Biomolecular Feedback Systems

Domitilla Del Vecchio MIT

Richard M. Murray Caltech

Classroom Copy v0.6d, July 20, 2013 © California Institute of Technology All rights reserved.

This manuscript is for review purposes only and may not be reproduced, in whole or in part, without written consent from the authors.

Contents

C	Contents			
P	refac	:e	iii	
N	otati	on	v	
1	Int	troductory Concepts	1	
	1.1	Systems Biology: Modeling, Analysis and Role of Feedback	1	
	1.2	The Cell as a System	8	
	1.3	Control and Dynamical Systems Tools	10	
	1.4	Input/Output Modeling	17	
	1.5	From Systems to Synthetic Biology	21	
	1.6	Further Reading	27	
2	Dy	vnamic Modeling of Core Processes	29	
	2.1	Modeling Techniques	29	
	2.2	Transcription and Translation	43	
	2.3	Transcriptional Regulation	53	
	2.4	Post-Transcriptional Regulation	68	
	2.5	Cellular Subsystems	79	
		Exercises	86	
3	Ar	nalysis of Dynamic Behavior	89	
	3.1	Analysis Near Equilibria	89	
	3.2	Robustness	102	
	3.3	Analysis of Reaction Rate Equations	113	
	3.4	Oscillatory Behavior	119	
	3.5	Bifurcations	129	
	3.6	Model Reduction Techniques	133	
		Exercises	139	
4	St	ochastic Modeling and Analysis	149	
•	4.1	Stochastic Modeling of Biochemical Systems	149	
		Simulation of Stochastic Systems	164	

i	CONTENTS

	4.3	Input/Output Linear Stochastic Systems	167 173
5	Fee	edback Examples	177
	5.1	The lac Operon	179
	5.2	Bacterial Chemotaxis	185
6	Bio	ological Circuit Components	197
	6.1	Introduction to Biological Circuit Design	197
	6.2	Negative Autoregulation	200
	6.3	The Toggle Switch	206
	6.4	The Repressilator	207
	6.5	Activator-Repressor Clock	211
	6.6	An Incoherent Feedforward Loop (IFFL)	215
		Exercises	218
7	Int	erconnecting Components	221
•	7.1	Input/Output Modeling and the Modularity Assumption	221
	7.2	Introduction to Retroactivity	222
	7.3	Retroactivity in Gene Circuits	225
	7.4	Retroactivity in Signaling Systems	230
	7.5	Insulation Devices: Retroactivity Attenuation	235
	7.5	Exercises	251
_	_		
8		sign Tradeoffs	255
	8.1	Competition for Shared Cellular Resources	255
	8.2	Stochastic Effects: Design Tradeoffs in Systems with Large Gains .	263
		Exercises	268
A	Al	Primer on Control Theory	269
	A.1	System Modeling	269
	A.2	Dynamic Behavior	270
	A.3	Linear Systems	272
	A.4	Reachability and observability	274
	A.5	Transfer Functions	276
	A.6	Frequency Domain Analysis	278
	A.7	PID Control	280
		Limits of Performance	
	A.9	Robust Performance	282
В	iblio	graphy	285
T,	ıdex		292
11	IUCX		474

Preface

This text is intended for researchers interested in the application of feedback and control to biomolecular systems. The material has been designed so that it can be used in parallel with the textbook *Feedback Systems* [1] as part of a course on biomolecular feedback and control systems, or as a standalone reference for readers who have had a basic course in feedback and control theory. The full text for this book, along with additional supplemental material, is available on a companion web site:

http://www.cds.caltech.edu/~murray/BFS

The material in this book is intended to be useful to three overlapping audiences: graduate students in biology and bioengineering interested in understanding the role of feedback in natural and engineered biomolecular systems; advanced undergraduates and graduate students in engineering disciplines who are interested the use of feedback in biological circuit design; and established researchers in the biological sciences who want to explore the potential application of principles and tools from control theory to biomolecular systems. We have written the text assuming some familiarity with basic concepts in feedback and control, but have tried to provide insights and specific results as needed, so that the material can be learned in parallel. We also assume some familiarity with cell biology, at the level of a first course for non-majors. The individual chapters in the text indicate the pre-requisites in more detail, most of which are covered either in AM08 or in the supplemental information available from the companion web site.

