
ELEMENTS OF COMPUTATIONAL SYSTEMS BIOLOGY

Edited by

Huma M. Lodhi
Stephen H. Muggleton

Imperial College London, London, UK



WILEY

A John Wiley & Sons, Inc., Publication

CONTENTS

PREFACE	xv
CONTRIBUTORS	xvii

PART I OVERVIEW

1 Advances in Computational Systems Biology	3
<i>Huma M. Lodhi</i>	
1.1 Introduction / 3	
1.2 Multiscale Computational Modeling / 4	
1.3 Proteomics / 7	
1.4 Computational Systems Biology and Aging / 8	
1.5 Computational Systems Biology in Drug Design / 9	
1.6 Software Tools for Systems Biology / 11	
1.7 Conclusion / 13	
References / 13	

PART II BIOLOGICAL NETWORK MODELING

2 Models in Systems Biology: The Parameter Problem and the Meanings of Robustness	21
<i>Jeremy Gunawardena</i>	
2.1 Introduction / 21	
2.2 Models as Dynamical Systems / 23	
2.2.1 Continuous Models / 24	
2.2.2 Discrete Models / 25	
2.3 The Parameter Problem / 26	
2.3.1 Parameterphobia / 27	
2.3.2 Measuring and Calculating / 27	

- 2.3.3 Counter Fitting / 28
- 2.3.4 Beyond Fitting / 29
- 2.4 The Landscapes of Dynamics / 29
 - 2.4.1 Qualitative Dynamics / 29
 - 2.4.2 Steady State Attractors of ODE Models / 33
- 2.5 The Meanings of Robustness / 35
 - 2.5.1 Parameter Biology / 35
 - 2.5.2 Robustness to Initial Conditions / 37
 - 2.5.3 Robustness in Reality / 38
 - 2.5.4 Structural Stability / 40
 - 2.5.5 Classifying Robustness / 42
- 2.6 Conclusion / 42
- References / 43

3 *In Silico* Analysis of Combined Therapeutics Strategy for Heart Failure

49

Sung-Young Shin, Tae-Hwan Kim, Kwang-Hyun Cho, and Sang-Mok Choo

- 3.1 Introduction / 49
- 3.2 Materials and Methods / 50
 - 3.2.1 Model Construction and Validation / 50
 - 3.2.2 Classification of Different Heart Failure Cases / 51
 - 3.2.3 Simulation Protocol / 54
- 3.3 Results / 54
 - 3.3.1 β -Adrenergic Receptor Antagonists / 54
 - 3.3.2 β -Adrenergic Receptor Kinase Inhibitor / 56
 - 3.3.3 Phosphodiesterase Inhibitor / 57
 - 3.3.4 Combined Therapies / 58
- 3.4 Discussion / 61
- Acknowledgment / 61
- 3A.1 Appendix / 62
 - 3A.1.1 Model Validation / 62
 - 3A.1.2 The Mathematical Model Used for Simulations / 64
- References / 80

4 Rule-Based Modeling and Model Refinement

83

Elaine Murphy, Vincent Danos, Jérôme F  ret, Jean Krivine, and Russell Harmer

- 4.1 Kappa, Briefly / 84
- 4.2 Refinement, Practically / 85
 - 4.2.1 A Simple Cascade / 85
 - 4.2.2 Another Cascade / 89
 - 4.2.3 The SSA Convention / 94
 - 4.2.4 A Less Obvious Refinement / 96

4.3	Rule-Based Modeling / 98	
4.3.1	Notation / 99	
4.3.2	Objects and Arrows / 99	
4.3.3	Extensions / 101	
4.3.4	Actions and Rules / 101	
4.3.5	Events and Probabilities / 102	
4.4	Refinement, Theoretically / 103	
4.4.1	Growth Policies / 103	
4.4.2	Simple Growth Policies / 104	
4.4.3	Neutral Refinements / 105	
4.4.4	Example Concluded / 107	
4.4.5	Growth Policies, Concretely / 107	
4.4.6	A Weakly Homogeneous Refinement / 109	
4.4.7	Nonhomogeneous Growth Policies / 111	
4.5	Conclusion / 113	
	References / 114	
5	A (Natural) Computing Perspective on Cellular Processes	115
	<i>Matteo Cavaliere and Tommaso Mazza</i>	
5.1	Natural Computing and Computational Biology / 115	
5.2	Membrane Computing / 116	
5.3	Formal Languages Preliminaries / 118	
5.4	Membrane Operations with Peripheral Proteins / 119	
5.5	Membrane Systems with Peripheral Proteins / 122	
5.5.1	Dynamics of the System / 122	
5.5.2	Reachability in Membrane Systems / 124	
5.6	Cell Cycle and Breast Tumor Growth Control / 126	
5.6.1	Cell Cycle Progression Inhibition in G1/S / 127	
5.6.2	Cell-Cycle Progression Inhibition in G2/M / 132	
	References / 138	
6	Simulating Filament Dynamics in Cellular Systems	141
	<i>Wilbur E. Channels and Pablo A. Iglesias</i>	
6.1	Introduction / 141	
6.2	Background: The Roles of Filaments within Cells / 142	
6.2.1	The Actin Network / 142	
6.2.2	Intermediate Filaments / 143	
6.2.3	Microtubules / 144	
6.3	Examples of Filament Simulations / 145	
6.3.1	Actin-Based Motility in <i>Listeria</i> / 145	
6.3.2	Kinetochores Positioning in Budding Yeast / 146	
6.3.3	Spindle Positioning in <i>Caenorhabditis Elegans</i> Embryos / 146	
6.3.4	Other Examples / 147	

