

Under Construction:

In the meantime, please email or write for a hardcopy reprint:

Felix T. Hong  
Department of Physiology  
Wayne State University  
Detroit, Michigan 48201 USA

Email: [fhong@med.wayne.edu](mailto:fhong@med.wayne.edu)

# INFORMATION PROCESSING AND LIVING SYSTEMS

EDITORS

VLADIMIR B BAJIC & TAN TIN WEE



Imperial College Press

# INFORMATION PROCESSING AND LIVING SYSTEMS

## Key Features

- Contains a systematic and comprehensive survey of biocomputing not existing in the current literature
- Provides a broad overview of bioinformatics with a number of novel bioinformatics applications that illustrate some of the principles of biocomputing
- A unique source of information on the biological/physiological background on the biological "computing" processes that are performed in living systems, including higher cognitive processes
- Shows how some of these computing examples in biology have found their way into useful computing applications such as genetic algorithms

Information processing and information flow occur in the course of an organism's development and throughout its lifespan. Organisms do not exist in isolation, but interact with each other constantly within a complex ecosystem. The relationships between organisms, such as those between prey or predator, host and parasite, and between mating partners, are complex and multidimensional. In all cases, there is constant communication and information flow at many levels.

This book focuses on information processing by life forms and the use of information technology in understanding them. Readers are first given a comprehensive overview of biocomputing before navigating the complex terrain of natural processing of biological information using physiological and analogous computing models. The remainder of the book deals with "artificial" processing of biological information as a human endeavor in order to derive new knowledge and gain insight into life forms and their functioning. Specific innovative applications and tools for biological discovery are provided as the link and complement to biocomputing.

Since "artificial" processing of biological information is complementary to natural processing, a better understanding of the former helps us improve the latter. Consequently, readers are exposed to both domains and, when dealing with biological problems of their interest, will be better equipped to grasp relevant ideas.

Imperial College Press

[www.icpress.co.uk](http://www.icpress.co.uk)

P391 hc

ISBN 1-86094-563-5



9 781860 945632

**INFORMATION  
PROCESSING AND  
LIVING SYSTEMS**

**SERIES ON ADVANCES IN BIOINFORMATICS  
AND COMPUTATIONAL BIOLOGY**

**Series Editors:**

Ying XU (*University of Georgia, USA*)

Limsoon WONG (*Institute for Infocomm Research, Singapore*)

**Associate Editors:**

Ruth Nussinov (*NCI, USA*)

Rolf Apweiler (*EBI, UK*)

Ed Wingender (*BioBase, Germany*)

See-Kiong Ng (*Inst for Infocomm Res, Singapore*)

Kenta Nakai (*Univ of Tokyo, Japan*)

Mark Ragan (*Univ of Queensland, Australia*)

---

**Vol. 1: Proceedings of the 3rd Asia-Pacific Bioinformatics Conference**

*Eds: Yi-Ping Phoebe Chen and Limsoon Wong*

**Vol. 2: Information Processing and Living Systems**

*Eds: Vladimir B. Bajic and Tan Tin Wee*

# INFORMATION PROCESSING AND LIVING SYSTEMS

EDITORS

**Vladimir B Bajic**

INSTITUTE FOR INFOCOMM RESEARCH, SINGAPORE

**TAN TIN WEE**

NATIONAL UNIVERSITY OF SINGAPORE, SINGAPORE



Imperial College Press

*Published by*

Imperial College Press  
57 Shelton Street  
Covent Garden  
London WC2H 9HE

*Distributed by*

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

*USA office:* 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

*UK office:* 57 Shelton Street, Covent Garden, London WC2H 9HE

**British Library Cataloguing-in-Publication Data**

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

**INFORMATION PROCESSING AND LIVING SYSTEMS**

Copyright © 2005 by Imperial College Press

*All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.*

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 1-86094-563-5

Printed in Singapore by B & JO Enterprise

# Preface

This is a book on information processing by and information processing for living systems. It is about the Information of Life, of living computers and the human endeavour of computing life.