Domitilla Del Vecchio Cambridge, Massachusetts Richard M. Murray Pasadena, California iv Contents

Notation

This is an internal chapter that is intended for use by the authors in fixing the notation that is used throughout the text. In the first pass of the book we are anticipating several conflicts in notation and the notes here may be useful to early users of the text.

Protein dynamics

For a gene 'genX', we write genX for the gene, m_{genX} for the mRNA and GenX for the protein when they appear in text or chemical formulas. Superscripts are used for covalent modifications, e.g., X^p for phosphorylation. We also use superscripts to differentiate between isomers, so m_{genX}^* might be used to refer to mature RNA or $GenX^f$ to refer to the folded versions of a protein, if required. Mathematical formulas use the italic version of the variable name, but roman font for the gene or isomeric state. The concentration of mRNA is written in text or formulas as m_{genX} (m_{genX}^* for mature) and the concentration of protein as p_{genX} (p_{genX}^f for folded). The same naming conventions are used for common gene/protein combinations: the mRNA concentration of tetR is m_{tetR} , the concentration of the associated protein is p_{tetR} and parameters are α_{tetR} , γ_{tetR} , etc.

For generic genes and proteins, use X to refer to a protein, m_x to refer to the mRNA associated with that protein and x to refer to the gene that encodes X. The concentration of X can be written either as X, p_x or [X], with that order of preference. The concentration of m_x can be written either as m_x (preferred) or $[m_x]$. Parameters that are specific to gene p are written with a subscripted p: α_p , γ_p , etc. Note that although the protein is capitalized, the subscripts are lower case (so indexed by the gene, not the protein) and also in roman font (since they are not a variable).

Transcription and translation. The dynamics of protein production are given by

$$\frac{dm_P}{dt} = \alpha_P \underbrace{-\mu m_P - \delta_P m_P}_{-\delta_P m_P}, \qquad \frac{dP}{dt} = \kappa_P m_P \underbrace{-\mu P - \gamma_P P}_{-\gamma_P P},$$

where $\alpha_{P,0}$ is the (basal) rate of production, δ_P parameterizes the rate of degradation of the mRNA m_p , β_P is the kinetic rate of protein production, μ is the growth rate that leads to dilution of concentrations and γ_P parameterizes the rate of degradation

vi CONTENTS

of the protein P. Since dilution and degradation enter in a similar fashion, we use $\delta = \delta + \mu$ and $\gamma = \gamma + \mu$ to represent the aggregate degradation and dilution rate. If we are looking at a single gene/protein, the various subscripts can be dropped.

When we ignore the mRNA concentration, we write the simplified protein dynamics as

$$\frac{dP}{dt} = \beta_P - \gamma_P P.$$

Assuming that the mRNA dynamics are fast compared to protein production, then the constant β_P is given by

$$\beta_P = \kappa_P \frac{\alpha_P}{\delta_P}.$$

In general, if this does not create confusion, we remove the subscripts "P" from the parameters.

Hill functions. For regulated production of proteins using Hill functions, we modify the constitutive rate of production to be F(Q), in which Q is a transcription factor, instead of α_p or β_p as appropriate. The Hill function is written in the forms

$$F(Q) = \frac{\alpha}{1 + (Q/K)^n}, \qquad F(Q) = \frac{\alpha (Q/K)^n}{1 + (Q/K)^n} + \alpha_0.$$

The notation for F mirrors that of transfer functions in AM08: $F_{p,q}$ represents the input/output relationship between input Q and output P (rate). If the target gene is not particularly relevant, the subscript can represent just the transcription factor:

$$F_{\text{lac}}(Q) = \frac{\alpha_{\text{lac}}}{1 + (Q/K_{\text{laq}})^{n_{\text{lac}}}}.$$

The subscripts can be dropped completely if there is only one Hill function in use.

Concentrations. For a species A, A is its concentration, that is, A := [A]. n_A is the number of A molecules and m_A is the mRNA.

For complexes ES (complex of E and S), we denote C = [ES] and write differential equations with C only or [ES], that is, $\frac{dC}{dt}$ or $\frac{d[ES]}{dt}$.

