
Biomolecular Feedback Systems

Domitilla Del Vecchio
MIT

Richard M. Murray
Caltech

Classroom Copy v0.6c, July 11, 2012
© 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

Contents	i
Preface	iii
Notation	v
1 Introductory Concepts	1
1.1 Systems Biology: Modeling, Analysis and Role of Feedback	1
1.2 Dynamics and Control in the Cell	8
1.3 Control and Dynamical Systems Tools [AM08]	21
1.4 Input/Output Modeling [AM08]	30
1.5 From Systems to Synthetic Biology	36
1.6 Further Reading	41
2 Dynamic Modeling of Core Processes	43
2.1 Modeling Techniques	43
2.2 Transcription and Translation	56
2.3 Transcriptional Regulation	61
2.4 Post-Transcriptional Regulation	71
2.5 Cellular subsystems	80
Exercises	87
3 Analysis of Dynamic Behavior	91
3.1 Analysis Near Equilibria	91
3.2 Robustness	104
3.3 Analysis of Reaction Rate Equations	116
3.4 Oscillatory Behavior	121
3.5 Bifurcations	132
3.6 Model Reduction Techniques	137
Exercises	142
4 Stochastic Modeling and Analysis	151
4.1 Stochastic Modeling of Biochemical Systems	151
4.2 Simulation of Stochastic Systems	166

4.3	Input/Output Linear Stochastic Systems	169
	Exercises	175
5	Feedback Examples	179
5.1	The <i>lac</i> Operon	181
5.2	Bacterial Chemotaxis	187
6	Biological Circuit Components	199
6.1	Introduction to Biological Circuit Design	199
6.2	Negative Autoregulation	202
6.3	The Toggle Switch	207
6.4	The Repressilator	209
6.5	Activator-repressor clock	213
6.6	An Incoherent Feedforward Loop (IFFL)	217
	Exercises	219
7	Interconnecting Components	221
7.1	Input/Output Modeling and the Modularity Assumption	221
7.2	Introduction to Retroactivity	222
7.3	Retroactivity in Gene Circuits	226
7.4	Retroactivity in Signaling Systems	231
7.5	Insulation Devices: Retroactivity Attenuation	236
	Exercises	253
8	Design Tradeoffs	255
8.1	Metabolic Burden	255
8.2	Stochastic Effects: Design Tradeoffs between Retroactivity and Noise	269
	Exercises	273
A	Cell Biology Primer	275
A.1	What is a Cell	276
A.2	What is a Genome	303
A.3	Molecular Genetics: Piecing It Together	320
B	Probability and Random Processes	329
B.1	Random Variables	329
B.2	Continuous-State Random Processes	336
B.3	Discrete-State Random Processes	343
	Bibliography	345
	Index	352

Preface

This text serves as a supplement to *Feedback Systems* by Åström and Murray [1] (referred to throughout the text as AM08) and is intended for researchers interested in the application of feedback and control to biomolecular systems. The text has been designed so that it can be used in parallel with *Feedback Systems* 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 AM08, along with additional supplemental material and a copy of these notes, is available on a companion web site:

<http://www.cds.caltech.edu/~murray/amwiki/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 in 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

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. Review

Protein dynamics

For a gene ‘gent’, 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} , δ_{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 , δ_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).

The dynamics of protein production are given by

$$\frac{dm_p}{dt} = \alpha_{p,0} - \underbrace{\mu m_p - \gamma_p m_p}_{-\gamma_p m_p}, \quad \frac{dP}{dt} = \beta_p m_p - \underbrace{\mu P - \delta_p P}_{-\delta_p P},$$

where $\alpha_{p,0}$ is the (basal) rate of production, γ_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 δ_p parameterizes the rate of degradation

of the protein P . Since dilution and degradation enter in a similar fashion, we use $\gamma = \gamma + \mu$ and $\delta = \delta + \mu$ 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

$$\frac{dP}{dt} = \beta_{p,0} - \delta_p P.$$

Assuming that the mRNA dynamics are fast compared to protein production, then the constant $\beta_{p,0}$ is given by

$$\beta_{p,0} = \beta_p \frac{\gamma_p}{\alpha_{p,0}}.$$

For regulated production of proteins using Hill functions, we modify the constitutive rate of production to be $f_p(Q)$ instead of $\alpha_{p,0}$ or $\beta_{p,0}$ as appropriate. The Hill function is written in the forms

$$F_{p,q}(Q) = \frac{\alpha_{p,q}}{1 + (Q/K_{p,q})^{n_{p,q}}}, \quad F_{p,q}(Q) = \frac{\alpha_{p,q}(Q/K_{p,q})^{n_{p,q}}}{1 + (Q/K_{p,q})^{n_{p,q}}}.$$

The notation for F mirrors that of transfer functions: $F_{p,q}$ 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: single letters:

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

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

Some common symbols:

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

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

$$\text{R1:} \quad A + B \xrightleftharpoons[d_1]{a_1} C \quad \frac{dC}{dt} = a_1 AB - d_1 C$$

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

$$\text{R2:} \quad S + E \xrightleftharpoons[d_2]{a_2} C \xrightarrow{k} P + E$$

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 $S_i, i = 1, \dots, n$ for the species and $R_j, j = 1, \dots, m$ 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

$$\frac{dx_i}{dt} = \sum_{j=1}^m m_{i,j} k_j x_j.$$

The collection of reactions are written as

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

where x_i is the concentration of species S_i , $N \in \mathbb{R}^{n \times m}$ is the stoichiometry matrix, v_j is the reaction flux vector for reaction j , and θ is the collection of parameters that 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 k th term of the j th reaction and ϵ_l^{jk} is the stoichiometry coefficient for the species x_l .

Generally speaking, coefficients for propensity functions and reaction rate constants are written using lower case (c_ξ, k_i , 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 [18]
- **GNM93** - J. Greenblatt, J. R. Nodwell and S. W. Mason, “Transcriptional antitermination” [38]
- **Mad07** - *From a to alpha: Yeast as a Model for Cellular Differentiation* by H. Madhani [61]
- **MBoC** - *The Molecular Biology of the Cell* by Alberts et al. [2]
- **PKT08** - *Physical Biology of the Cell* by Phillips, Kondev and Theriot [76]

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

Review

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