“... No man is an island, entire of itself; every man is a piece of the continent, a part of the main. If a clod be washed away by the sea, Europe is the less, as well as if a promontory were, as well as if a manor of thy friend’s or of thine own were: any man’s death diminishes me, because I am involved in mankind, and therefore never send to know for whom the bell tolls; it tolls for thee...”

*Meditation XVII, John Donne (1623)*

All living systems reproduce after their kind to perpetuate their lineage. The offsprings inherit characteristics from both parents and the basis for this inheritance lies with genetic material, whether it is DNA for most unicellular or higher organisms, or RNA in some viruses.

From the genetic material inherited from the parents, the progeny must decode the information within the genetic material which then manifests as various traits that we observe amongst the great diversity of organisms in the living world. This involves information processing and information flow at the most fundamental level throughout the course of the organisms development and lifespan.

Organisms do not exist isolated, but interact with each other constantly within a complex ecosystem. The relationships between organisms such as those between prey or predator, host and parasite, mating partners, or amongst members of a colony of social insects, etc., are both complex and multi-dimensional. In all cases, there is constant communication and infor-

mation flow at many levels. Therefore it is important for us to appreciate that living systems need to compute and process information.

For example, the hottest area in microbiology today is quorum sensing in bacteria. How does a bacterium know how many of its kind is present in its surroundings before it can launch an attack successfully against a host? How does a cell calculate a density gradient and propel its motion towards a food source at the right speed without overshooting the target? How does transcription of DNA know when there is enough messenger RNA, and how does a ribosome know when to stop producing proteins? Where are the feedback loops and how is the regulation tied to the computation or sensing of how much is out there? How do biological systems calculate time - time to live, time to spawn, circadian rhythms and so on? How does the brain process information, count, sense time or store information in memory? How do living systems retain information and transmit it to the next generation? How do living systems share information with others? Thus our book focuses on information processing by life forms and the use of information technology to understand the wonder and fascination of living things.

## **Overview of the Book**

Today, far too many books have been written about bioinformatics. Most of them are highly mathematical and emphasize the underlying mathematical principles and how they may be applied to biological data; or else, they take a superficial but practical approach towards processing and analyzing biological data, covering the so-called “how-to” approach. Few books come close to dealing with computing challenges and computing phenomena in nature, but none combines these with the complementary bioinformatics topics and useful bioinformatics applications. In this book we attempted to do so and we believe that the benefits to the reader will be multiple, going from a simple-to-grasp presentation, to sophisticated and innovative applications.

This book is organised into two major parts, focussing on Biocomputing and Bioinformatics. Both are facets of the information of life - the flow of information in life forms, as well as the use of information technology and computing to understand the mystery of living things.

In the first part, the first two chapters present a comprehensive overview of biocomputing. This constitutes the biocomputing part “Living Computers” that attempts to explain natural processing of biological information using physiological models and analogous models in computing.

The bioinformatics part “Computing Life” - deals with “artificial” processing of biological information as a human endeavour in order to derive new knowledge and insights into life forms and how they function. This part provides overviews of different bioinformatics topics and a glimpse of specific innovative applications for biological discovery as the link and complement to biocomputing.

### **Why are we putting these two domains together?**

Artificial processing of biological information is complementary to natural processing, and by juxtaposing the two, we attempt to enhance our understanding of the natural processing by elucidating and discovering new relevant biological information in a way not commonly done in the literature today.

Our thesis is that a better understanding of the natural processing of biological information, as coming from the biocomputing part, helps us improve the way of processing biological information in deriving new knowledge and insights into life forms and how they function. Consequently, readers will be exposed to complementary domains and will be better equipped to grasp ideas of biocomputing and bioinformatics in tandem when dealing with the biological problems of their interest.

Overall, this book contains a systematic and comprehensive survey of biocomputing not existing in the current literature, and combines it with overviews of different bioinformatics topics complemented by a number of novel bioinformatics applications that illustrate some of the principles of biocomputing.

The book represents a unique source of information on the biological and physiological background against which biological “computing” processes are performed in living systems, including higher cognitive processes. It also shows how some of these computing exemplars in biology have found their way into useful computing applications, many of these applications useful in themselves for dealing with biological information. In particular, we focussed on representative, easy-to-read overviews complemented by a few illustrative applications in dealing with biological information in the bioinformatics realm.