For names of proteins, such as TetR, we write T := [TetR] and everything fol-

For names of proteins, such as TetR, we write T := [TetR] and everything follows the rules of the species A.

Vector fields. $\dot{x} = f(x)$ or $\dot{x} = f(x, u, \theta)$: all are lower case. Upper case F is reserved for Hill functions.

Some common symbols:

Symbol	LaTeX	Comment
X_{tot}	X_\tot	Total concentration of a species
K_d	\Kd	Dissociation constant
K_m	\Km	Michaelis-Menten constant

CONTENTS vii

Chemical reactions

We write the symbol for a chemical species A using roman type. The number of molecules of a species A is written as n_a . The concentration of the species is occasionally written as [A], but we more often use the notation A, as in the case of proteins, or x_a . For a reaction $A + B \longleftrightarrow C$, we use the notation

R1:
$$A + B \rightleftharpoons C$$
 $\frac{dC}{dt} = a_1 AB - d_1 C$

This notation is primarily intended for situations where we have multiple reactions and need to distinguish between many different constants. Enzymatic reactions have the form

R2:
$$S + E \stackrel{a_2}{\rightleftharpoons} C \stackrel{k}{\rightarrow} P + E$$

For a small number of reactions, the reaction number can be dropped.

It will often be the case that two species A and B will form a molecular bond, in which case we write the resulting species as AB. If we need to distinguish between covalent bonds and hydrogen bonds, we write the latter as A:B. Finally, in some situations we will have labeled section of DNA that are connected together, which we write as A-B, where here A represents the first portion of the DNA strand and B represents the second portion. When describing (single) strands of DNA, we write A' to represent the Watson-Crick complement of the strand A. Thus A-B:B'-A' would represent a double stranded length of DNA with domains A and B.

The choice of representing covalent molecules using the conventional chemical notation AB can lead to some confusion when writing the reaction dynamics using A and B to represent the concentrations of those species. Namely, the symbol AB could represent either the concentration of A times the concentration of B or the concentration of AB. To remove this ambiguity, when using this notation we write [A][B] as $A \cdot B$.

When working with a system of chemical reactions, we write S_i , i = 1,...,n for the species and R_j , j = 1,...,m for the reactions. We write n_i to refer to the molecular count for species i and $x_i = [S_i]$ to refer to the concentration of the species. The individual equations for a given species are written

$$\frac{dx_i}{dt} = \sum_{j=1}^m k_{i,jk} x_j x_k.$$

The collection of reactions are written as

$$\frac{dx}{dt} = Nv(x, \theta), \qquad \frac{dx_i}{d} = N_{ij}v_j(x, \theta),$$

where x_i is the concentration of species S_i , $N \in \mathbb{R}^{n \times m}$ is the stochiometry matrix, v_j is the reaction flux vector for reaction j, and θ is the collection of parameters that

VIII CONTENTS

the define the reaction rates. Occasionally it will be useful to write the fluxes as polynomials, in which case we use the notation

$$v_j(x,\theta) = \sum_k E_{jk} \prod_l x_l^{\epsilon_l^{jk}}$$

where E_{jk} is the rate constant for the kth term of the jth reaction and ϵ_l^{jk} is the stochiometry coefficient for the species x_l .

Generally speaking, coefficients for propensity functions and reaction rate constants are written using lower case $(c_{\xi}, k_i, \text{etc})$. Two exceptions are the dissociation constant, which we write as K_d , and the Michaelis-Menten constant, which we write as K_m .

Figures

In the public version of the text, certain copyrighted figures are missing. The filenames for these figures are listed and many of the figures can be looked up in the following references:

- Cou08 Mechanisms in Transcriptional Regulation by A. J. Courey [17]
- GNM93 J. Greenblatt, J. R. Nodwell and S. W. Mason, "Transcriptional antitermination" [34]
- Mad07 From a to alpha: Yeast as a Model for Cellular Differentiation by H. Madhani [58]
- MBoC The Molecular Biology of the Cell by Alberts et al. [2]
- PKT08 Physical Biology of the Cell by Phillips, Kondev and Theriot [72]

The remainder of the filename lists the chapter and figure number.

Comments intended for reviewers are marked as in this paragraph. These comments generally explain missing material that will be included in the final text.

Review