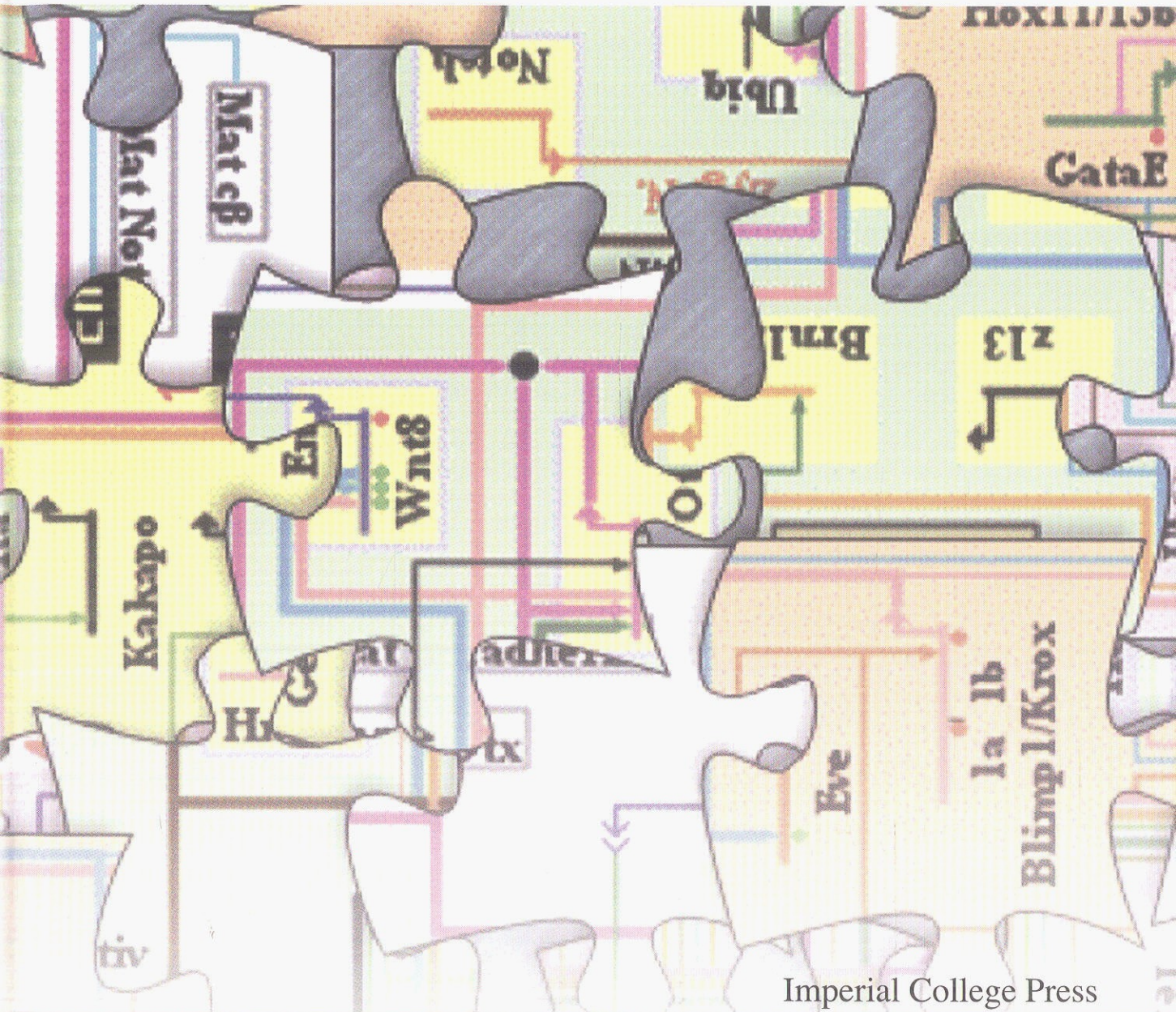


Computational Modeling of Gene Regulatory Networks

— a Primer

Hamid Bolouri



Imperial College Press

Contents

1. Introduction	1
The increasing role of computational analysis in biology	1
What this book tries to achieve	3
Who should read this book	4
How this book is organized	6
Acknowledgments	7
Feedback	7
2. What Is a System, and Why Should We Care?	9
Linearity versus nonlinearity	9
Nonlinear systems	13
Nonlinear systems are the norm, not the exception, in biology	14
3. What Models Can and Cannot Predict	17
Interpolation versus extrapolation	17
Iterative model refinement by experimental falsification of model extrapolations	21
The importance of remembering the limitations of data	22
Cross-validation	23
Function approximation versus classification	25
Appendix: A model of biphasic kinetics	26

4. Why Make Computational Models of Gene Regulatory Networks?	29
What is a model?	29
What is the goal of GRN modeling?	31
Why make computational models of GRNs?	32
Serendipitous benefits of computational GRN modeling	33
Some pitfalls of modeling	34
Good practice guidelines	35
Appendix: Working definitions of ‘genes’ and ‘Gene Regulatory Networks’	36
5. Graphical Representations of Gene Regulatory Networks	39
Desirable features of computational GRN representations	39
Graphical representation of GRN activity in multiple compartments	43
Computational network building, editing, and topological analysis	46
6. Implicit Modeling via Interaction Network Maps	49
Data interpretation through implicit modeling	49
Global molecular interaction maps — Guilt by association	50
Why do we need <i>global</i> molecular interaction maps?	53
Example uses of interaction maps as predictive models	54
7. The Biochemical Basis of Gene Regulation	61
The probability of a chemical reaction	61
<i>A simple method for modeling stochastic molecular reaction events</i>	63
Chemical kinetics in cells are different from <i>in vitro</i> kinetics	65
Compared to transcription, most signaling events are instantaneous	66
How transcription factors find their targets on DNA	67
DNA bending and looping by transcription factors	70
Spatial localization: multi-compartment modeling	71