- 6.4 Overview of Filament Simulation / 147
- 6.5 Changing Filament Length / 149
 - 6.5.1 Resegmenting Filament / 149
- 6.6 Forces on Filaments / 151
 - 6.6.1 Brownian Forces / 151
 - 6.6.2 Straightening Force / 152
 - 6.6.3 Forces from Motor Complexes / 154
- 6.7 Imposing Constraints / 154
 - 6.7.1 Motivation / 154
 - 6.7.2 Derivation / 155
 - 6.7.3 Implementation / 156
 - 6.7.4 State Equation / 157
- 6.8 Solver / 158
- 6.9 Conclusion / 159
- References / 160

PART III BIOLOGICAL NETWORK INFERENCE

7 Reconstruction of Biological Networks by Supervised Machine Learning Approaches **165**

Jean-Philippe Vert

- 7.1 Introduction / 165
- 7.2 Graph Reconstruction as a Pattern Recognition Problem / 168
 - 7.2.1 Problem Formalization / 168
 - 7.2.2 Pattern Recognition / 169
 - 7.2.3 Graph Inference as a Pattern Recognition Problem / 171
 - 7.2.4 Graph Inference with Local Models / 171
 - 7.2.5 Graph Inference with Global Models / 173
 - 7.2.6 Remarks / 178
- 7.3 Examples / 181
 - 7.3.1 Reconstruction of a Metabolic Network / 181
 - 7.3.2 Reconstruction of a PPI Network / 182
 - 7.3.3 Reconstruction of Gene Regulatory Networks / 183
- 7.4 Discussion / 185
- References / 186

8 Supervised Inference of Metabolic Networks from the Integration of Genomic Data and Chemical Information **189**

Yoshihiro Yamanishi

- 8.1 Introduction / 189
- 8.2 Materials / 192
 - 8.2.1 Metabolic Network / 192
 - 8.2.2 Genomic Data / 192

- 8.2.3 Chemical Information / 193
- 8.2.4 Kernel Representation / 193
- 8.3 Supervised Network Inference with Metric Learning / 194
 - 8.3.1 Formalism of the Problem / 194
 - 8.3.2 From Metric Learning to Graph Inference / 195
- 8.4 Algorithms for Supervised Network Inference / 196
 - 8.4.1 Kernel Canonical Correlation Analysis (KCCA) / 196
 - 8.4.2 Distance Metric Learning (DML) / 197
 - 8.4.3 Kernel Matrix Regression (KMR) / 199
 - 8.4.4 Penalized Kernel Matrix Regression (PKMR) / 200
 - 8.4.5 Relationship with Kernel Matrix Completion and *em*-algorithm / 201
- 8.5 Data Integration / 203
 - 8.5.1 Genomic Data Integration / 203
 - 8.5.2 Chemical Compatibility Network / 203
 - 8.5.3 Incorporating Chemical Constraint / 204
- 8.6 Experiments / 204
- 8.7 Discussion and Conclusion / 207
- References / 209

9 Integrating Abduction and Induction in Biological Inference Using CF-Induction 213

Yoshitaka Yamamoto, Katsumi Inoue, and Andrei Doncescu

- 9.1 Introduction / 213
- 9.2 Logical Modeling of Metabolic Flux Dynamics / 215
 - 9.2.1 Metabolic Pathways / 215
 - 9.2.2 Regulation of Enzymatic Activities / 217
- 9.3 CF-induction / 218
 - 9.3.1 Inductive Logic Programming / 218
 - 9.3.2 Abduction and Induction in CF-induction / 221
- 9.4 Experiments / 224
 - 9.4.1 A Simple Pathway / 224
 - 9.4.2 A Metabolic Pathway of Pyruvate / 227
- 9.5 Related Work / 230
- 9.6 Conclusion and Future Work / 232
- Acknowledgments / 232
- References / 233