The recent phenomenon of genomics, where large amounts of information stored in the genetic code of living organisms have been elucidated, the accompanying wave of proteomics, metabolomics, transcriptomics, systems biology and other newer ‘omics, urgently calls for a quantum leap of advance in information processing needed for deciphering the meaning of

all this information in living systems. In dealing with this deluge, we hope that you will enjoy this eclectic combination of thoughts found in this book and take time out from the rush of the Internetised world today, to ponder over the intriguing issues raised in this book. Enjoy!

*Vladimir B Bajic and Tin Wee Tan*  
*September 2004*  
*Singapore*

# Contents

*Preface*

v

## **Chapter 1. A Multi-Disciplinary Survey of Biocomputing:**

### **1. Molecular and Cellular Levels**

1

1	Introduction . . . . .	2
2	Lock-Key Paradigm versus Switch-Based Processing . . . . .	3
3	Absolute versus Relative Determinism . . . . .	10
4	Nested Hierarchy of Biocomputing Dynamics . . . . .	13
5	Membrane as a Mesoscopic Substrate . . . . .	16
5.1	Localized and Delocalized Potentials in Biomembranes . . . . .	17
5.2	Role of Membrane Fluidity in the Mesoscopic Dynamics . . . . .	24
5.3	Electrostatic Interactions as a Molecular Switching Mechanism . . . . .	29
5.4	Lateral Mobility of Protons on Membrane Surfaces: the “Pacific Ocean” Effect . . . . .	34
5.5	Role and Specificity of Phospholipid Polar Head- Groups . . . . .	36
5.6	Effect of Transmembrane Diffusion Potentials and Compartmentalization . . . . .	39
5.7	Vesicular Transport, Exocytosis and Synaptic Transmission . . . . .	40
6	Shape-Based Molecular Recognition . . . . .	42
6.1	Role of Short-Range Non-Covalent Bond Interactions in Molecular Recognition . . . . .	43
6.2	Molecular Recognition between Ferredoxin and FNR . . . . .	49
6.3	Comparison of Plastocyanin and Cytochrome $c_6$ . . . . .	51
6.4	Molecular Recognition of Transducin and Arrestin . . . . .	54

	6.5	Electronic-Conformational Interactions . . . . .	60
7		Intracellular and Intramolecular Dynamics . . . . .	61
	7.1	Electrostatic Interactions between a Small Molecule and a Macromolecule . . . . .	62
	7.2	Effect of Phosphorylation . . . . .	64
	7.3	Concept of Intelligent Materials . . . . .	67
	7.4	Concept of Calcium-Concentration Microdomain . . . . .	75
	7.5	Errors, Gradualism and Evolution . . . . .	77
	7.6	Protein Folding . . . . .	80
8		Stochastic Nature of Neural Events: Controlled Randomness of Macroscopic Dynamics . . . . .	88
9		Long-Term Potentiation and Synaptic Plasticity . . . . .	100
10		Role of Dendrites in Information Processing . . . . .	103
11		Efficiency of Biocomputing . . . . .	105
12		General Discussion and Conclusion . . . . .	110
		References . . . . .	115

**Chapter 2. A Multi-Disciplinary Survey of Biocomputing:  
2. Systems and Evolutionary Levels, and  
Technological Applications** 141

1		Introduction . . . . .	142
2		Background . . . . .	147
	2.1	Key Conclusions to Part 1 . . . . .	147
	2.2	Element of Non-Equilibrium Thermodynamics . . . . .	148
	2.3	Element of Cellular Automata . . . . .	149
	2.4	Element of Nonlinear Dynamic Analysis . . . . .	151
3		Biocomputing at the Evolutionary Level . . . . .	153
	3.1	Is Evolution Deterministic? . . . . .	153
	3.2	Explanatory Power of Evolution . . . . .	155
	3.3	Evolution as Problem Solving . . . . .	156
	3.4	Random Search, Exhaustive Search and Heuristic Search . . . . .	157
	3.5	Enigma of Homochirality of Biomolecules . . . . .	158
	3.6	Damage Control and Opportunity Invention . . . . .	160
	3.7	Analogues and Homologues . . . . .	163
	3.8	Co-Evolution and Perpetual Novelty . . . . .	164
	3.9	Punctuated Equilibrium and Cambrian Explosion . . . . .	165
4		Cognitive Aspects of Biocomputing . . . . .	166
	4.1	Models of Creative Problem Solving . . . . .	166

