Biomolecular Feedback Systems

Domitilla Del Vecchio
MIT

Richard M. Murray
Caltech

DRAFT v0.4a, January 1, 2011

Chapter 2
Dynamic Modeling of Core Processes

The goal of this chapter is to describe basic biological mechanisms in a way that can be represented by simple dynamic models. We begin the chapter a discussion of the basic modeling formalisms that we will utilize to model biomolecular feedback systems. We then proceed to study a number of core processes within the cell, providing different model-based descriptions of the dynamics that will be used in later chapters to analyze and design biomolecular systems. The focus in this chapter and the next is on deterministic models using ordinary differential equations; Chapter 4 describes how to model the stochastic nature of biomolecular systems.

Prerequisites. Readers should have some basic familiarity with cell biology, at the level of the description in Section 1.2 (see also Appendix A), and a basic understanding of ordinary differential equations, at the level of Chapter 2 of AM08 (see also Appendix B).

2.1 Modeling Techniques

In order to develop models for some of the core processes of the cell, we will need to build up a basic description of the biochemical reactions that take place, including production and degradation of proteins, regulation of transcription and translation, intracellular sensing, action and computation, and intercellular signaling. As in other disciplines, biomolecular systems can be modeled in a variety of different ways, at many different levels of resolution, as illustrated in Figure 2.1. The choice of which model to use depends on the questions that we want to answer, and good modeling takes practice, experience and iteration. We must properly capture the aspects of the system that are important, reason about the appropriate temporal and spatial scales to be included, and take into account the types of simulation and analysis tools be be applied. Models that are to be used for analyzing existing systems should make testable predictions and provide insight into the underlying dynamics. Design models must additionally capture enough of the important behavior to allow decisions to be made regarding how to interconnect subsystems, choose parameters and design regulatory elements.

In this section we describe some of the basic modeling frameworks that we will build on throughout the rest of the text. We begin with brief descriptions of the relevant physics and chemistry of the system, and then quickly move to models that focus on capturing the behavior using reaction rate equations. In this chapter

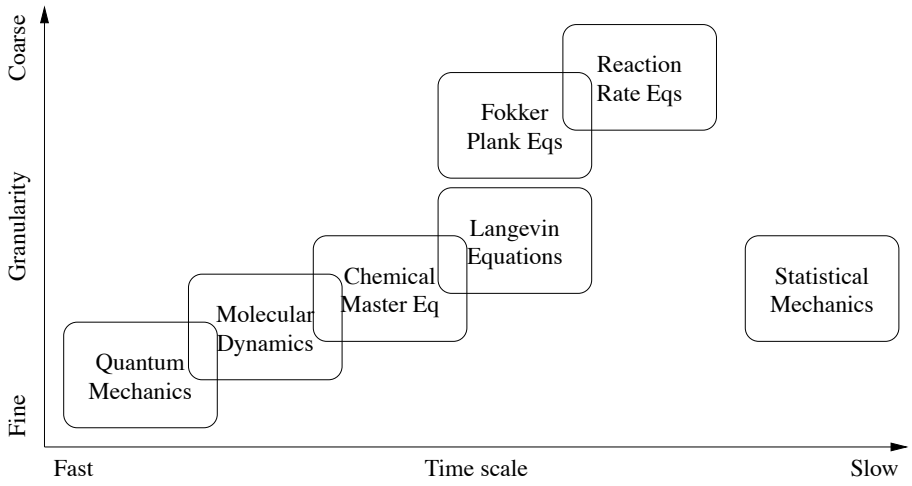


Figure 2.1: Different methods of modeling biomolecular systems.

our emphasis will be on dynamics with time scales measured in seconds to hours and mean behavior averaged across a large number of molecules. We touch only briefly on modeling in the case where stochastic behavior dominates and defer a more detailed treatment until Chapter 4.

Statistical mechanics and chemical kinetics

At the fine end of the modeling scale depicted in Figure 2.1, we can attempt to model the *molecular dynamics* of the cell, in which we attempt to model the individual proteins and other species and their interactions via molecular-scale forces and motions. At this scale, the individual interactions between protein domains, DNA and RNA are resolved, resulting in a highly detailed model of the dynamics of the cell.

For our purposes in this text, we will not require the use of such a detailed scale. Instead, we will start with the abstraction of molecules that interact with each other through stochastic events that are guided by the laws of thermodynamics. We begin with an equilibrium point of view, commonly referred to as *statistical mechanics*, and then briefly describe how to model the (statistical) dynamics of the system using chemical kinetics. We cover both of these points of view very briefly here, primarily as a stepping stone to more deterministic models, and present a more detailed description in Chapter 4.

The underlying representation for both statistical mechanics and chemical kinetics is to identify the appropriate *microstates* of the system. A microstate corresponds to a given configuration of the components (species) in the system relative to each other and we must enumerate all possible configurations between the molecules that are being modeled. As an example, consider the distribution of RNA

Figure 2.2: Microstates for RNA polymerase. Each microstate of the system corresponds to the RNA polymerase being located at some position in the cell. If we discretize the possible locations on the DNA and in the cell, the microstates corresponds to all possible non-overlapping locations of the RNA polymerases. Figure from Phillips, Kondev and Theriot [56]; used with permission of Garland Science.

polymerase in the cell. It is known that most RNA polymerases are bound to the DNA in a cell, either as they produce RNA or as they diffuse along the DNA in search of a promoter site. Hence we can model the microstates of the RNA polymerase system as all possible locations of the RNA polymerase in the cell, with the vast majority of these corresponding to the RNA polymerase at some location on the DNA. This is illustrated in Figure 2.2.

In statistical mechanics, we model the configuration of the cell by the probability that system is in a given microstate. This probability can be calculated based on the energy levels of the different microstates. The laws of statistical mechanics state that if we have a set of microstates Q, then the steady state probability that the system is in a particular microstate q is given by

$$P(q) = \frac{1}{Z} e^{-E_q/(k_B T)}, \tag{2.1}$$

where E_q is the energy associated with the microstate $q \in Q$, k_B is the Boltzmann constant, T is the temperature in degrees Kelvin, and Z is a normalizing factor, known as the *partition function*,

$$Z = \sum_{q \in Q} e^{-E_q/(k_B T)}.$$

(These formulas are described in more detail in Chapter 4.)

Table 2.1: Configurations for a combinatorial promoter with an activator and a repressor. Each row corresponds to a specific macrostate of the promoter in which the listed molecules are bound to the target region. The relative energy of state compared with the ground state provides a measure of the likelihood of that state occurring, with more negative numbers corresponding to more energetically favorable configurations.

State	OR1	OR2	Prom	ΔG	Comment
S_1	–	–	–	$E_0 = 0$	No binding (ground state)
S_2	–	–	RNAP	$E_{\text{RNAP}} = -5$	RNA polymerase bound
S_3 the	R	–	–	$E_{\text{R}} = -10$	Repressor bound
S_4	–	A	–	$E_{\text{A}} = -12$	Activator bound
$S_4 5$	–	A	RNAP	$E_{\text{A:RNAP}} = -15$	Activator and RNA polymerase

By keeping track of those microstates that correspond to a given system state (also called a *macrostate*), we can compute the overall probability that a given macrostate is reached. Thus, if we have a set of states $S \subset Q$ that correspond to a given macrostate, then the probability of being in the set S is given by

$$P(S) = \frac{1}{Z} \sum_{q \in S} e^{-E_q/(k_B T)} = \frac{\sum_{q \in S} e^{-E_q/(k_B T)}}{\sum_{q \in Q} e^{-E_q/(k_B T)}}. \tag{2.2}$$

This can be used, for example, to compute the probability that some RNA polymerase is bound to a given promoter, averaged over many independent samples, and from this we can reason about the rate of expression of the corresponding gene.

Example 2.1 (Combinatorial promoter)**.** A combinatorial promoter is a region of DNA in which multiple transcription factors can bind and influence the subsequent binding of RNA polymerase. Combinatorial promoters appear in a number of natural and engineered circuits and represent a mechanism for creating switch-like behavior, for example by having a gene that controls expression of its own transcription factors.

One method to model a combinatorial promoter is to use the binding energies of the different combinations of proteins to the operator region, and then compute the probability of being in a given promoter state given the concentration of each of the transcription factors. Table 2.1 shows the possible states of a notional promoter that has two operator regions—one that binds a repressor protein R and another that binds an activator protein A. As indicated in the table, the promoter has three (possibly overlapping) regions of DNA: OR1 and OR2 are binding sites for the repressor and activator proteins, and Prom is the location where RNA polymerase binds. (The individual labels are primarily for bookkeeping purposes and may not correspond to physically separate regions of DNA.)

To determine the probabilities of being in a given macrostate, we must compute the individual microstates that occur at a given concentrations of repressor, activator and RNA polymerase. Each microstate corresponds to an individual set of molecules binding in a specific configuration. So if we have n_R repressor molecules, then there is one microstate corresponding to *each* different repressor molecule that is bound, resulting in n_R individual microstates. In the case of configuration S_5, where two different molecules are bound, the number of combinations is given by the product of the numbers of individual molecules, $n_A \cdot n_{RNAP}$, reflecting the possible combinations of molecules that can occupy the promoter sites. The overall partition function is given by summing up the contributions from each microstate:

$$Z = e^{-E_0/(k_B T)} + n_{RNAP}\, e^{-E_{RNAP}/(k_B T)} + n_R\, e^{-E_R/(k_B T)} + n_A\, e^{-E_A/(k_B T)} + n_A n_{RNAP}\, e^{-E_{A:RNAP}/(k_B T)}. \quad (2.3)$$

The probability of a given macrostate is determined using equation (2.2). For example, if we define the promoter to be "active" if RNA polymerase is bound to the DNA, then the probability of being in this macrostate as a function of the various molecular counts is given by

$$\begin{aligned} P_{\text{active}}(n_R, n_A, n_{RNAP}) &= \frac{1}{Z}\left(n_{RNAP}\, e^{-E_{RNAP}/(k_B T)} + n_A\, n_{RNAP} e^{-E_{A:RNAP}/(k_B T)}\right) \\ &= \frac{k_{A:RNAP}\, n_A + k_{RNAP}}{1 + k_{RNAP} + k_R\, n_R + (k_A + k_{A:RNAP}) n_A}, \end{aligned}$$

where

$$k_X = e^{-(E_X - E_0)/(k_B T)}.$$

From this expression we see that if $n_R \gg n_A$ then P_{active} tends to 0 while if $n_A \gg n_R$ then P_{active} tends to 1, as expected. ∇

Statistical mechanics describes the steady state distribution of microstates, but does not tell us how the microstates evolve in time. To include the dynamics, we must consider the *chemical kinetics* of the system and model the probability that we transition from one microstate to another in a given period of time. Let q represent the microstate of the system, which we shall take as a vector of integers that represents the number of molecules of a specific types in given configurations or locations. We describe the kinetics of the system by making use of the *propensity function* $a(\xi; q, t)$, which captures the instantaneous probability that at time t a system will transition between state q and state $q + \xi$, where ξ is the change in the vector of integers representing the microstate.