Morphogen gradients	72
Appendix: Stochastic simulation using Gillespie's algorithm	73
8. A Single-Cell Model of Transcriptional Regulation	77
Modeling strategy	77
Modeling framework and notation	78
A single-cell stochastic model of transcriptional regulation	79
Recruitment of RNA polymerase II complex and transcription initiation	82
Appendix: Simulation of the distribution of gene expression levels in a population of genetically identical cells	89
9. Simplified Models: Mass-Action Kinetics	99
Why model with mass-action kinetics?	99
The fundamentals of Ordinary Differential Equations (ODEs)	100
Steady states	103
Average promoter occupancy by a single transcription factor	104
<i>Promoter occupancy by two or more factors</i>	105
A two-step kinetic model of mRNA and protein concentration	107
mRNA and protein levels at steady state	109
Promoter occupancy as a function of regulator concentration	109
Analytical solution of mRNA and protein time-course kinetics for genes regulated by posttranscriptionally activated factors	110
The time-course behavior of genes regulated by other genes	112
The Boolean approximation to transcription kinetics	114
In the absence of feedback, transcription factors in animals do not reach steady state	115
Positive and negative feedback loops can drive gene expression to fixed steady-state levels	117
Gene expression as a function of DNA-bound regulator activity	117

Appendix A: ODE modeling with Berkeley Madonna	119
Appendix B: Derivation of mathematical expressions for mRNA and protein levels as a function of changing occupancy levels	120
Appendix C: Time to steady state for genes not regulated by feedback	122
10. Simplified Models: Boolean and Multi-valued Logic	123
Background	123
Discrete-variable piecewise linear ODEs	125
Multi-valued logic networks	129
Implicit-time logic networks (a.k.a. kinetic logic)	132
Learning discrete logic models directly from data	135
Linear ODE models of transcriptional regulation	136
Process algebras	139
Appendix: Logic simulation model files	140
11. Simplified Models: Bayesian Networks	143
A preview	145
Probabilities: A brief review	146
Continuous and discrete probability distributions	148
The theoretical foundation of BNs: Conditional probabilities	149
Making predictions with a given BN	151
Modeling networks with feedback as Dynamic Bayesian Networks	154
Constructing BNs directly from data	156
<i>Causality in BNs</i>	161
Computational efficiency in BNs	162
Current limitations of Bayesian Networks	163
Resources for BNs	164
Appendix: Exploring BNs with Hugin	165

12. The Relationship between Logic and Bayesian Networks	167
Noisy logic networks	167
Probabilistic Boolean Networks	169
Learning PBNs from data	171
Some useful properties of PBNs	172
13. Network Inference in Practice	175
A summary of the general approach to network reconstruction	175
Learning logic models from gene expression data alone	178
Learning continuous-valued network models from expression data	182
Network structure building by data integration	184
14. Searching DNA Sequences for Transcription Factor Binding Sites	189
Consensus sequences	189
Position Weight Matrices	191
Visualizing PWMs with sequence logos	194
<i>A taxonomy of TFBS prediction algorithms</i>	196
Resources for TFBS prediction	201
Some good practice guidelines	202
Measuring the performance of binding site prediction algorithms	204
Extracting predicted TFBSs from ChIP-chip data	206
Appendix: DNA sequence processing	211
15. Model Selection Theory	213
Fitting error versus generalization error	213
Model misspecification	214
Model invalidation	215

Model selection criteria	216
How to calculate the log-likelihood value for a regression model	219
Parameter counts of common modeling frameworks	221
The effect of function complexity	222
Multi-model averaging	223
Other approaches to model refinement	224
16. Simplified Models — GRN State Signatures in Data	225
Principal Component Analysis	226
Nonlinear PCA	232
Multi-dimensional Scaling (MDS)	235
Partial Least Squares (PLS)	237
The implicit approach to pattern detection in complex data	237
Appendix: Step-by-step example PCA transformations	239
17. System Dynamics	243
Transients and steady states	243
Phase portraits	245
Parameter analysis	249
Parameter optimization and the evolution of optimal dynamics	252
Bistability through mutual inhibition	254
Negative auto-regulation	255
Mixed positive and negative feedback	258
Appendix: Analyzing feedback dynamics	260
18. Robustness Analysis	265
Robustness and sensitivity	265
Perturbations in system state variables versus perturbations in system parameters	266

Failure tolerance versus graceful degradation	266
Global and local perspectives	268
Local sensitivity analysis	268
Global sensitivity analysis	270
The role of network topology in robustness	273
Evolution of robustness	275
Robustness to transcriptional noise	277
Context and completeness of models	277
19. GRN Modules and Building Blocks	279
Hierarchical modularity in engineered systems	279
Organizational principles in GRNs	281
Network motifs in GRNs	283
Functional building blocks	288
Using network motifs and functional building blocks to decode GRNs	290
20. Notes on Data Processing for GRN Modeling	293
What type of data is best for modeling?	293
Beware of the side-effects of the methods used to collect data	294
How many time points are sufficient for modeling dynamics?	295
<i>In vivo</i> versus <i>ex vivo</i> and <i>in vitro</i> data	296
Using meaningful units to quantify data	297
Misinterpreting data	297
21. Applications of Computational GRN Modeling	299
Overview	299
GRN modeling challenges in medical systems biology	301
Modeling hierarchical, distributed processing in the immune system	305

22. Quo Vadis	311
The US\$1000 genome and its challenges	311
Single-cell biology	313
Multi-scale modeling	315
Software engineering challenges	316
Becoming bilingual	318
Molecular biology is still in the discovery phase	319
Index	321