4.1.1	Wallas' Four-Phase Model . . . . .	167
4.1.2	Koestler's Bisociation Model . . . . .	167
4.1.3	Simonton's Chance-Configuration Model . . . . .	168
4.2	Parallel Processing versus Sequential Processing in Pattern Recognition . . . . .	170
4.3	Random Search versus Heuristic Search . . . . .	175
4.4	Dogmatism and Self-Imposed Constraint . . . . .	177
4.5	Retention Phase: The Need of Sequential Verification . . . . .	179
4.6	Picture-Based Reasoning versus Rule-Based Reasoning in Pattern Recognition . . . . .	181
4.7	Advantages and Disadvantages of Rule-Based Reasoning . . . . .	183
4.8	Contemporary Interpretation of Freud's Concept of the Unconscious and Poincaré's Introspective Account . . . . .	191
4.9	Interpretation of Hypnagogia and Serendipity . . . . .	205
4.10	Gray Scale of Understanding and Interpretation of Intuition and "Aha" Experience . . . . .	215
4.11	Pseudo-Parallel Processing . . . . .	226
4.12	Need of Conceptualization and Structured Knowledge . . . . .	229
4.13	Koestler's Bisociation versus Medawar's Hypothetico- Deduction Scheme . . . . .	231
4.14	Behaviorism versus Cognitivism . . . . .	234
4.15	Cerebral Lateralization . . . . .	236
4.16	Innovation versus Imitation: Gray Scale of Creativity . . . . .	241
4.17	Elements of Anticipation and Notion of Planning Ahead . . . . .	244
4.18	Intelligence of Nonhuman Animals: Planning Ahead, Versatility and Language Capability . . . . .	248
4.19	Multiple Intelligences: Role of Working Memory . . . . .	256
4.20	Creativity in Music, Art and Literary Works . . . . .	261
4.21	Complex and Interacting Factors in the Creative Process: Role of Motivation, Hard Work and Intelligence . . . . .	274
4.22	Education and Training: Present Educational Problem . . . . .	284
4.23	Substituted Targets and Goals in Social Engineering . . . . .	297
4.24	Cognitive Development: Nature versus Nurture . . . . .	300
4.25	Is the Crisis in the U.S. Science Education False? . . . . .	306
4.26	Simulations of Gestalt Phenomena in Creativity . . . . .	311
5	Consciousness and Free Will . . . . .	325
5.1	Consciousness . . . . .	326
5.2	Controversy of the Free Will Problem . . . . .	328
5.3	Conflict between Free Will and Classical Determinism . . . . .	330

5.4	One-to-One versus One-to-Many Temporal Mapping . .	332
5.5	Compatibilists versus Incompatibilists . . . . .	335
5.6	Randomness and Determinism in Microscopic Dynamics	338
5.7	Randomness and Determinism in Mesoscopic and Macroscopic Dynamics . . . . .	341
5.8	Endogenous Noise . . . . .	343
5.9	“Controlled” Randomness in a Hierarchical Biocomput- ing System . . . . .	354
5.10	Impossibility of Proving or Disproving the Existence of Free Will . . . . .	355
5.11	Quantum Indeterminacy at the Biological Level . . . .	356
5.12	Microscopic Reversibility and Physical Determinism . .	358
5.13	Incompatibility of Microscopic Reversibility and Macro- scopic Irreversibility . . . . .	363
5.14	Origin of Macroscopic Irreversibility . . . . .	376
5.15	Enigmas of Alternativism, Intelligibility and Origination	386
5.16	Laplace’s “Hidden Cause” Argument . . . . .	394
5.17	Physical Determinism and Cosmology . . . . .	397
5.18	Free Will and Simulations of Consciousness . . . . .	399
5.19	Critique of the New-Mysterian View . . . . .	403
5.20	Readiness Potential and Subjective Feeling of Volition .	413
6	Digression on Philosophy and Sociology of Science . . . . .	419
6.1	Falsifiability and Non-Uniqueness of Scientific Theories	419
6.2	Rise of Postmodernism . . . . .	423
6.3	Gauch’s Analysis . . . . .	424
6.4	Fallibility of Falsification . . . . .	425
6.5	Science of Conjecture . . . . .	427
6.6	Role of Subjectivity in Creative Problem Solving and Value Judgement . . . . .	435
6.7	Critiques of Science Fundamentalism and Post- modernism . . . . .	442
6.8	Level of Confidence in Scientific Knowledge . . . . .	446
6.9	Sociological Aspects of Science . . . . .	447
6.10	Logical Inconsistencies of Antirealism . . . . .	448
6.11	Objective Knowledge: Popper’s Third World . . . . .	449
6.12	Method of Implicit Falsification: Is Psychoanalysis Unscientific? . . . . .	449
6.13	Life Itself: Epistemological Considerations . . . . .	451