More specifically, the propensity function is defined such that

$$a(\xi; q, t)dt \quad = \quad \text{Probability that the microstate will transition from state } q \text{ to state } q + \xi \text{ between time } t \text{ and time } t + dt.$$

We will give more detail in Chapter 4 regarding the validity of this functional form, but for now we simply assume that such a function can be defined for our system.

Using the propensity function, we can keep track of the probability distribution for the state by looking at all possible transitions into and out of the current state. Specifically, given $P(q,t)$, the probability of being in state q at time t, we can compute the time derivative $\dot{P}(q,t)$ as

$$\frac{d}{dt}P(q,t) = \sum_{\xi} a(\xi; q-\xi, t)P(q-\xi,t) - \sum_{\xi} a(\xi; q, t)P(q,t). \tag{2.4}$$

This equation (and its many variants) is called the *chemical master equation* (CME). The first sum on the right hand side represents the transitions into the state q from some other state $q-\xi$ and the second sum represents that transitions out of the state q into some other state $q+\xi$. The variable ξ in the sum ranges over all possible transitions between microstates.

Clearly the dynamics of the distribution $P(q,t)$ depends on the form of the propensity function $a(\xi)$. Consider a simple reaction of the form

$$\mathrm{A} + \mathrm{B} \rightleftharpoons \mathrm{AB} \quad \equiv \quad \begin{aligned} \mathrm{Rf}&: \quad \mathrm{A} + \mathrm{B} \longrightarrow \mathrm{AB} \\ \mathrm{Rr}&: \quad \mathrm{AB} \longrightarrow \mathrm{A} + \mathrm{B}. \end{aligned} \tag{2.5}$$

We assume that the reaction takes place in a well-stirred volume Ω and let the configurations q be represented by the number of each species that is present. The forward reaction R_f is a bimolecular reaction and we will see in Chapter 4 that it has a propensity function

$$a(\xi^{\mathrm{f}}; q) = (k^{\mathrm{f}}_{\xi}/\Omega)\, n_{\mathrm{A}} n_{\mathrm{B}},$$

where ξ^{f} represents the forward reaction, n_{A} and n_{B} are the number of molecules of each species and k^{f}_{ξ} is a constant coefficient that depends on the properties of the specific molecules involved. The reverse reaction R_r is a unimolecular reaction and we will see that it has a propensity function

$$a(\xi^{\mathrm{r}}, q) = k^{\mathrm{r}}_{\xi} n_{\mathrm{AB}},$$

where ξ^{r} represents the reverse reaction, k^{r}_{ξ} is a constant coefficient and n_{AB} is the number of molecules of AB that are present.

Example 2.2 (Repression of gene expression)**.** We consider a simple model of repression in which we have a promoter that contains binding sites for RNA polymerase and a repressor protein R. RNA polymerase only binds when the repressor is absent, after which it can undergo an isomerization reaction to form an open complex and initiate transcription. Once the RNA polymerase begins to create mRNA, we assume the promoter region is uncovered, allowing another repressor or RNA polymerase to bind.

The following reactions describe this process:

$$\begin{aligned}
&\text{R1:} && \mathrm{R} + \mathrm{DNA} \rightleftharpoons \mathrm{R{:}DNA} \\
&\text{R2:} && \mathrm{RNAP} + \mathrm{DNA} \rightleftharpoons \mathrm{RNAP{:}DNA^{c}} \\
&\text{R3:} && \mathrm{RNAP{:}DNA^{c}} \longrightarrow \mathrm{RNAP{:}DNA^{o}} \\
&\text{R4:} && \mathrm{RNAP{:}DNA^{o}} \longrightarrow \mathrm{RNAP} + \mathrm{DNA} \quad (+\mathrm{mRNA}),
\end{aligned}$$

where RNAP:DNA^{c} represents the closed complex and RNAP:DNA^{o} represents the open complex. The states for the system depend on the number of molecules of each species and complex that are present. If we assume that we start with n_{R} repressors and n_{RNAP} RNA polymerases, then the possible states for our system are given by

State	DNA	R	RNAP	R:DNA	RNAP:DNA^{c}	RNAP:DNA^{o}
q_1	1	n_{R}	n_{RNAP}	0	0	0
q_2	0	$n_{\mathrm{R}}-1$	n_{RNAP}	1	0	0
q_3	0	n_{R}	$n_{\mathrm{RNAP}}-1$	0	1	0
q_4	0	n_{R}	$n_{\mathrm{RNAP}}-1$	0	0	1

Note that we do not keep of each individual repressor or RNA polymerase molecule that binds to the DNA, but simply keep track of whether they are bound or not.

We can now rewrite the chemical reactions as a set of transitions between the possible microstates of the system. Assuming that all reactions take place in a volume Ω, we use the propensity functions for unimolecular and bimolecular reactions to obtain:

$$\begin{aligned}
&\xi_1^f: \quad \mathrm{q}_1 \longrightarrow \mathrm{q}_2; \quad a(\xi_1^f) = (k_1^f/\Omega)\, n_{\mathrm{R}} && \xi_1^r: \quad \mathrm{q}_2 \longrightarrow \mathrm{q}_1; \quad a(\xi_1^r) = k_1^r \\
&\xi_2^f: \quad \mathrm{q}_1 \longrightarrow \mathrm{q}_3; \quad a(\xi_2^f) = (k_2^f/\Omega)\, n_{\mathrm{RNAP}} && \xi_2^r: \quad \mathrm{q}_3 \longrightarrow \mathrm{q}_1; \quad a(\xi_2^r) = k_2^r \\
&\xi_3: \quad \mathrm{q}_3 \longrightarrow \mathrm{q}_4; \quad a(\xi_3) = k_3 && \xi_4: \quad \mathrm{q}_4 \longrightarrow \mathrm{q}_1; \quad a(\xi_4^r) = k_4
\end{aligned}$$

The chemical master equation can now be written down using the propensity functions for each reaction:

$$\frac{d}{dt}\begin{pmatrix} P(q_1,t) \\ P(q_2,t) \\ P(q_3,t) \\ P(q_4,t) \end{pmatrix} = \begin{pmatrix} -(k_1^f/\Omega)n_{\mathrm{R}} - (k_2^f/\Omega)n_{\mathrm{RNAP}} & k_1^r & k_2^r & k_4 \\ (k_1^f/\Omega)n_{\mathrm{R}} & -k_1^r & 0 & 0 \\ (k_2^f/\Omega)n_{\mathrm{RNAP}} & 0 & -k_2^r - k_3 & 0 \\ 0 & 0 & k_3 & -k_4 \end{pmatrix} \begin{pmatrix} P(q_1,t) \\ P(q_2,t) \\ P(q_3,t) \\ P(q_4,t) \end{pmatrix}.$$

The initial condition for the system can be taken as $P(q,0) = (1,0,0,0)$, corresponding to the state q_1. A simulation showing the evolution of the probabilities is shown in Figure 2.3.

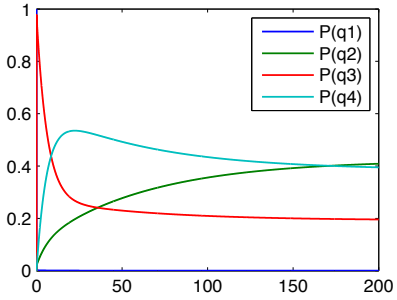


Figure 2.3: Numerical solution of chemical master equation for simple repression model.

The equilibrium solution for the probabilities can be solved by setting $\dot{P} = 0$, which yields:

$$P_e(q_1) = \frac{k_1^r k_4 \Omega (k_2^r + k_3)}{k_1^f k_4 n_{\mathrm{R}} (k_2^r + k_3) + k_1^r k_2^f n_{\mathrm{RNAP}} (k_3 + k_4) + k_1^r k_4 \Omega (k_2^r + k_3)}$$

$$P_e(q_2) = \frac{k_1^f k_4 n_{\mathrm{R}} (k_2^r + k_3)}{k_1^f k_4 n_{\mathrm{R}} (k_2^r + k_3) + k_1^r k_2^f n_{\mathrm{RNAP}} (k_3 + k_4) + k_1^r k_4 \Omega (k_2^r + k_3)}$$

$$P_e(q_3) = \frac{k_1^r k_2^f k_4 n_{\mathrm{RNAP}}}{k_1^f k_4 n_{\mathrm{R}} (k_2^r + k_3) + k_1^r k_2^f n_{\mathrm{RNAP}} (k_3 + k_4) + k_1^r k_4 \Omega (k_2^r + k_3)}$$

$$P_e(q_4) = \frac{k_1^r k_2^f k_3 n_{\mathrm{RNAP}}}{k_1^f k_4 n_{\mathrm{R}} (k_2^r + k_3) + k_1^r k_2^f n_{\mathrm{RNAP}} (k_3 + k_4) + k_1^r k_4 \Omega (k_2^r + k_3)}$$

We see that the functional dependencies are similar to the case of the combinatorial promoter of Example 2.1, but with the binding energies replaced by kinetic rate constants. ∇

The primary difference between the statistical mechanics description given by equation (2.1) and the chemical kinetics description in equation (2.4) is that the master equation formulation describes how the probability of being in a given microstate evolves over time. Of course, if the propensity functions and energy levels are modeled properly, the steady state, average probabilities of being in a given microstate should be the same for both formulations.

Mass action kinetics

Although very general in form, the chemical master equation suffers from being a very high dimensional representation of the dynamics of the system. We shall see in Chapter 4 how to implement simulations that obey the master equation, but in many instances we will not need this level of detail in our modeling. In particular,

there are many situations in which the number of molecules of a given species is such that we can reason about the behavior of a chemically reacting system by keeping track of the *concentration* of each species as a real number. This is of course an approximation, but if the number of molecules is sufficiently large, then the approximation will generally be valid and our models can be dramatically simplified.

To go from the chemical master equation to a simplified form of the dynamics, we begin by making a number of assumptions. First, we assume that we can represent the state of a given species by its concentration $c_A = n_A/\Omega$, where n_A is the number of molecules of A in a given volume Ω. We also treat this concentration as a real number, ignoring the fact that the real concentration is quantized. Finally, we assume that our reactions take place in a well-stirred volume, so that the rate of interactions between two species is solely determined by the concentrations of the species.