10 Analysis and Control of Deterministic and Probabilistic Boolean Networks 235

Tatsuya Akutsu and Wai-Ki Ching

- 10.1 Introduction / 235
- 10.2 Boolean Network / 236

- 10.3 Identification of Attractors / 238
 - 10.3.1 Definition of BN-ATTRACTOR / 238
 - 10.3.2 Basic Recursive Algorithm / 239
 - 10.3.3 On the Worst Case Time Complexity of BN-ATTRACTOR / 241
- 10.4 Control of Boolean Network / 241
 - 10.4.1 Definition of BN-CONTROL / 242
 - 10.4.2 Dynamic Programming Algorithms for BN-CONTROL / 243
 - 10.4.3 NP-hardness Results on BN-CONTROL / 244
- 10.5 Probabilistic Boolean Network / 246
- 10.6 Computation of Steady States of PBN / 248
 - 10.6.1 Exact Computation of PBN-STEADY / 248
 - 10.6.2 Approximate Computation of PBN-STEADY / 249
- 10.7 Control of Probabilistic Boolean Networks / 250
 - 10.7.1 Dynamic Programming Algorithm for PBN-CONTROL / 250
 - 10.7.2 Variants of PBN-CONTROL / 251
- 10.8 Conclusion / 253
- Acknowledgments / 254
- References / 254

11 Probabilistic Methods and Rate Heterogeneity

257

Tal Pupko and Itay Mayrose

- 11.1 Introduction to Probabilistic Methods / 257
- 11.2 Sequence Evolution is Described Using Markov Chains / 258
 - 11.2.1 Estimating Pairwise Distances / 260
 - 11.2.2 Calculating the Likelihood of a Tree / 261
 - 11.2.3 Extending the Basic Model / 262
- 11.3 Among-site Rate Variation / 263
- 11.4 Distribution of Rates Across Sites / 266
 - 11.4.1 The Gamma Distribution / 266
 - 11.4.2 Numerical Approximation of the Continuous Gamma Distribution / 268
 - 11.4.3 Alternative Rate Distributions / 269
- 11.5 Site-specific Rate Estimation / 271
- 11.6 Tree Reconstruction Using Among-site Rate Variation Models / 272
- 11.7 Dependencies of Evolutionary Rates Among Sites / 274
- 11.8 Related Works / 275
- References / 276

PART IV GENOMICS AND COMPUTATIONAL SYSTEMS BIOLOGY

- 12 From DNA Motifs to Gene Networks: A Review of Physical Interaction Models** **283**
Panayiotis V. Benos and Alain B. Tchagang
- 12.1 Introduction / 283
 - 12.2 Fundamentals of Gene Transcription / 286
 - 12.2.1 Physical Basis of Transcription Regulation and Representation of DNA Patterns / 286
 - 12.2.2 High-Throughput Data: Microarrays, Deep Sequencing, ChIP-chip, and ChIP-seq / 287
 - 12.3 Physical Interaction Algorithms / 289
 - 12.3.1 Basic Definitions / 289
 - 12.3.2 Problem Formulation / 289
 - 12.3.3 Clustering-Based Approaches / 290
 - 12.3.4 Sequence- or ChIP-Based Regression Methods / 293
 - 12.3.5 Network Component Analysis Methods / 296
 - 12.3.6 Factor Analysis Methods / 299
 - 12.4 Conclusion / 301
 - 12.4.1 Future Prospects and Challenges / 304
 - Acknowledgments / 304
 - References / 305
- 13 The Impact of Whole Genome *In Silico* Screening for Nuclear Receptor-Binding Sites in Systems Biology** **309**
Carsten Carlberg and Merja Heinäniemi
- 13.1 Introduction / 309
 - 13.2 Nuclear Receptors / 310
 - 13.2.1 NRs as a Link between Nutrition Sensing and Inflammation Prevention / 311
 - 13.2.2 NRs and System Biology / 312
 - 13.3 The PPAR Subfamily / 313
 - 13.3.1 Global Datasets that Identify a Central Role for PPARs in Disease Progression / 314
 - 13.3.2 PPAR Response Elements / 315
 - 13.4 Methods for *in Silico* Screening of Transcription Factor-Binding Sites / 315
 - 13.5 Binding Dataset of PPREs and the Classifier Method / 317
 - 13.6 Clustering of Known PPAR Target Genes / 318
 - 13.6.1 A Look at PPREs in their Genomic Context: Putative Target Genes and Binding Modules / 319