6.14	Unity of Knowledge or Great Divide: The Case of Harris versus Edwards . . . . .	475
7	Technological Applications . . . . .	481
7.1	Expert Systems in Artificial Intelligence . . . . .	481
7.2	Neural Network Computing . . . . .	483
7.3	Animat Path to Artificial Intelligence . . . . .	488
7.4	Agent Technology . . . . .	489
7.5	Neuromolecular Brain Model: Multi-Level Neural Network . . . . .	490
7.6	Embryonics: Evolvable Hardware . . . . .	493
7.7	A Successful Example of Molecular Computing: Solving the Direct Hamiltonian Path Problem . . . . .	493
7.8	Prospect of Molecular Electronics in Biocomputing . . . . .	495
8	General Discussion and Conclusion . . . . .	499
	References . . . . .	532

### Chapter 3. Models for Complex Eukaryotic Regulatory DNA Sequences

575

1	Introduction . . . . .	575
2	Some Biology of Transcription Regulation . . . . .	576
2.1	The Basal Transcription Machinery . . . . .	576
2.2	Chromatin Structure in Regulatory Regions . . . . .	579
2.3	Specific Gene Regulation: Sequence Elements and Transcription Factors . . . . .	581
3	Core Promoter Recognition . . . . .	585
3.1	<i>Ab Initio</i> Prediction . . . . .	585
3.2	Alignment Approaches . . . . .	592
4	Prediction of Regulatory Regions by Cross-Species Conservation . . . . .	593
5	Searching for Motif Clusters . . . . .	596
6	Perspective . . . . .	599
	References . . . . .	602

### Chapter 4. An Algorithm for *Ab Initio* DNA Motif Detection

611

1	Introduction . . . . .	611
2	Algorithm . . . . .	612

3	Experiments . . . . .	613
	References . . . . .	614

<b>Chapter 5. Detecting Molecular Evidence of Positive Darwinian Selection</b>		<b>615</b>
1	Introduction . . . . .	615
1.1	Molecular Evolution Research in a Time of Genomes . . . . .	616
1.2	Some Examples . . . . .	616
1.3	Chapter Overview . . . . .	617
2	Types of Adaptive Evolution . . . . .	618
2.1	Episodic Positive Selection . . . . .	618
2.2	Diversifying Selection — The Biological Arms Races . . . . .	618
3	The Neutral Theory of Molecular Evolution . . . . .	619
3.1	Cost of Natural Selection . . . . .	619
3.2	Recent Tests of the Neutral Theory . . . . .	620
3.3	Detecting Departures from Neutrality . . . . .	621
4	Selective Sweeps and Genetic Hitchhiking . . . . .	622
4.1	Detecting Selective Sweeps . . . . .	623
4.2	Correlation between Local Recombination Rates and Diversity . . . . .	624
4.3	Distinguishing Complex Demographic Histories or Background Selection from Positive Selection . . . . .	625
5	Codon-Based Methods to Detect Positive Selection . . . . .	626
5.1	Counting Methods . . . . .	628
5.1.1	Window-Based Methods . . . . .	629
5.1.2	Site-Specific Methods . . . . .	629
5.2	Probabilistic Methods . . . . .	630
5.2.1	Site-Specific Methods . . . . .	631
5.2.2	Lineage-Specific Methods . . . . .	634
5.2.3	Detecting Selection in Non-Coding Regions . . . . .	634
5.3	Comparison of Counting and Probabilistic Approaches to Comparative Methods . . . . .	635
5.4	Codon Volatility . . . . .	636
5.5	Codon-Based Methods that use Polymorphism Data . . . . .	637
6	Discussion and Future Prospects . . . . .	638
	References . . . . .	639