Before proceeding, we should recall that in many (and perhaps most) situations inside of cells, these assumptions are *not* particularly good ones. Biomolecular systems often have very small molecular counts and are anything but well mixed. Hence, we should not expect that models based on these assumptions should perform well at all. However, experience indicates that in many cases the basic form of the equations provides a good model for the underlying dynamics and hence we often find it convenient to proceed in this manner.

Putting aside our potential concerns, we can now proceed to write the dynamics of a system consisting of a set of species S_i, $i = 1,\ldots,N$ undergoing a set of reactions R_j, $j = 1,\ldots,M$. We write $x_i = [S_i]$ for the concentration of species i (viewed as a real number). Because we are interested in the case where the number of molecules is large, we no longer attempt to keep track of every possible configuration, but rather simply assume that the state of the system at any given time is given by the concentrations x_i. Hence the state space for our system is given by $x \in \mathbb{R}^N$ and we seek to write our dynamics in the form of a differential equation

$$\dot{x} = f(x, \theta)$$

where $f : \mathbb{R}^N \to \mathbb{R}^N$ describes the rate of change of the concentrations as a function of the instantaneous concentrations and θ represents the parameters that govern the dynamic behavior.

To illustrate the general form of the dynamics, we consider again the case of a basic bimolecular reaction

$$A + B \rightleftharpoons AB.$$

Each time the forward reaction occurs, we decrease the number of molecules of A and B by 1 and increase the number of molecules of AB (a separate species) by 1. Similarly, each time the reverse reaction occurs, we decrease the number of molecules of AB by one and increase the number of molecules of A and B.

Using our discussion of the chemical master equation, we know that the likelihood that the forward reaction occurs in a given interval dt is given by $a(\xi^{\mathrm{f}}; x, t)dt = (k_\xi^{\mathrm{f}}/\Omega) n_{\mathrm{A}} n_{\mathrm{B}} dt$ and the reverse reaction has likelihood $a(\xi^{\mathrm{r}}; q, t) = k_\xi^{\mathrm{r}} n_{\mathrm{AB}}$. It follows that the concentration of the complex AB satisfies

$$\begin{aligned}
[\mathrm{AB}](t+dt) - [\mathrm{AB}](t) &= \mathbb{E}\{n_{\mathrm{AB}}(t+dt)/\Omega - n_{\mathrm{AB}}(t)/\Omega\} \\
&= \Big(a(\xi^{\mathrm{f}}; q - \xi^{\mathrm{f}}, t) - a(\xi^{\mathrm{r}}; q, t)\Big)/\Omega \cdot dt \\
&= \Big(k_\xi^{\mathrm{f}} n_{\mathrm{A}} n_{\mathrm{B}}/\Omega^2 - k_\xi^{\mathrm{r}} n_{\mathrm{AB}}/\Omega\Big) dt \\
&= \Big(k_\xi^{\mathrm{f}} [\mathrm{A}][\mathrm{B}] - k_\xi^{\mathrm{r}} [\mathrm{AB}]\Big) dt.
\end{aligned}$$

Taking the limit as dt approaches zero (but remains large enough that we can still average across multiple reactions, as described in more detail in Chapter 4), we obtain

$$\frac{d}{dt}[\mathrm{AB}] = k_\xi^{\mathrm{f}} [\mathrm{A}][\mathrm{B}] - k_\xi^{\mathrm{r}} [\mathrm{AB}].$$

In a similar fashion we can write equations to describe the dynamics of A and B and the entire system of equations is given by

$$\begin{aligned}
\frac{d}{dt}[\mathrm{A}] &= k_\xi^{\mathrm{r}} [\mathrm{AB}] - k_\xi^{\mathrm{f}} [\mathrm{A}][\mathrm{B}] \\
\frac{d}{dt}[\mathrm{B}] &= k_\xi^{\mathrm{r}} [\mathrm{AB}] - k_\xi^{\mathrm{f}} [\mathrm{A}][\mathrm{B}] \\
\frac{d}{dt}[\mathrm{AB}] &= k_\xi^{\mathrm{f}} [\mathrm{A}][\mathrm{B}] - k_\xi^{\mathrm{r}} [\mathrm{AB}]
\end{aligned}
\qquad \text{or} \qquad
\begin{aligned}
\dot{A} &= k_\xi^{\mathrm{r}} C - k_\xi^{\mathrm{f}} A \cdot B \\
\dot{B} &= k_\xi^{\mathrm{r}} C - k_\xi^{\mathrm{f}} A \cdot B \\
\dot{C} &= k_\xi^{\mathrm{f}} A \cdot B - k_\xi^{\mathrm{r}} C,
\end{aligned}$$

where $C = [\mathrm{AB}]$. These equations are known as the *mass action kinetics* or the *reaction rate equations* for the system. The parameters k_ξ^{f} and k_ξ^{r} are called the *rate constants* and they match the parameters that were used in the underlying propensity functions.

Note that the same rate constants appear in each term, since the rate of production of AB must match the rate of depletion of A and B and vice versa. We adopt the standard notation for chemical reactions with specified rates and write the individual reactions as

$$\mathrm{A} + \mathrm{B} \xrightarrow{k_\xi^{\mathrm{f}}} \mathrm{AB}, \qquad \mathrm{AB} \xrightarrow{k_\xi^{\mathrm{r}}} \mathrm{A} + \mathrm{B},$$

where k_ξ^{f} and k_ξ^{r} are the reaction rates. For bidirectional reactions we can also write

$$\mathrm{A} + \mathrm{B} \underset{k_\xi^{\mathrm{r}}}{\overset{k_\xi^{\mathrm{f}}}{\rightleftharpoons}} \mathrm{AB}.$$

It is easy to generalize these dynamics to more complex reactions. For example, if we have a reversible reaction of the form

$$\mathrm{A} + 2\,\mathrm{B} \underset{k^{\mathrm{r}}}{\overset{k^{\mathrm{f}}}{\rightleftharpoons}} 2\,\mathrm{C} + \mathrm{D},$$

where A, B, C and D are appropriate species and complexes, then the dynamics for the species concentrations can be written as

$$\begin{aligned} \frac{d}{dt}A &= k^{\mathrm{r}}C^2 \cdot D - k^{\mathrm{f}}A \cdot B^2, & \frac{d}{dt}C &= 2k^{\mathrm{f}}A \cdot B^2 - 2k^{\mathrm{r}}C^2 \cdot D, \\ \frac{d}{dt}B &= 2k^{\mathrm{r}}C^2 \cdot D - 2k^{\mathrm{f}}A \cdot B^2, & \frac{d}{dt}D &= k^{\mathrm{f}}A \cdot B^2 - k^{\mathrm{r}}C^2 \cdot D. \end{aligned} \tag{2.6}$$

Rearranging this equation, we can write the dynamics as

$$\frac{d}{dt}\begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ -2 & 2 \\ 2 & -2 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} k^{\mathrm{f}}A \cdot B^2 \\ k^{\mathrm{r}}C^2 \cdot D \end{pmatrix}. \tag{2.7}$$

We see that in this decomposition, the first term on the right hand side is a matrix of integers reflecting the stoichiometry of the reactions and the second term is a vector of rates of the individual reactions.

More generally, given a chemical reaction consisting of a set of species S_i, $i = 1, \ldots, n$ and a set of reactions R_j, $j = 1, \ldots, M$, we can write the mass action kinetics in the form

$$\frac{dx}{dt} = Nv(x),$$

where $N \in \mathbb{R}^{n \times m}$ is the *stoichiometry matrix* for the system and $v(x) \in \mathbb{R}^M$ is the *reaction flux vector*. Each row of $v(x)$ corresponds to the rate at which a given reaction occurs and the corresponding column of the stoichiometry matrix corresponds to the changes in concentration of the relevant species. As we shall see in the next chapter, the structured form of this equation will allow us to explore some of the properties of the dynamics of chemically reacting systems.

Example 2.3 (Covalent modification of a protein)**.** Consider the set of reactions involved in the phosphorylation of a protein by a kinase, as shown in Figure 1.14. Let S represent the substrate, K represent the kinase and S^{p} represent the phosphorylated (activated) substrate. The sets of reactions illustrated in Figure 1.14 are

$$\begin{aligned} &\mathrm{R1}: && \mathrm{K} + \mathrm{ATP} \rightleftharpoons \mathrm{K{:}ATP} \\ &\mathrm{R2}: && \mathrm{S} + \mathrm{K{:}ATP} \rightleftharpoons \mathrm{S{:}K{:}ATP} \\ &\mathrm{R3}: && \mathrm{S{:}K{:}ATP} \longrightarrow \mathrm{S^p{:}K{:}ADP} \\ &\mathrm{R4}: && \mathrm{S^p{:}K{:}ADP} \longrightarrow \mathrm{S^p} + \mathrm{K{:}ADP} \\ &\mathrm{R5}: && \mathrm{K{:}ADP} \rightleftharpoons \mathrm{K} + \mathrm{ADP} \end{aligned}$$

We now write the kinetics for each reaction:

$$\begin{aligned} v_1^{\mathrm{f}} &= k_1^{\mathrm{f}}[\mathrm{K}][\mathrm{ATP}] & v_1^{\mathrm{r}} &= k_1^{\mathrm{r}}[\mathrm{K{:}ATP}] \\ v_2^{\mathrm{f}} &= k_2^{\mathrm{f}}[\mathrm{S}][\mathrm{K{:}ATP}] & v_2^{\mathrm{r}} &= k_2^{\mathrm{r}}[\mathrm{S{:}K{:}ATP}] \\ v_3 &= k_3[\mathrm{S{:}K{:}ATP}] & v_4 &= k_4[\mathrm{S^p{:}K{:}ADP}] \\ v_5^{\mathrm{f}} &= k_5^{\mathrm{f}}[\mathrm{K{:}ADP}] & v_5^{\mathrm{r}} &= k_5^{\mathrm{r}}[\mathrm{K}][\mathrm{K{:}ADP}] \end{aligned}$$

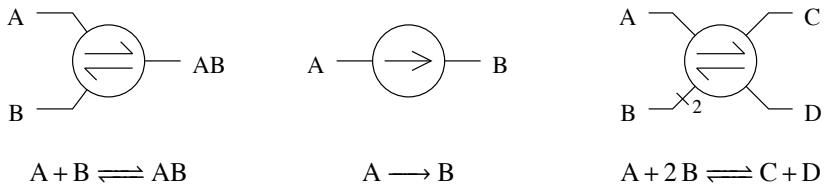


Figure 2.4: Diagrams for chemical reactions.

We treat [ATP] as a constant (regulated by the cell) and hence do not directly track its concentration. (If desired, we could similarly ignore the concentration of ADP since we have chosen not to include the many additional reactions in which it participates.)

