# Biomolecular Feedback Systems

Domitilla Del Vecchio
MIT

Richard M. Murray
Caltech

Classroom Copy v0.6d, July 20, 2013

# Contents

# 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

# 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. Superscripts are used for covalent modifications, e.g., $\mathrm{X}^{\mathrm{p}}$ for phosphorylation. We also 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 $m_{\mathrm{genX}}$ ($m^{*}_{\mathrm{genX}}$ for mature) and the concentration of protein as $p_{\mathrm{genX}}$ ($p^{\mathrm{f}}_{\mathrm{genX}}$ 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 $p_{\mathrm{tetR}}$ and parameters are $\alpha_{\mathrm{tetR}}$, $\gamma_{\mathrm{tetR}}$, etc.

For generic genes and proteins, use X to refer to a protein, $\mathrm{m}_{\mathrm{x}}$ 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 $X$, $p_{\mathrm{x}}$ or [X], with that order of preference. The concentration of $\mathrm{m}_{\mathrm{x}}$ can be written either as $m_{\mathrm{x}}$ (preferred) or $[\mathrm{m}_{\mathrm{x}}]$. Parameters that are specific to gene *p* are written with a subscripted p: $\alpha_{\mathrm{p}}$, $\gamma_{\mathrm{p}}$, 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, $\delta_P$ parameterizes the rate of degradation of the mRNA $\mathrm{m}_{\mathrm{p}}$, $\beta_P$ is the kinetic rate of protein production, $\mu$ is the growth rate that leads to dilution of concentrations and $\gamma_P$ parameterizes the rate of degradation

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 $\beta_P$ 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 $\alpha_\mathrm{p}$ or $\beta_\mathrm{p}$ 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: $F_\mathrm{p,q}$ 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:

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

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]. $n_\mathrm{A}$ is the number of A molecules and $m_\mathrm{A}$ 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 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 |
|---|---|---|
| $X_\mathrm{tot}$ | `X_\tot` | Total concentration of a species |
| $K_d$ | `\Kd` | Dissociation constant |
| $K_m$ | `\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 $\mathrm{A} + \mathrm{B} \longleftrightarrow \mathrm{C}$, we use the notation

$$\texttt{R1}: \quad \mathrm{A} + \mathrm{B} \underset{d_1}{\overset{a_1}{\rightleftharpoons}} \mathrm{C} \qquad \frac{dC}{dt} = a_1 AB - d_1 C$$

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

$$\texttt{R2}: \quad \mathrm{S} + \mathrm{E} \underset{d_2}{\overset{a_2}{\rightleftharpoons}} \mathrm{C} \xrightarrow{k} \mathrm{P} + \mathrm{E}$$

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 $\mathrm{S}_i$, $i = 1, \ldots, n$ for the species and $\mathrm{R}_j$, $j = 1, \ldots, m$ for the reactions. We write $n_i$ to refer to the molecular count for species $i$ and $x_i = [\mathrm{S}_i]$ 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

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

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

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 $k$th term of the $j$th reaction and $\epsilon_l^{jk}$ is the stochiometry coefficient for the species $x_l$.

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

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.