<b>Chapter 6. Molecular Phylogenetic Analysis: Understanding Genome Evolution</b>	<b>645</b>
1 What is Phylogenetics? . . . . .	645
2 What is a Phylogenetic Tree? . . . . .	646
3 Identifying Duplicate Genes . . . . .	646
3.1 Generate Protein Families . . . . .	647
3.2 Multiple Sequence Alignments . . . . .	647
3.3 Reconstructing Phylogenetic Trees . . . . .	647
4 Assessing the Accuracy of Phylogenetic Trees . . . . .	649
5 High-Throughput Screening of Tree Topologies . . . . .	649
6 Concluding Remarks . . . . .	649
References . . . . .	650
<b>Chapter 7. Constructing Biological Networks of Protein-Protein Interactions</b>	<b>653</b>
1 Introduction . . . . .	653
2 Bioinformatic Approaches . . . . .	654
2.1 Homology . . . . .	655
2.2 Fusion Events . . . . .	656
2.3 Co-Localization . . . . .	657
2.4 Co-Evolution . . . . .	659
2.5 Literature Mining . . . . .	660
3 From Interactions to Networks . . . . .	664
3.1 False Negatives . . . . .	665
3.2 False Positives . . . . .	666
4 Conclusion . . . . .	668
References . . . . .	669
<b>Chapter 8. Computational Modelling of Gene Regulatory Networks</b>	<b>673</b>
1 Introduction . . . . .	673
2 A Novel Approach . . . . .	674
3 Modelling Application with Integrated Approach of First- Order Differential Equations, State Space Representation and Kalman Filter . . . . .	676
3.1 Discrete-Time Approximation of First-Order Differen- tial Equations . . . . .	676

3.2	State Space Representation . . . . .	677
3.3	Kalman Filter . . . . .	678
3.4	Using GA for the Selection of Gene Subset for a GRN . . . . .	678
3.5	GA Design for Gene Subset Selection . . . . .	679
3.6	Procedure of the GA-Based Method for Gene Subset Selection . . . . .	679
4	Experiments and Results . . . . .	681
4.1	Building a Global GRN of the Whole Gene Set out of the GRNs of Smaller Number of Genes (Putting the Pieces of the Puzzle Together) . . . . .	683
5	Conclusions . . . . .	684
	References . . . . .	685

## **Chapter 9. Overview of Text-Mining in Life-Sciences** . . . . . 687

1	Introduction . . . . .	687
2	Overview of Text-Mining . . . . .	689
3	Scope and Nature of Text-Mining in Life-Sciences Domain . . . . .	689
3.1	Characteristics of Text-Mining Systems . . . . .	690
3.2	Systems Aimed at Life-Sciences Applications . . . . .	690
4	Conclusions . . . . .	692
	References . . . . .	692

## **Chapter 10. Integrated Prognostic Profiles: Combining Clinical and Gene Expression Information through Evolving Connectionist Approach** . . . . . 695

1	Introduction . . . . .	696
2	Methods . . . . .	697
2.1	Data sets . . . . .	698
2.2	Data Integration . . . . .	698
2.3	Common Feature Set Selection . . . . .	698
2.4	Algorithm of Integrated Feature Selection . . . . .	699
2.5	Experimental Design . . . . .	701
3	Results . . . . .	701
3.1	Classification Accuracy Test and Profile Verification . . . . .	702
4	Discussion . . . . .	704
4.1	Discovering Genotype Phenotype Relationships through Integrated Profiles . . . . .	704