The kinetics for each species are thus given by

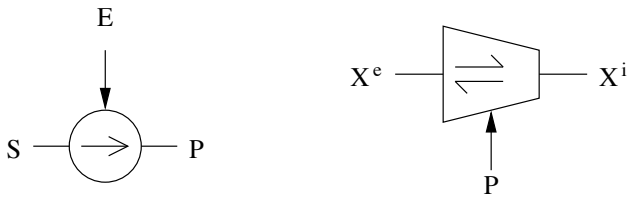
$$\begin{aligned}
\frac{d}{dt}[\mathrm{K}] &= -v_1^{\mathrm{f}} + v_1^{\mathrm{r}} + v_5^{\mathrm{f}} - v_5^{\mathrm{r}} & \frac{d}{dt}[\mathrm{K{:}ATP}] &= v_1^{\mathrm{f}} - v_1^{\mathrm{r}} - v_2^{\mathrm{f}} + v_2^{\mathrm{r}} \\
\frac{d}{dt}[\mathrm{S}] &= -v_2^{\mathrm{f}} + v_2^{\mathrm{r}} & \frac{d}{dt}[\mathrm{S{:}K{:}ATP}] &= v_2^{\mathrm{f}} - v_2^{\mathrm{r}} - v_3 \\
\frac{d}{dt}[\mathrm{S^p}] &= v_4 & \frac{d}{dt}[\mathrm{S^p{:}K{:}ADP}] &= v_3 - v_4 \\
\frac{d}{dt}[\mathrm{ADP}] &= v_5^{\mathrm{f}} - v_5^{\mathrm{r}} & \frac{d}{dt}[\mathrm{K{:}ADP}] &= v_4 - v_5^{\mathrm{f}} + v_5^{\mathrm{r}}.
\end{aligned}$$

In standard stochiometric form, we write

$$\underbrace{\frac{d}{dt}\begin{pmatrix} [\mathrm{K}] \\ [\mathrm{K{:}ATP}] \\ [\mathrm{S}] \\ [\mathrm{S{:}K{:}ATP}] \\ [\mathrm{S^p}] \\ [\mathrm{S^p{:}K{:}ADP}] \\ [\mathrm{ADP}] \\ [\mathrm{K{:}ADP}] \end{pmatrix}}_{x} = \underbrace{\begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 1 & -1 \\ 1 & -1 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 1 \end{pmatrix}}_{N} \underbrace{\begin{pmatrix} v_1^{\mathrm{f}} \\ v_1^{\mathrm{r}} \\ v_2^{\mathrm{f}} \\ v_2^{\mathrm{r}} \\ v_3 \\ v_4 \\ v_5^{\mathrm{f}} \\ v_5^{\mathrm{r}} \end{pmatrix}}_{v(x)}$$

∇

We will often find it convenient to represent collections of chemical reactions using simple diagrams, so that we can see the basic interconnection between various chemical species and properties. A set of diagrams for standard chemical reactions is shown in Figure 2.4.



$S + E \rightleftharpoons ES \longrightarrow E + P$ $\qquad$ $X^e + P \rightleftharpoons X^e{:}P \longrightarrow X^i + P$

(a) Enzymatic reaction (b) Permease-modulated transport

Figure 2.5: Diagrams for enzymatic reactions.

Reduced order mechanisms

In this section, we look at dynamics of some common reactions that occur in biomolecular systems. Under some assumptions on the relative rates or reactions and concentrations of species, it is possible to derive reduced order expressions for the dynamics of the system. We focus here on an informal derivation of the relevant results, but return to these examples in the next chapter to illustrate that the same results can derived using a more formal and rigorous approach.

Simple binding reaction. Consider the reaction

$$A + B \underset{k^r}{\overset{k^f}{\rightleftharpoons}} C, \tag{2.8}$$

where C is the complex AB. Assume that B is a species that is controlled by other reactions in the cell and that the total concentration of A is conserved, so that $A + C = [A] + [AB] = A_{tot}$. If the dynamics of this reaction are fast compared to other reactions in the cell, then the amount of A and C present can be computed as a (steady state) function of B.

To compute how A and C depend on the concentration of B, we must solve for the equilibrium concentrations of A and C. The rate equation for C is given by

$$\frac{dC}{dt} = k^f B \cdot (A_{tot} - C) - k^r C.$$

By setting $\dot{C} = 0$ and letting $K_d := k^r/k^f$, we obtain the expressions

$$C = \frac{BA_{tot}}{B + K_d}, \qquad A = \frac{A_{tot}K_d}{B + K_d}.$$

The constant K_d is the inverse of the affinity of A to B. The steady state value of C increases with B while the steady state value of A decreases with B as more of A is found in the complex C.

Cooperative binding reaction. Assume now that B binds to A only after dimerization, that is, only after binding another molecule of B. Then, we have that reactions (2.8) become

$$\mathrm{B} + \mathrm{B} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{B_d}, \qquad \mathrm{B_d} + \mathrm{A} \underset{k^{\mathrm{r}}}{\overset{k^{\mathrm{f}}}{\rightleftharpoons}} \mathrm{C}, \qquad A + C = A_{\mathrm{tot}},$$

in which $\mathrm{B_d}$ denotes the dimer of B. The corresponding ODE model is given by

$$\frac{dB_d}{dt} = k_1 B^2 - k_2 B_d, \qquad \frac{dC}{dt} = k^{\mathrm{f}} B_d \cdot (A_{\mathrm{tot}} - C) - k^{\mathrm{r}} C.$$

By setting $\dot{B}_d = 0, \dot{C} = 0$, and by defining $K_m := k_1/k_2$, we we obtain that

$$B_d = K_m B^2, \qquad C = \frac{B_d A_{\mathrm{tot}}}{B_d + K_d}, \qquad A = \frac{A_{\mathrm{tot}} K_d}{B_d + K_d},$$

so that

$$C = \frac{K_m A_{\mathrm{tot}} B^2}{K_m B^2 + K_d}, \qquad A = \frac{A_{\mathrm{tot}} K_d}{K_m B^2 + K_d}.$$

As an exercise, the reader can verify that if B binds to A only as a complex of n copies of B, that is,

$$\mathrm{B} + \mathrm{B} + \ldots + \mathrm{B} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{B_n}, \qquad \mathrm{B_n} + \mathrm{A} \underset{k^{\mathrm{r}}}{\overset{k^{\mathrm{f}}}{\rightleftharpoons}} \mathrm{C}, \qquad A + C = A_{\mathrm{tot}},$$

then we have that

$$C = \frac{K_m A_{\mathrm{tot}} B^n}{K_m B^n + K_d}, \qquad A = \frac{A_{\mathrm{tot}} K_d}{K_m B^n + K_d}.$$

In this case, one says that the binding of B to A is *cooperative* with cooperativity n. Figure 2.6 shows the above functions, which are often referred to as *Hill functions*.

Competitive binding reaction. Finally, consider the case in which two species B_a and B_r both bind to A competitively, that is, they cannot be bound to A at the same time. Let C be the complex formed between B_a and A and let $\bar{\mathrm{C}}$ be the complex formed between B_r and A. Then, we have the following reactions

$$\mathrm{B_a} + \mathrm{A} \underset{k^{\mathrm{r}}}{\overset{k^{\mathrm{f}}}{\rightleftharpoons}} \mathrm{C}, \qquad \mathrm{B_r} + \mathrm{A} \underset{\bar{k}^{\mathrm{r}}}{\overset{\bar{k}^{\mathrm{f}}}{\rightleftharpoons}} \bar{C}, \qquad A + C + \bar{C} = A_{\mathrm{tot}},$$

for which we can write the ODE system as

$$\frac{dC}{dt} = k^{\mathrm{f}} B_a \cdot (A_{\mathrm{tot}} - C - \bar{C}) - k^{\mathrm{r}} C, \qquad \frac{d\bar{C}}{dt} = \bar{k}^{\mathrm{f}} B_r \cdot (A_{\mathrm{tot}} - C - \bar{C}) - k^{\mathrm{r}} \bar{C}.$$

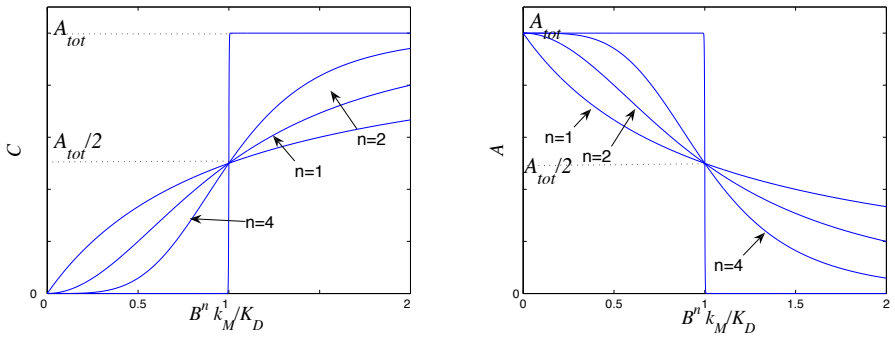


Figure 2.6: Steady state concentrations of the complex C and of A as functions of the concentration of B.

By setting the derivatives to zero, we obtain that

$$C(k^{\mathrm{f}}B_a + k^{\mathrm{r}}) = k^{\mathrm{f}}B_a(A_{\mathrm{tot}} - \bar{C}), \qquad \bar{C}(\bar{k}^{\mathrm{f}}B_r + \bar{k}^{\mathrm{r}}) = \bar{k}^{\mathrm{f}}B_r(A_{\mathrm{tot}} - C),$$

and defining $\bar{K}_d := \bar{k}^{\mathrm{r}}/\bar{k}^{\mathrm{f}}$ leads to

$$\bar{C} = \frac{B_r(A_{\mathrm{tot}} - C)}{B_r + \bar{K}_d}, \qquad C\left(B_a + K_d - \frac{B_aB_r}{B_r + \bar{K}_d}\right) = B_a\left(\frac{\bar{K}_d}{B_r + \bar{K}_d}\right)A_{\mathrm{tot}},$$

from which we finally obtain that

$$C = \frac{B_aA_{\mathrm{tot}}\bar{K}_d}{\bar{K}_dB_a + K_dB_r + K_d\bar{K}_d}, \qquad \bar{C} = \frac{B_rA_{\mathrm{tot}}K_d}{K_dB_r + \bar{K}_dB_a + K_d\bar{K}_d}.$$

Note that in this derivation, we have assumed that both B_a and B_r bind A as monomers. If they were binding as dimers, the reader should verify that they would appear in the final expressions with a power of two. Note also that in this derivation we have assumed that B_a and B_r cannot simultaneously bind to A. If they were binding simultaneously to A, we would have included another complex comprising B_a and B_r and A. Denoting this new complex by C', we would have added also the two additional reactions

$$\mathrm{C} + \mathrm{B_r} \underset{k'^{\mathrm{r}}}{\overset{k'^{\mathrm{f}}}{\rightleftharpoons}} \mathrm{C}', \qquad \bar{C} + \mathrm{B_a} \underset{\bar{k}'^{\mathrm{r}}}{\overset{\bar{k}'^{\mathrm{f}}}{\rightleftharpoons}} \mathrm{C}'$$

and we would have modified the conservation law for A to $A_{\mathrm{tot}} = A + C + \bar{C} + C'$. The reader can verify that in this case a mixed term B_rB_a would appear in the equilibrium expressions.

