Biomolecular Feedback Systems

Domitilla Del Vecchio
MIT

Richard M. Murray
Caltech

DRAFT v0.4b, January 23, 2011
© 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 v

Notation vii

1 Introductory Concepts 1-1
1.1 Systems Biology: Modeling, Analysis and the Role of Feedback . . 1-1
1.2 Dynamics and Control in the Cell 1-9
1.3 Control and Dynamical Systems Tools [AM08] 1-21
1.4 From Systems to Synthetic Biology 1-32
1.5 Further Reading 1-36

I Modeling and Analysis

2 Dynamic Modeling of Core Processes 2-1
2.1 Modeling Techniques 2-1
2.2 Transcription and Translation 2-17
2.3 Transcriptional Regulation 2-21
2.4 Post-Transcriptional Regulation 2-28
2.5 Cellular subsystems (TBD) 2-30
Exercises 2-30

3 Analysis of Dynamic Behavior 3-1
3.1 Input/Output Modeling [AM08] 3-1
3.2 Analysis Near Equilibria 3-5
3.3 Robustness 3-14
3.4 Analysis of Reaction Rate Equations 3-19
3.5 Oscillatory Behavior 3-26
3.6 Analysis Using Describing Functions 3-35
3.7 Bifurcations 3-41
3.8 Model Reduction Techniques 3-42
Exercises 3-46

4 Stochastic Modeling and Analysis **4-1**
4.1 Stochastic Modeling of Biochemical Systems 4-1
4.2 Simulation of Stochastic sections 4-15
4.3 Analysis of Stochastic Systems 4-15
4.4 Linearized Modeling and Analysis 4-15
4.5 Markov chain modeling and analysis 4-15
4.6 System identification techniques 4-15
4.7 Model Reduction 4-15
Exercises 4-15

5 Feedback Examples **5-1**
5.1 The *lac* Operon 5.1-1
5.2 Heat Shock Response in Bacteria 5.2-1
5.3 Bacteriophage λ 5.3-1
5.4 Bacterial Chemotaxis 5.4-1
5.5 Yeast mating response 5.5-1

II Design and Synthesis

6 Biological Circuit Components **6-1**
6.1 Biological Circuit Design 6-1
6.2 Self-repressed gene 6-2
6.3 The Toggle Switch 6-5
6.4 The repressilator 6-6
6.5 Activator-repressor clock 6-9
Exercises 6-14

7 Interconnecting Components **7-1**
7.1 Input/Output Modeling and the Modularity Assumption 7-1
7.2 Beyond the Modularity Assumption: Retroactivity 7-3
7.3 Insulation Devices to Enforce Modularity 7-10
7.4 Design of genetic circuits under the modularity assumption 7-14
7.5 Biological realizations of an insulation component 7-14

8 Design Tradeoffs **8-1**

9 Design Examples **9-1**

III Appendices

A Cell Biology Primer **A-1**
A.1 What is a Cell A-2

A.2 What is a Genome A-28
A.3 Molecular Genetics: Piecing It Together A-44

B A Primer on Control Theory **B-1**
B.1 System Modeling B-1
B.2 Dynamic Behavior B-2
B.3 Linear Systems B-4
B.4 Reachability and observability B-6
B.5 Transfer Functions B-9
B.6 Frequency Domain Analysis B-10
B.7 PID Control B-12
B.8 Limits of Performance B-13
B.9 Robust Performance B-14

C Random Procesess **C-1**
C.1 Random Variables C-1
C.2 Continuous-State Random Processes C-8
C.3 Discrete-State Random Processes C-15
C.4 Input/Output Linear Stochastic Systems C-16

Bibliography **B-1**

Index **I-1**

Preface

This text serves as a supplement to *Feedback Systems* by Åström and Murray [1] (refered 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 text 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 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.

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, $\mathrm{m}_{\mathrm{genX}}$ for the mRNA and GenX for the protein when they appear in text or chemical formulas. We use superscripts to differentiate between isomers, so $\mathrm{m}^{*}_{\mathrm{genX}}$ might be used to refer to mature RNA or $\mathrm{GenX}^{\mathrm{f}}$ 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} ($p^{\mathrm{f}}_{\mathrm{genX}}$ for folded). The same naming conventions are used for common gene/protein combinations: the mRNA concentration of *tetR* is $\mathrm{m}_{\mathrm{tetR}}$, 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 $[\mathrm{m}_{\mathrm{x}}]$. 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

$$\frac{dm_{\mathrm{p}}}{dt} = \alpha_{\mathrm{p},0} - \mu m_{\mathrm{p}} - \gamma_{\mathrm{p}} m_{\mathrm{p}}, \qquad \frac{dP}{dt} = \beta_{\mathrm{p}} m_{\mathrm{p}} - \mu P - \delta_{\mathrm{p}} P,$$

where $\alpha_{\mathrm{p},0}$ is the (basal) rate of production, γ_{p} parameterizes the rate of dilution and 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 $\bar{\gamma} = \gamma + \mu$ and $\bar{\delta} = \delta + \mu$ 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

$$\frac{dP}{dt} = \beta_{\mathrm{p},0} - \bar{\delta}_{\mathrm{p}} P.$$

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

$$\beta_{\mathrm{p},0} = \beta_{\mathrm{p}} \frac{\bar{\gamma}_{\mathrm{p}}}{\alpha_{\mathrm{p},0}}.$$

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

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

The notation for F mirrors that of transfer functions: $F_{\mathrm{p,q}}$ represents the input/output relationship between input Q and output P (rate). The comma can be dropped when the genes in question are single letters:

$$F_{\mathrm{pq}}(Q) = \frac{\alpha_{\mathrm{pq}}}{K_{\mathrm{pq}} + Q^{n_{\mathrm{pq}}}}.$$

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

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 $\mathrm{A} + \mathrm{B} \longleftrightarrow \mathrm{C}$, we use the notation

$$\mathrm{R1}: \quad \mathrm{A} + \mathrm{B} \underset{k^{\mathrm{r}}_{r1}}{\overset{k^{f}_{r1}}{\rightleftharpoons}} \mathrm{C} \qquad \frac{dC}{dt} = k^{f}_{r1} AB - k^{\mathrm{r}}_{r1} C$$

This notation is primarily intended for situations where we have multiple reactions and need to distinguish between many different constants. For a small number of reactions, the reaction number can be dropped or replaced with a single digit (k^{f}_1, k^{r}_2, etc).

It will often be the case that two species A and B will form a covalent bond, in which case we write the resulting species as AB. We will distinguish covalent bonds from much weaker hydrogen bonding by writing 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 convential 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, $i = 1, \ldots, n$ for the species and R_j, $j = 1, \ldots, m$ 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

Missing. Figure out notation here. BST?

The collection of reactions are written as

$$\dot{x} = Nv(x, \theta), \qquad \dot{x}_i = N_{ij} v_j(x, \theta),$$

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 the define the reaction rates. Occassionaly 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 reation 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 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 [33]
- `Mad07` - *From a to alpha: Yeast as a Model for Cellular Differentiation* by H. Madhani [50]
- `MBoC` - *The Molecular Biology of the Cell* by Alberts et al. [2]
- `PKT08` - *Physical Biology of the Cell* [59]

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