5	Conclusion . . . . .	705
	References . . . . .	706
<b>Chapter 11. Databases on Gene Regulation</b>		<b>709</b>
1	Introduction . . . . .	709
2	Brief Overview of Common Databases Presenting General Information on Genes and Proteins . . . . .	711
3	Specialized Databases on Transcription Regulation . . . . .	712
	3.1 TRANSFAC <sup>®</sup> . . . . .	713
	3.2 TRANSCompel <sup>®</sup> — A Database on Composite Regula- tory Elements . . . . .	715
	3.3 SMART DB — A Database on Scaffold/Matrix Attached Regions . . . . .	715
4	Databases on Biomolecular Interactions and Signaling Networks . . . . .	716
	4.1 Regulatory Networks: General Properties and Peculiarities . . . . .	716
	4.2 Variety of Databases on Protein Interactions and Signaling Networks . . . . .	717
	4.3 TRANSPATH <sup>®</sup> — A Database on Signal Transduction Pathways . . . . .	719
5	Application of the Databases for Causal Interpretation of Gene Expression Data . . . . .	721
	5.1 Analysis of Promoters . . . . .	721
	5.2 Identification of Key Nodes in Signaling Networks . . . . .	723
	References . . . . .	724
<b>Chapter 12. On the Search of Better Validation and Statistical Methods in Microarray Data Analysis</b>		<b>729</b>
1	Introduction . . . . .	729
2	Microarray Analysis Steps . . . . .	730
3	Preprocessing . . . . .	730
4	Normalization . . . . .	732
5	Identification of Differentially Expressed Genes . . . . .	733
6	Validation Strategies . . . . .	734
7	Experimental Validation Methods . . . . .	736
	7.1 Self-Hybridization or Identical Replicates . . . . .	736

7.2	Quantitative RT-PCR . . . . .	736
7.3	Mutant versus Wild Type . . . . .	737
7.4	Gene Spike-In Experiments . . . . .	738
7.5	Other Validation Experiments . . . . .	739
8	Summary . . . . .	739
	References . . . . .	740

### **Chapter 13. Information Extraction from Dynamic Biological Web Sources**

		741
1	Introduction . . . . .	741
2	Information Extraction from Dynamic Web Sources . . . . .	743
3	Survey of Wrapper Maintenance Systems . . . . .	745
3.1	Wrapper Verification Methods . . . . .	745
3.1.1	RAPTURE . . . . .	745
3.1.2	Forward-Backward Scanning Algorithm . . . . .	745
3.2	Wrapper Reinduction Methods . . . . .	746
3.2.1	ROADRUNNER . . . . .	746
3.2.2	DataProg . . . . .	747
3.2.3	Schema-Guided Wrapper Maintenance (SG-WRAM) . . . . .	747
3.2.4	ReInduce Algorithm . . . . .	748
4	Conclusion . . . . .	748
	References . . . . .	749

### **Chapter 14. Computer Aided Design of Signaling Networks**

1	Introduction . . . . .	751
2	Signaling Pathways: A Prickly Proposition . . . . .	751
3	Challenges of Signaling Modeling . . . . .	754
4	The Goals and Features of Cellware . . . . .	756
5	Concluding Remarks . . . . .	757
	References . . . . .	759

### **Chapter 15. Analysis of DNA Sequences: Hunting for Genes**

		761
1	Introduction . . . . .	761
2	DNA and Genes . . . . .	762
2.1	DNA . . . . .	762
2.2	Coding Genes . . . . .	762

2.3	The Genetic Code and Proteins . . . . .	764
2.4	Structure of Coding Genes . . . . .	764
2.5	Complementary DNA (cDNA) . . . . .	765
2.6	Non-Coding Genes . . . . .	765
3	Genomes . . . . .	766
3.1	Computational Analysis of the Genome: Coding Gene Prediction . . . . .	767
3.2	Computational Analysis of the Genome: Non-Coding Gene Prediction . . . . .	768
4	Closing Remarks . . . . .	769
	References . . . . .	769

**Chapter 16. Biological Databases and Web Services:**

**Metrics for Quality**

771

1	Introduction . . . . .	771
2	Growing Need for Quality Control . . . . .	772
3	Metrics for Quality Analysis . . . . .	773
3.1	Content . . . . .	773
3.2	Availability . . . . .	774
3.3	Combining Different Metrics . . . . .	775
4	Discussion and Conclusion . . . . .	776
	References . . . . .	776