Enzymatic reaction. A general enzymatic reaction can be written as

$$\mathrm{E} + \mathrm{S} \underset{k^{\mathrm{r}}}{\overset{k^{\mathrm{f}}}{\rightleftharpoons}} \mathrm{C} \xrightarrow{k_{\mathrm{cat}}} \mathrm{E} + \mathrm{P},$$

in which E is an enzyme, S is the substrate to which the enzyme binds to form the complex C, and P is the product resulting from the modification of the substrate S due to the binding with the enzyme E. The rate k^{f} is referred to as association constant, k^{r} as dissociation constant, and k_{cat} as the catalytic rate. Enzymatic reactions are very common and we will see specific instances of them in the sequel, e.g., phosphorylation and dephosphorylation reactions. The corresponding ODE system is given by

$$\frac{dE}{dt} = -k^{\mathrm{f}}E\cdot S + k^{\mathrm{r}}C + k_{\mathrm{cat}}C, \qquad \frac{dC}{dt} = k^{\mathrm{f}}E\cdot S - (k^{\mathrm{r}} + k_{\mathrm{cat}})C,$$
$$\frac{dS}{dt} = -k^{\mathrm{f}}E\cdot S + k^{\mathrm{r}}C, \qquad \frac{dP}{dt} = k_{\mathrm{cat}}C.$$

The total enzyme concentration is usually constant and denoted by E_{tot}, so that $E + C = E_{\mathrm{tot}}$. Substituting in the above equations $E = E_{\mathrm{tot}} - C$, we obtain

$$\frac{dE}{dt} = -k^{\mathrm{f}}(E_{\mathrm{tot}} - C)\cdot S + k^{\mathrm{r}}C + k_{\mathrm{cat}}C, \qquad \frac{dC}{dt} = k^{\mathrm{f}}(E_{\mathrm{tot}} - C)\cdot S - (k^{\mathrm{r}} + k_{\mathrm{cat}})C,$$
$$\frac{dS}{dt} = -k^{\mathrm{f}}(E_{\mathrm{tot}} - C)\cdot S + k^{\mathrm{r}}C, \qquad \frac{dP}{dt} = k_{\mathrm{cat}}C.$$

This system cannot be solved analytically, therefore assumptions have been used in order to reduce it to a simpler form. Michaelis and Menten assumed that the conversion of E and S to C and *vice versa* is much faster than the decomposition of C into E and P. This approximation is called the *quasi-equilibrium* approximation between the enzyme and the complex. This assumption can be translated into the condition

$$k^{\mathrm{f}}, k^{\mathrm{r}} \gg k_{\mathrm{cat}}$$

on the rate constants. Under this assumption and assuming that $S \gg E$ (at least at time 0), C immediately reaches its steady state value (while P is still changing). The steady state value of C is given by solving $k^{\mathrm{f}}(E_{\mathrm{tot}} - C)S - (k^{\mathrm{r}} + k_{\mathrm{cat}})C = 0$ for C, which gives

$$C = \frac{E_{\mathrm{tot}}S}{S + K_m}, \quad \text{with} \quad K_m = \frac{k^{\mathrm{r}} + k_{\mathrm{cat}}}{k^{\mathrm{f}}},$$

in which the constant K_m is called the *Michaelis constant*. Letting $V_{max} = k_{\mathrm{cat}}E_{\mathrm{tot}}$, the resulting kinetics

$$\frac{dP}{dt} = \frac{V_{max}S}{S + K_m}$$

is called *Michaelis-Menten kinetics*. The constant V_{max} is called the maximal velocity (or maximal flux) and it represents the maximal rate that can be obtained when the enzyme is completely saturated by the substrate.

Chemical reaction networks (TBD)

2.2 Transcription and Translation

In this section we consider the processes of transcription and translation, using the modeling techniques described in the previous section to capture the fundamental dynamic behavior. Models of transcription and translation can be done at a variety of levels of detail and which model to use depends on the questions that one wants to consider. We present several levels of modeling here, starting with a fairly detailed set of reactions and ending with highly simplified models that can be used when we are only interested in average production rate of proteins at relatively long time scales.

The basic reactions that underly transcription include the diffusion of RNA polymerase from one part of the cell to the promoter region, binding of an RNA polymerase to the promoter, isomerization from the closed complex to the open complex and finally the production of mRNA, one base pair at a time. To capture this set of reactions, we keep track of the various forms of RNA polymerase according to its location and state: $\mathrm{RNAP}^{\mathrm{c}}$ represents RNA polymerase in the cytoplasm and $\mathrm{RNAP}^{\mathrm{d}}$ is non-specific binding of RNA polymerase to the DNA. We must similarly keep track of the state of the DNA, to insure that multiple RNA polymerases do not bind to the same section of DNA. Thus we can write $\mathrm{DNA}^{\mathrm{p}}$ for the promoter region, $\mathrm{DNA}^{\mathrm{g},i}$ for the ith section of a gene g (whose length can depend on the desired resolution) and $\mathrm{DNA}^{\mathrm{t}}$ for the termination sequence. We write RNAP:DNA to represent RNA polymerase bound to DNA (assumed closed) and $\mathrm{RNAP{:}DNA}^{\mathrm{o}}$ to indicate the open complex. Finally, we must keep track of the mRNA that is produced by transcription: we write $\mathrm{mRNA}^{\mathrm{i}}$ to represent an mRNA strand of length i and assume that the length of the gene of interest is N.

Using these various states of the RNA polymerase and locations on the DNA, we can write a set of reactions modeling the basic elements of transcription as

$$
\begin{aligned}
\text{Binding to DNA:} \quad & \mathrm{RNAP}^{\mathrm{c}} \rightleftharpoons \mathrm{RNAP}^{\mathrm{d}} \\
\text{Diffusion along DNA:} \quad & \mathrm{RNAP}^{\mathrm{d}} \rightleftharpoons \mathrm{RNAP}^{\mathrm{p}} \\
\text{Binding to promoter:} \quad & \mathrm{RNAP}^{\mathrm{p}} + \mathrm{DNA}^{\mathrm{p}} \rightleftharpoons \mathrm{RNAP{:}DNA}^{\mathrm{p}} \\
\text{Isomerization:} \quad & \mathrm{RNAP{:}DNA}^{\mathrm{p}} \rightleftharpoons \mathrm{RNAP{:}DNA}^{\mathrm{o}} \\
\text{Start of transcription:} \quad & \mathrm{RNAP{:}DNA}^{\mathrm{o}} \longrightarrow \mathrm{RNAP{:}DNA}^{\mathrm{g},1} + \mathrm{DNA}^{\mathrm{p}} \\
\text{mRNA creation (index } k\text{):} \quad & \mathrm{RNAP{:}DNA}^{\mathrm{g},1} \longrightarrow \mathrm{RNAP{:}DNA}^{\mathrm{g},2} + \mathrm{mRNA}_{\mathrm{k}}^{1} \\
\text{Elongation, } i = 1, \ldots, N\text{:} \quad & \mathrm{RNAP{:}DNA}^{\mathrm{g},\mathrm{i}+1} + \mathrm{mRNA}_{\mathrm{k}}^{\mathrm{i}} \longrightarrow \mathrm{RNAP{:}DNA}^{\mathrm{g},\mathrm{i}+2} + \mathrm{mRNA}_{\mathrm{k}}^{\mathrm{i}+1} \\
\text{Binding to terminator:} \quad & \mathrm{RNAP{:}DNA}^{\mathrm{g},\mathrm{N}} + \mathrm{mRNA}_{\mathrm{k}}^{\mathrm{N}-1} \longrightarrow \mathrm{RNAP{:}DNA}^{\mathrm{t}} + \mathrm{mRNA}_{\mathrm{k}}^{\mathrm{N}} \\
\text{Termination:} \quad & \mathrm{RNAP{:}DNA}^{\mathrm{t}} \longrightarrow \mathrm{RNAP}^{\mathrm{c}} \\
\text{Degradation:} \quad & \mathrm{mRNA}_{\mathrm{k}}^{\mathrm{N}} \longrightarrow \emptyset
\end{aligned}
\tag{2.9}
$$

This reaction has been written for prokaryotes, but a similar set of reactions could be written for eukaryotes: the main differences would be that the RNA polymerase remains in the nucleus and the mRNA must be spliced and transported to the cytosol. Note that at the start of transcription we "release" the promoter region of the DNA, thus allowing a second RNA polymerase to bind to the promoter while the first RNA polymerase is still transcribing the gene.

A similar set of reactions can be written to model the process of translation. Here we must keep track of the binding of the ribosome to the mRNA, translation of the mRNA sequence into a polypeptide chain and folding of the polypeptide chain into a functional protein. Let $\text{Ribo:mRNA}^{\text{RBS}}$ indicate the ribosome bound to the ribosome binding site, $\text{Ribo:mRNA}^{\text{AA}i}$ the ribosome bound to the ith codon, $\text{Ribo:mRNA}^{\text{start}}$ and $\text{Ribo:mRNA}^{\text{stop}}$ for the start and stop codons, and PPC^{i} for a polypeptide chain consisting of i amino acids. The reactions describing translation can then be written as

$$
\begin{aligned}
\text{Binding to RBS}: &\quad \text{Ribo} + \text{mRNA}_k^{\text{RBS}} \rightleftharpoons \text{Ribo:mRNA}_k^{\text{RBS}} \\
\text{Start of translation}: &\quad \text{Ribo:mRNA}_k^{\text{RBS}} \longrightarrow \text{Ribo:mRNA}_k^{\text{start}} + \text{mRNA}_k^{\text{RBS}} \\
\text{Polypeptide chain creation}: &\quad \text{Ribo:mRNA}_k^{\text{start}} \longrightarrow \text{Ribo:mRNA}_k^{\text{AA2}} + \text{PPC}^{1} \\
\text{Elongation, } i = 1, \ldots, M: &\quad \text{Ribo:mRNA}_k^{\text{AA(i+1)}} + \text{PPC}^{\text{i}} \longrightarrow \text{Ribo:mRNA}_k^{\text{AA(i+2)}} + \text{PPC}^{\text{i+1}} \\
\text{Stop codon}: &\quad \text{Ribo:mRNA}_k^{\text{M}} + \text{PPC}^{\text{M}-1} \longrightarrow \text{Ribo:mRNA}_k^{\text{stop}} + \text{ppc}^{\text{M}} \\
\text{Release of mRNA}: &\quad \text{Ribo:mRNA}_k^{\text{stop}} \longrightarrow \text{Ribo} \\
\text{Folding}: &\quad \text{PPC}^{\text{M}} \longrightarrow \text{protein} \\
\text{Degradation}: &\quad \text{protein} \longrightarrow \emptyset
\end{aligned}
$$

As in the case of transcription, we see that these reactions allow multiple ribosomes to translate the same piece of mRNA by freeing up the ribosome binding site (RBS) when translation begins.

