactions within the context of subcellular, cellular, tissue, organ, and

system structures." Kirkwood and colleagues (2003) observe a need to apply "e-biology" on aging in order to integrate theory and data.

There is no need to add another definition of systems biology. More important than such a definition is the operational meaning and the *modus vivendi*. However, we would like to emphasize the view that although the *new* property of systems biology is the computational aspect, the trinity of experimentation, data handling, and mathematical modeling is crucial for further successful development of biological science.

Although deciphering of the DNA sequences of many organisms including man has been acknowledged as an important step towards the exact representation of biology, it is currently not possible to calculate the phenotype of an organism from genotype or to simulate a living cell using only the information encoded in these sequences. We will show in the following chapters what can be achieved at present. An old proverb states, "What you expect is what you will get." Knowledge of different concepts, methodologies, and sources of information will support researchers in interpreting their data in a broader context.

This book is divided into three parts. The first part gives an introduction to three main foundations of systems biology – cell biology, mathematics, and experimental techniques. This will be very basic for advanced readers but will prove helpful for those approaching systems biology from a different scientific discipline.

The second part of the book presents current strategies of computational modeling and data mining. It covers in detail various cellular processes such as metabolism, signaling, the cell cycle, and gene expression, as well as the interactions between them. We introduce different concepts of modeling and discuss how the different models can be used to tackle a number of frequent problems, including such questions as how regulation is organized, how data can be interpreted, or which model to apply under specific settings.

The third part gives an overview on currently available help and resources from the Internet. We represent modeling tools that we frequently use ourselves. We also give an overview on databases that are indispensable for information exchange and therefore constitute an essential support for systems biology.

The ideas presented in this book rely on the work of many colleagues currently or formerly active in the field. Our contribution to systems biology has been influenced by many other scientists and our teachers, whom we wish to acknowledge.

We also thank a number of people who helped us in finishing this book. We are especially grateful to Bente Kofahl, Dr. Wolfram Liebermeister, and Dr. Damini Tapadar for reading and commenting on the manuscript. Hendrik Hache and Mario Drungowski contributed with data analysis. Parts of the experimental data used throughout the book were generated in collaboration with Dr. Marie-Laure Yaspo, Dr. James Adjaye and Dr. Pia Aanstad. We thank Monica Shevack for the artistic preparation of many figures.

E.K. wishes to thank her family for support, especially her sons for patience and hot dinners. R.H. thanks his family for supporting him throughout the course of writing. Funding from the following sources is appreciated: E.K. and A.K. are supported by the German Federal Ministry for Education and Research and by the Berlin Center of Genome Based Bioinformatics. C.W. is financed by the EU FP6 grant (LSHG-CT-2003–503269) and R.H. and H.L. by the Max Planck Society.

References

- AEBERSOLD, R., HOOD, L.E. and WATTS, J.D. Equipping scientists for the new biology (2000) *Nat. Biotechnol.* 18, 359
- FRASER, S.E. and HARLAND, R.M. The molecular metamorphosis of experimental embryology (2000) *Cell* 100, 41–55
- HARTWELL, L.H., HOPFIELD, J.J., LEIBLER, S. and MURRAY, A.W. From molecular to modular cell biology (1999) *Nature* 402, C47–52
- HOOD, L. Systems biology: integrating technology, biology, and computation (2003) *Mech. Ageing Dev.* 124, 9–16
- IDEKER, T., GALITSKI, T. and HOOD, L. A new approach to decoding life: systems biology (2001) *Annu. Rev. Genomics Hum. Genet.* 2, 343–72
- KIRKWOOD, T.B., BOYS, R.J., GILLESPIE, C.S., PROCTOR, C.J., SHANLEY, D.P. and WILKINSON, D.J. Towards an e-biology of ageing: integrating theory and data (2003) *Nat. Rev. Mol. Cell Biol.* 4, 243–9
- KITANO, H. Computational systems biology (2002a) *Nature* 420, 206–10
- KITANO, H. Systems biology: a brief overview (2002b) *Science* 295, 1662–4
- LEVCHENKO, A. Dynamical and integrative cell signaling: challenges for the new biology (2003) *Biotechnol. Bioeng.* 84, 773–82
- NOBLE, D. Modeling the heart—from genes to cells to the whole organ (2002) *Science* 295, 1678–82
- NURSE, P. A long twentieth century of the cell cycle and beyond (2000) *Cell* 100, 71–8

Foreword

Systems biology is an emergent discipline that is gaining increased attention. A desire to understand systems of living organisms is not a new one. It can be traced back a few decades. Walter Cannon's homeostasis, Norbert Wiener's cybernetics, and Ludwig von Bertalanffy's general systems theory all points to essentially the same direction – system-level understanding of biological systems. Since the discovery of double helix structure of DNA and a series of efforts that gave birth to molecular biology, astonishing progress has been made on our understanding on living forms as molecular machinery. The climax came as completion of human genome sequencing.

With accumulating knowledge of genes and proteins, the next natural question to ask is how they are working together? What are principles that govern at the system-level? With the progress of molecular biology, genomics, computer science, and control theory, the old question is now being revisited with new concepts and methodologies.

A system is not just an assembly of components. There are principles that govern at the system-level. Unlike genes and proteins that are rather tangible objects, a system is no tangible. The essence of the system lies in dynamics that is not tangible. This makes the game of systems biology complicated, and may sound alien to many molecular biologists who are accustomed to a molecular-oriented view of the world. Needless to say system-level understanding has to be grounded onto molecular-level so that a continuous spectrum of knowledge can be established.

The enterprise of systems biology research requires both breadth and depth of understanding for various aspects of biological, computational, mathematical, and even engineering issues. So far, there has not been a coherent textbook in the field that covers broad aspects of systems biology. (I wrote a textbook in 2001 perhaps the first textbook in systems biology, but it was only in Japanese.) In this textbook, the authors have successfully covered sufficiently broad aspects of biology and computation that is essential in getting started in systems biology research. It is essential that both computational and experimental aspects of biology are described consistently and seamlessly. The students who learned through this textbook will make no barrier between computation and experiments. They would use advanced computational tools just like using PCR. I am expecting to see a new generation of systems biologists who get the first touch of the field from this book.

Bon voyage

Tokyo, Japan, September 26 2004

Hiroaki Kitano

Systems Biology in Practice. Concepts, Implementation and Application.

E. Klipp, R. Herwig, A. Kowald, C. Wierling, H. Lehrach

Copyright © 2005 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

ISBN: 3-527-31078-9

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

XII | Contents

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	Endoplasmic 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

XIV | Contents

- 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	<i>k</i> -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
- Subject Index 451