- 13.7 Conclusion / 319
- Acknowledgments / 320
- References / 321

14 Environmental and Physiological Insights from Microbial Genome Sequences **325**

Alessandra Carbone and Anthony Mathelier

- 14.1 Some Background, Motivation, and Open Questions / 325
- 14.2 A First Statistical Glimpse to Genomic Sequences / 328
- 14.3 An Automatic Detection of Codon Bias in Genes / 329
- 14.4 Genomic Signatures and a Space of Genomes for Genome Comparison / 330
- 14.5 Study of Metabolic Networks Through Sequence Analysis and Transcriptomic Data / 331
- 14.6 From Genome Sequences to Genome Synthesis: Minimal Gene Sets and Essential Genes / 332
- 14.7 A Chromosomal Organization of Essential Genes / 333
- 14.8 Viral Adaptation to Microbial Hosts and Viral Essential Genes / 334
- 14A.1 Appendix / 335
 - 14A.1.1 Some Comments on the Mathematical Methods / 335
 - 14A.1.2 Complete Genomes Available / 336
- References / 336

PART V SOFTWARE TOOLS FOR SYSTEMS BIOLOGY

15 ALI BABA: A Text Mining Tool for Systems Biology **343**

Jörg Hakenberg, Conrad Plake, and Ulf Leser

- 15.1 Introduction to Text Mining / 343
- 15.2 ALI BABA as a Tool for Mining Biological Facts from Literature / 346
 - 15.2.1 Which Proteins are Involved in the Activation or Recognition Mechanism for PmrD? / 347
 - 15.2.2 What are the Risk Factors of Treating G6PD-Deficient Malaria Patients with Primaquine? / 347
 - 15.2.3 The Wnt Signaling Pathway / 349
- 15.3 Components and usage of ALI BABA / 349
 - 15.3.1 Views / 350
 - 15.3.2 Navigation / 351
 - 15.3.3 Filters / 352
 - 15.3.4 Storage and Retrieval / 353
 - 15.3.5 Graph Editor / 353

15.4	ALI BABA'S Approach to Text Mining / 354	
15.4.1	Natural Language Processing / 354	
15.4.2	Named Entity Recognition and Identification / 355	
15.4.3	Word Sense Disambiguation / 356	
15.4.4	Relation Mining / 357	
15.4.5	Evaluation of Components / 359	
15.5	Related Biomedical Text Mining Tools / 361	
15.5.1	Document Retrieval and Summarization / 361	
15.5.2	Entity Extraction / 361	
15.5.3	Relation Mining / 362	
15.6	Conclusions and Future Perspectives / 362	
	Acknowledgments / 364	
	References / 364	
16	Validation Issues in Regulatory Module Discovery	369
	<i>Alok Mishra and Duncan Gillies</i>	
16.1	Introduction / 369	
16.2	Data Types / 370	
16.3	Data Integration / 371	
16.4	Validation Approaches / 374	
16.4.1	Internal Validation Methods / 374	
16.4.2	Measures of Biological Significance / 376	
16.5	Conclusions / 378	
	References / 378	
17	Computational Imaging and Modeling for Systems Biology	381
	<i>Ling-Yun Wu, Xiaobo Zhou, and Stephen T.C. Wong</i>	
17.1	Bioinformatics / 383	
17.1.1	Gene Microarray / 383	
17.1.2	Mass Spectrometry / 383	
17.1.3	Molecular Networks and Pathways / 384	
17.2	Bioimage Informatics of High-Content Screening / 385	
17.2.1	Image Pre-processing / 386	
17.2.2	Cell Detection, Segmentation, and Centerline Extraction / 386	
17.2.3	Cell Tracking and Registration / 388	
17.2.4	Feature Extraction / 388	
17.3	Connecting Bioinformatics and Biomedical Imaging / 389	
17.3.1	Cellular Networks Analysis by Using HCS / 389	
17.3.2	Gene Function Annotation by Using HCS / 390	
17.3.3	Association Studies with Clinical Imaging Traits / 390	

xiv CONTENTS

- 17.3.4 *In Vivo* Genomics Analysis / 391
- 17.3.5 *In Vivo* Proteomics Analysis / 392
- 17.3.6 *In Vivo* Genetic Analysis / 393
- 17.3.7 *In Vivo* RNAi Experiments / 393

- 17.4 Summary / 394
- Acknowledgments / 394
- References / 394

INDEX 403

SERIES INFORMATION 413