As complex as these reactions are, they are still missing many important effects. For example, we have not accounted for the existence and effects of the 5' and 3' untranslated regions (UTRs) of a gene and we have also left out various error correction mechanisms in which ribosomes can step back and release an incorrect amino acid that has been incorporated into the polypeptide chain. We have also left out the many chemical species that must be present in order for a variety of the reactions to happen (NTPs for mRNA production, amino acids for protein production, etc). Incorporation of these effects requires additional reactions that track the many possible states of the molecular machinery that underlies transcription and translation.

Given a set of reactions, the various stochastic processes that underly detailed models of transcription and translation can be specified using the stochastic modeling framework described briefly in the previous section. In particular, using either models of binding energy or measured rates, we can construct propensity functions

for each of the many reactions that lead to production of proteins, including the motion of RNA polymerase and the ribosome along DNA and RNA. For many problems in which the detailed stochastic nature of the molecular dynamics of the cell are important, these models are the most relevant and they are covered in some detail in Chapter 4.

Alternatively, we can move to the reaction rate formalism and model the reactions using differential equations. To do so, we must compute the various reaction rates, which can be obtained from the propensity functions or measured experimentally. In moving to this formalism, we approximate the concentrations of various species as real numbers, which may not be accurate since some species exist at low molecular counts in the cell. Despite all of these approximations, in many situations the reaction rate equations are perfectly sufficient, particularly if we are interested in the average behavior of a large number of cells.

In some situations, an even simpler model of the transcription, translation and folding processes can be utilized. If we assume that RNA polymerase binds to DNA at some average rate (which includes both the binding and isomerization reactions) and that transcription takes some fixed time (depending on the length of the gene), then the process of transcription can be described using the delay differential equation

$$\frac{dm_p}{dt} = \alpha_{p,0} - \mu m_p - \gamma_p m_p, \qquad m_p^*(t) = e^{-\mu\tau_p^m} m_p(t-\tau_p^m), \tag{2.10}$$

where m_p is the concentration of mRNA for protein P, m_p^* is the concentration of "active" mRNA, $\alpha_{p,0}$ is the rate of production of the mRNA for protein P, μ is the growth rate of the cell (which results in dilution of the concentration) and γ_p is the rate of degradation of the mRNA. Since the dilution and degradation terms are of the same form, we will often combine these terms in the mRNA dynamics and use a single coefficient $\bar{\gamma}_p$.

The active mRNA is the mRNA that is available for translation by the ribosome. We model its concentration through a simple time delay of length τ_p^m that accounts for the transcription of the ribosome binding site in prokaryotes or splicing and transport from the nucleus in eukaryotes. The exponential factor accounts for dilution due to the change in volume of the cell, where μ is the cell growth rate. The constants $\alpha_{p,0}$ and $\bar{\gamma}_p$ capture the average rates of production and degradation, which in turn depend on the more detailed biochemical reactions that underlie transcription.

Once the active mRNA is produced, the process of translation can be described via a similar ordinary differential equation the describes the production of a functional protein:

$$\frac{dP}{dt} = \beta_{p,0} m_p^* - \bar{\delta}_p P, \qquad P^f(t) = e^{-\mu\tau_p^f} P(t-\tau_p^f). \tag{2.11}$$

Here P represents the concentration of the polypeptide chain for the protein, P^f represents the concentration of functional protein (after folding). The parameters that govern the dynamics are $\beta_{p,0}$, the rate of translation of mRNA; $\bar{\delta}_p$ the rate of degradation and dilution of P; and τ_p^f, the time delay associated with folding and other processes required to make the protein functional. The exponential term again accounts for dilution due to cell growth. The degradation and dilution term, parameterized by $\bar{\delta}_p$, captures both rate at which the polypeptide chain is degraded and the rate at which the concentration is diluted due to cell growth.

It will often be convenient to write the dynamics for transcription and translation in terms of the functional mRNA and functional protein. Differentiating the expression for m_p^*, we see that

$$\begin{aligned}\frac{dm_p^*(t)}{dt} &= e^{-\mu\tau_p^m}\dot{m}_p(t-\tau_p^m) \\ &= e^{-\mu\tau_p^m}(\alpha_{p,0}-\bar{\gamma}_p m_p(t-\tau_p^m)) = \bar{\alpha}_{p,0}-\bar{\gamma}_p m_p^*(t),\end{aligned} \tag{2.12}$$

where $\bar{\alpha}_{p,0} = e^{-\mu\tau_p^m}\alpha_{p,0}$. A similar expansion for the active protein dynamics yields

$$\frac{dP^f(t)}{dt} = \bar{\beta}_{p,0}m_p^*(t-\tau_p^f) - \bar{\delta}P^f(t), \tag{2.13}$$

where $\bar{\beta}_{p,0} = e^{-\mu\tau_p^f}\beta_{p,0}$. We shall typically use equations (2.12) and (2.13) as our (reduced) description of protein folding, dropping the superscript f and overbars when there is no risk of confusion.

In many situations the time delays described in the dynamics of protein production are small compared with the time scales at which the protein concentration changes (depending on the values of the other parameters in the system). In such cases, we can simplify our model of the dynamics of protein production and write

$$\frac{dm_p}{dt} = \alpha_{p,0} - \bar{\gamma}_p m_p, \qquad \frac{dP}{dt} = \beta_{p,0}m_p - \bar{\delta}_p P. \tag{2.14}$$

Note that we here have dropped the superscripts $*$ and f since we are assuming that all mRNA is active and proteins are functional and dropped the overbar on α and β since we are assuming the time delays are negligible. We retain the overbars on γ and δ since dilution due to cell growth is still a potentially important factor.

Finally, the simplest model for protein production is one in which we only keep track of the basal rate of production of the protein, without including the mRNA dynamics. This essentially amounts to assuming the mRNA dynamics reach steady state quickly and replacing the first differential equation in equation (2.14) with its equilibrium value. Thus we obtain

$$\frac{dP}{dt} = \beta_{p,0}m_p^e - \delta_p P = \beta_{p,0}\frac{\alpha_{p,0}}{\gamma_p} - \delta_p P =: \beta_p - \delta_p P.$$

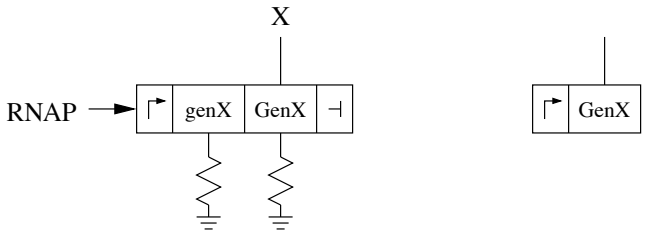


Figure 2.7: Simplified diagrams for protein production. The diagram on the left shows a section of DNA with RNA polymerase as an input, protein concentration as an output and degradation of mRNA and protein. The figure on the right is a simplified view in which only the protein output is indicated.

This model represents a simple first order, linear differential equation for the rate of production of a protein. In many cases this will be a sufficiently good approximate model, although we will see that in many cases it is too simple to capture the observed behavior of a biological circuit.

We will often find it convenient to represent protein production using a simple diagram that hides the details of the particular model that we decide to use. Figure 2.7 shows the symbol that we will use through the text. The diagram is intended to resemble a section of double stranded DNA, with a promoter and terminator at the ends, and then a list of the gene and protein in the middle. The boxes labeled by the gene and protein schematically represent the mRNA and protein concentration, with the line at the left of the DNA represent the input of RNA polymerase and the line on the top representing the the (folded) protein. The symbols at the bottom represent the degradation and dilution of mRNA and protein.

2.3 Transcriptional Regulation

The operation of a cell is governed by the selective expression of genes in the DNA of the organism, which control the various functions the cell is able to perform at any given time. Regulation of protein activity is a major component of the molecular activities in a cell. By turning genes on and off, and modulating their activity in more fine-grained ways, the cell controls the many metabolic pathways in the cell, responds to external stimuli, differentiates into different cell types as it divides, and maintains the internal state of the cell required to sustain life.

The regulation of gene expression and protein activity is accomplished through a variety of molecular mechanisms, as illustrated in Figure 2.8. We see that at each stage of the processing from a gene to a protein, there are potential mechanisms for regulating the production processes. The remainder of this section will focus on transcriptional control, the next section on control between transcription and translation, and the third section on post-translational control mechanisms. We begin with a description of regulation mechanisms in prokaryotes (bacterial) and then

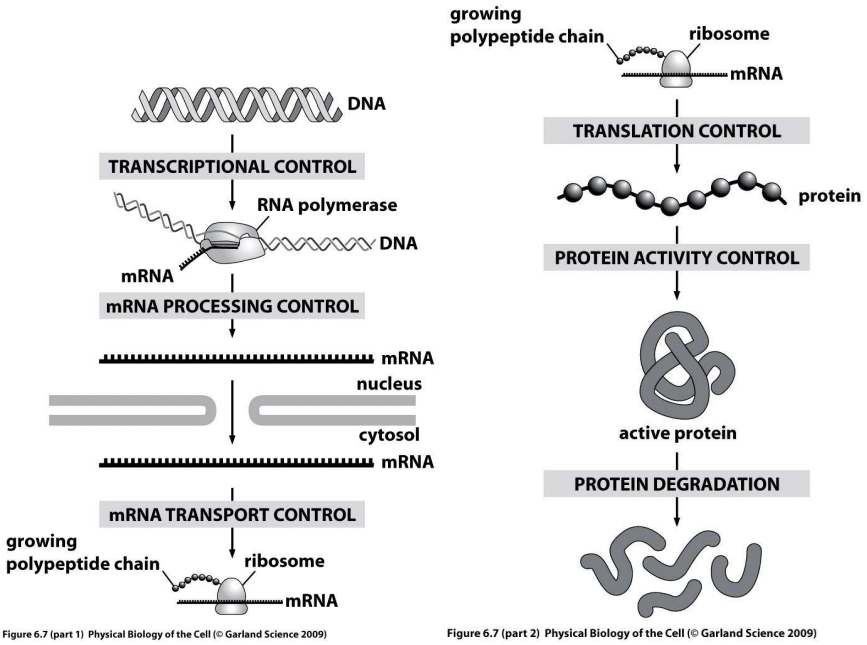


Figure 2.8: Regulation of proteins. Figure from Phillips, Kondev and Theriot [56]; used with permission of Garland Science.

describe the additional mechanisms that are specific to eukaryotes.

Prokaryotic mechanisms

Transcriptional regulation refers to the selective expression of genes by activating or repressing the transcription of DNA into mRNA. The simplest such regulation occurs in prokaryotes, where proteins can bind to "operator regions" in the vicinity of the promoter region of a gene and affect the binding of RNA polymerase and the subsequent initiation of transcription. A protein is called a *repressor* if it blocks the transcription of a given gene, most commonly by binding to the DNA and blocking the access of RNA polymerase to the promoter. An *activator* operates in the opposite fashion: it recruits RNA polymerase to the promoter region and hence transcription only occurs when the activator (protein) is present.

We can capture this set of molecular interactions by modifying the RNA polymerase binding reactions in equation (2.9). For a repressor (Rep), we simply have to add a reaction that represents the repressor bound to the promoter:

$$\text{Repressor binding:} \qquad \text{DNA}^{\text{p}} + \text{Rep} \rightleftharpoons \text{DNA:Rep}$$

This reaction acts to "sequester" the DNA promoter site so that it is no longer available for binding by RNA polymerase (which requires DNA^{p}). The strength

of the repressor is reflected in the reaction rate constants for the repressor binding reaction and the equilibrium concentrations of $\mathrm{DNA^p}$ versus DNA:Rep model the "leakiness" of the repressor.

The modifications for an activator (Act) are a bit more complicated, since we have to modify the reactions to require the presence of the activator before RNA polymerase can bind. One possible mechanism is

$$\begin{aligned}
\text{Activator binding:} &\quad \mathrm{DNA^p + Act \rightleftharpoons DNA{:}Act} \\
\text{Diffusion along DNA:} &\quad \mathrm{RNAP^d \rightleftharpoons RNAP^p} \\
\text{Binding to promoter w/ activator:} &\quad \mathrm{RNAP^p + DNA{:}Act \rightleftharpoons RNAP{:}DNA^o + DNA{:}Act} \\
\text{Binding to promoter w/out activator:} &\quad \mathrm{RNAP^p + DNA^p \rightleftharpoons RNAP{:}DNA^p}
\end{aligned}$$

Here we model both the enhanced binding of the RNA polymerase to the promoter in the presence of the activator, as well as the possibility of binding without an activator. The relative reaction rates determine how strong the activator is and the "leakiness" of transcription in the absence of the activator.

As indicated earlier, many activators and repressors operate in the presence of inducers. To incorporate these dynamics in our description, we simply have to add the reactions that correspond to the interaction of the inducer with the relevant protein. For a negative inducer, we can simply add a reaction in which the inducer binds the regulator protein and effectively sequesters it so that it cannot interact with the DNA. For example, a negative inducer operating on a repressor could be modeled by adding the reaction

$$\mathrm{Rep + Ind \rightleftharpoons Rep{:}Ind.}$$

Positive inducers can be handled similarly, except now we have to modify the binding reactions to only work in the presence of a regulatory protein bound to an inducer. For example, a positive inducer on an activator would have the modified reactions

$$\begin{aligned}
\text{Inducer binding:} &\quad \mathrm{Act + Ind \rightleftharpoons Act{:}Ind} \\
\text{Activator binding:} &\quad \mathrm{DNA^p + Act{:}Ind \rightleftharpoons DNA{:}Act{:}Ind} \\
\text{Diffusion along DNA:} &\quad \mathrm{RNAP^d \rightleftharpoons RNAP^p} \\
\text{Binding to promoter w/ activator:} &\quad \mathrm{RNAP^p + DNA{:}Act{:}Ind \rightleftharpoons RNAP{:}DNA^o + DNA{:}Act{:}Ind}
\end{aligned}$$

A simplified version of the dynamics can be obtained by assuming that transcription factors bind to the DNA rapidly, so that they are in steady state configurations. In this case, we can make use of the steady state statistical mechanics techniques described in Section 2.1 and relate the expression of the gene to the probability that the activator or repressor is bound to the DNA (P_{bound}). This can be done at the level of the reaction rate equation by replacing the differential equations for activator or repressor binding with their steady state values. Here instead

we demonstrate how to account for this rapid binding in the simplified differential equation models presented at the end of Section 2.2.

Recall that given the relative energies of the different microstates of the system, we can compute the probability of a given configuration using equation (2.1):

$$P(q) = \frac{1}{Z} e^{-E_q/(k_B T)}.$$

Consider the regulation of a gene a with a protein concentration given by p_a and a corresponding mRNA concentration m_a. Let b be a second gene with protein concentration p_b that represses the production of protein A through transcriptional regulation. If we let q_bound represent the microstate corresponding to the appropriate activator or repressor bound to the DNA, then we can compute $P(q_\mathrm{bound})$ as a function of the concentration p_b, which we write as $P_\mathrm{bound}(p_\mathrm{b})$. For a repressor, the resulting mRNA dynamics can be written as

$$\frac{dm_a}{dt} = (1 - P_\mathrm{bound}(p_\mathrm{b}))\alpha_{a0} - \gamma_a m_a. \tag{2.15}$$

We see that the effect of the repression is modeled by a modification of the rate of transcription depending on the probability that the repressor is bound to the DNA.

In the case of an activator, we proceed similarly. The modified mRNA dynamics are given by

$$\frac{dm_a}{dt} = P_\mathrm{bound}(p_\mathrm{b})\alpha_{a0} - \gamma_a m_a, \tag{2.16}$$

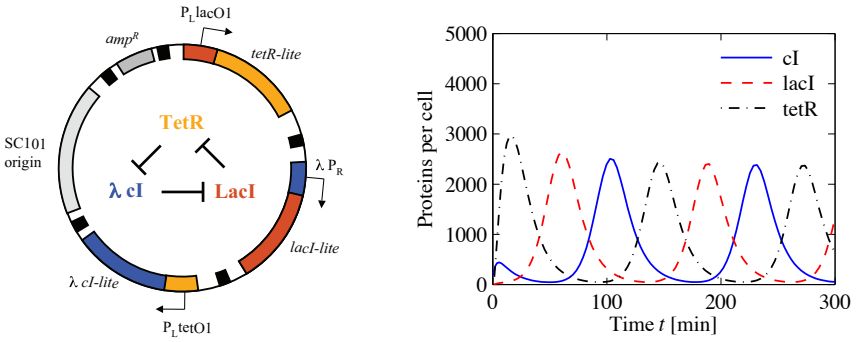
where now we see that B must be bound to the DNA in order for transcription to occur.

As we shall see in Chapter 4 (see also Exercise 2.1, the functional form of P_bound can be nicely approximated by a monotonic rational function, called a *Hill function* [17, 52]. For a repressor, the Hill function is given by

$$f_a^r(p_b) = 1 - P_\mathrm{bound}(p_b) = \frac{\alpha_{ab}}{k_{ab} + p_\mathrm{b}^{n_{ab}}} + \alpha_a,$$

where the subscripts correspond to a protein B repressing production of a protein A, and the parameters α_{ab}, k_{ab} and n_{ab} describe how B represses A. The maximum transcription rate occurs when $p_b = 0$ and is given by $\alpha_{ab}/k_{ab} + \alpha_{a0}$. The minimum rate of transcription occurs when $p_b \to \infty$, giving α_{a0}, which describes the "leakiness" of the promoter. The parameter n_{ab} is called the *Hill coefficient* and determines how close the Hill function is to a step function. The Hill coefficient is often called the *degree of cooperativity* of the reaction, as it often arises from molecular reactions that involve multiple ("cooperating") copies of the protein X.

Example 2.4 (Repressilator)**.** As an example of how these models can be used, we consider the model of a "repressilator," originally due to Elowitz and Leibler [23].



(a) Repressilator plasmid (b) Repressilator simulation

Figure 2.9: The repressilator genetic regulatory network. (a) A schematic diagram of the repressilator, showing the layout of the genes in the plasmid that holds the circuit as well as the circuit diagram (center). (b) A simulation of a simple model for the repressilator, showing the oscillation of the individual protein concentrations. (Figure courtesy M. Elowitz.)

The repressilator is a synthetic circuit in which three proteins each repress another in a cycle. This is shown schematically in Figure 2.9a, where the three proteins are TetR, λ cI and LacI.

The basic idea of the repressilator is that if TetR is present, then it represses the production of λ cI. If λ cI is absent, then LacI is produced (at the unregulated transcription rate), which in turn represses TetR. Once TetR is repressed, then λ cI is no longer repressed, and so on. If the dynamics of the circuit are designed properly, the resulting protein concentrations will oscillate.

We can model this system using three copies of equation (2.15), with A and B replaced by the appropriate combination of TetR, cI and LacI. The state of the system is then given by $x = (m_{\text{TetR}}, p_{\text{TetR}}, m_{\text{cI}}, p_{\text{cI}}, m_{\text{LacI}}, p_{\text{LacI}})$. Figure 2.9b shows the traces of the three protein concentrations for parameters $n = 2$, $\alpha = 0.5$, $k = 6.25 \times 10^{-4}$, $\alpha_0 = 5 \times 10^{-4}$, $\gamma = 5.8 \times 10^{-3}$, $\beta = 0.12$ and $\delta = 1.2 \times 10^{-3}$ with initial conditions $x(0) = (1, 0, 0, 200, 0, 0)$ (following [23]). ∇

For an activator the Hill function is given by

$$f_a^a(p_b) = P_{\text{bound}}(p_b) = \frac{\alpha_{ab} k_{ab} p_{\text{b}}^{n_{ab}}}{k_{ab} + p_{\text{b}}^{n_{ab}}} + \alpha_{a0},$$

where the variables are the same as described previously. Note that in the case of the activator, if p_{b} is zero, then the production rate is α_{a0} (versus $\alpha_{ab} + \alpha_{a0}$ for the repressor). As p_{b} gets large, the first term in the Hill function approaches α_{ab} and the transcription rate becomes $\alpha_{ab} + \alpha_{a0}$ (versus α_{a0} for the repressor). Thus we see that the activator and repressor act in opposite fashion from each other. Figure 2.10 shows the standard Hill functions for activation and repression.

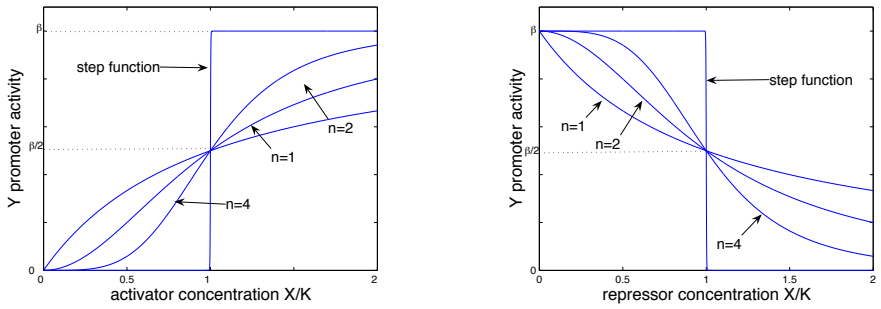


Figure 2.10: Hill function for an activator (left) and for a repressor (right).

In the case where there are inducers present, we can modify our model by adding the appropriate additional reactions. For example, if we have a repressor B with a negative inducer (such as LacI and IPTG), we can add a reaction

$$\mathrm{B} + \mathrm{I} \underset{k^{\mathrm{r}}}{\overset{k^{\mathrm{f}}}{\rightleftharpoons}} \mathrm{B{:}I}.$$

If we assume that this reaction is fast relative to the other dynamics in the system, we can solve for the equilibrium concentration of the inducer bound to the repressor,

$$[\mathrm{B{:}I}] = \frac{k^{\mathrm{f}}}{k^{\mathrm{r}}}[\mathrm{B}][\mathrm{I}],$$

where k^{f} and k^{r} are the forward and reverse reaction rates. We can now attempt to solve for $P_{\mathrm{bound}}(I)$ by computing the amount of repressor that is still free to bind to the DNA.

A simplified case occurs when we assume that most of the repressor is either bound to the inducer or free, so that the amount of B bound to the DNA is small. In this case we can solve for p_{b} in terms of I and then combine the expression for P_{bound} with the modified value of p_{b}. If we let B_T represent the total amount of B present and assume this is constant, we can write

$$B_T = [\mathrm{B{:}I}] + [B]$$

(ignoring any contributions from B:DNA) and solve for p_b as

$$p_b = [B] = \frac{A^T}{1 + (k^{\mathrm{f}}/k^{\mathrm{r}})I}.$$

The resulting expression for $P_{\mathrm{bound}}(I)$ is complicated, but easily computed.

We will often find it convenient to represent the process of regulation in a graphical fashion that hides the specific details of the model that we choose to use. Figure 2.11 shows the notation that we will use in this text to represent the process of transcription, translation and regulation.

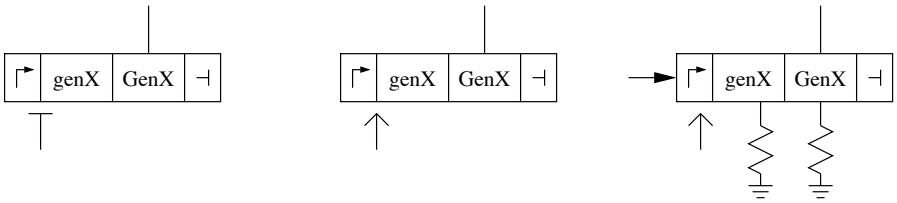


Figure 2.11: Circuit diagrams for transcriptional regulation of a gene. The first two figures represent repression and activation. If desired, additional mechanisms can also be indicated, as shown in the diagram on the right.

We have described how the Hill function can model the regulation of a gene by a single transcription factor. However, genes can also be regulated by multiple transcription factors, some of which may be activators and some may be repressors. The input function can thus take several forms depending on the roles (activators versus repressors) of the various transcription factors [3]. In general, the input function of a transcriptional module that takes as input transcription factors p_i for $i \in \{1, ..., N\}$ will be denoted $f(p_1, ..., p_n)$.

Consider a transcriptional module with input function $f(p_1, ..., p_n)$. The internal dynamics of the transcriptional module usually models mRNA and protein dynamics through the processes of transcription and translation. Protein production is balanced by decay, which can occur through degradation or dilution. Thus, the dynamics of a transcriptional module is often well captured by the ordinary differential equations

$$\frac{dm_y}{dt} = f(p_1, ..., p_n) - \gamma_y m_y, \qquad \frac{dp_y}{dt} = \beta_y m_y - \delta_y p_y, \tag{2.17}$$

where m_y denotes the concentration of mRNA translated by gene y, the constants γ_y and δ_y incorporate the dilution and degradation processes, and β_y is a constant that establishes the rate at which the mRNA is translated.

Several other methods of transcriptional regulation can exist in cells.

Antitermination. Antitermination can also be used as a transcriptional regulatory mechanism. To model its effects, assume that we have a coding region labeled h that occurs after an antitermination site. We modify the termination reactions from equation (2.9):

$$\begin{aligned}
\text{Termination (unchanged):} &\quad \text{RNAP:DNA}^{t} \longrightarrow \text{RNAP}^{c} \\
\text{Binding to utilization site:} &\quad \text{DNA}^{Nut} + \text{N} \rightleftharpoons \text{DNA}^{Nut}\text{:N} \\
\text{Antitermination:} &\quad \text{RNAP:DNA}^{t} + \text{DNA}^{Nut} \\
\text{Termination (unchanged):} &\quad \text{RNAP:DNA}^{t} \longrightarrow \text{RNAP:DNA}^{h,1}
\end{aligned}$$

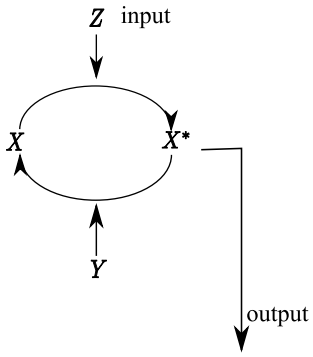


Figure 2.12: Diagram representing a covalent modification cycle.

Regulation in eukaryotes

Transcriptional regulation in eukaryotes is more complex than in prokaryotes. In many situations the transcription of a given gene is affected by many different transcription factors, with multiple molecules being required to initiate and/or suppress transcription.

2.4 Post-Transcriptional Regulation

In addition to regulation of expression through modifications of the process of transcription, cells can also regulate the production and activity of proteins via a collection of other post-transcriptional modifications. These include methods of modulating the translation of proteins, as well as affecting the activity of a protein via changes in its conformation, as shown in Figure 2.8.

RNA-based regulation (TBD)

Allosteric modifications to proteins (TBD)

Covalent modifications to proteins

Covalent modification is a post-translational protein modification that affects the activity of the protein. It plays an important role both in the control of metabolism and in signal transduction. Here, we focus on *reversible* cycles of modification, in which a protein is interconverted between two forms that differ in activity either because of effects on the kinetics relative to substrates or for altered sensitivity to effectors.

At high level, any covalent modification cycle involves a target protein, say X, an enzyme for modifying it, say Z, and one for reversing the modification, say Y (see Figure 2.12). We call X^* the activated protein. There are often allosteric

effectors or further covalent modification systems that regulate the activity of the modifying enzymes, but we do not consider here this added level of complexity. There are several types of covalent modification, depending on the type of activation of the protein. *Phosphorylation* is a covalent modification that takes place mainly in eukaryotes and involves activation of the inactive protein X by addition of a phosphate group. In this case, the enzyme Z is called a *kinase* while the enzyme Y is called *phosphatase*. Another type of covalent modification that is very common in both prokaryotes and eukaryotes is *methylation*. Here, the inactive protein is activated by the addition of a methyl group.

The reactions describing this system are given by the following two enzymatic reactions, also called a *two step reaction model*,

$$\mathrm{Z} + \mathrm{X} \underset{k_r}{\overset{k_f}{\rightleftharpoons}} \mathrm{C} \xrightarrow{k_{cat}} \mathrm{X}^* + \mathrm{Z}, \qquad \mathrm{Y} + \mathrm{X}^* \underset{k'_r}{\overset{k'_f}{\rightleftharpoons}} \mathrm{C}' \xrightarrow{k'_{cat}} \mathrm{X} + \mathrm{Y}.$$

The corresponding ODE model is given by

$$\begin{aligned}
\frac{dZ}{dt} &= -k_f Z \cdot X + (k_{cat} + k_r)C, & \frac{dY}{dt} &= -k'_f Y \cdot X^* + (k'_r + k'_{cat})C', \\
\frac{dX}{dt} &= -k_f Z \cdot X + k_r C + k'_{cat} C', & \frac{dX^*}{dt} &= k_{cat} C - k'_f Y \cdot X^* + k'_r C', \\
\frac{dC}{dt} &= k_f Z \cdot X - (k_r + k_{cat})C, & \frac{dC'}{dt} &= k'_f Y \cdot X^* - (k'_r + k'_{cat})C'.
\end{aligned}$$

Furthermore, we have that the total amounts of enzymes Z and Y are conserved. Denote the total concentrations of Z and Y by Z_{tot}, Y_{tot}, respectively. Then, we also have the conservation laws $Z + C = Z_{tot}$ and $Y + C' = Y_{tot}$. We can thus reduce the above system of ODEs to the following one, in which we have substituted $Z = Z_{tot} - C$ and $Y = Y_{tot} - C'$:

$$\begin{aligned}
\frac{dC}{dt} &= k_f (Z_{tot} - C) \cdot X - (k_r + k_{cat})C \\
\frac{dX^*}{dt} &= k_{cat} C - k'_f (Y_{tot} - C') \cdot X^* + k'_r C' \\
\frac{dC'}{dt} &= k'_f (Y_{tot} - C') \cdot X^* - (k'_r + k'_{cat})C'.
\end{aligned}$$

As for the case of the enzymatic reaction, this system cannot be analytically integrated. To simplify it, we can perform a similar approximation as done for the enzymatic reaction. In particular, the complexes C and C' are often assumed to reach their steady state values very quickly because $k_f, k_r, k'_f, k'_r \gg k_{cat}, k'_{cat}$. Therefore, we can approximate the above system by substituting for C and C' their steady state values given by the solutions to

$$k_f (Z_{tot} - C) \cdot X - (k_r + k_{cat})C = 0$$

and

$$k'_f(Y_{tot} - C') \cdot X^* - (k'_r + k'_{cat})C' = 0.$$

By solving these equations, we obtain that

$$C' = \frac{Y_{tot}X^*}{X^* + K'_m}, \text{ with } K'_m = \frac{k'_r + k'_{cat}}{k'_f}$$

and that

$$C = \frac{Z_{tot}X}{X + K_m}, \text{ with } K_m = \frac{k_r + k_{cat}}{k_f}.$$

As a consequence, the ODE model of the phosphorylation system can be well approximated by

$$\frac{dX^*}{dt} = k_{cat}\frac{Z_{tot}X}{X + K_m} - k'_f\frac{Y_{tot}K'_m}{X^* + K'_m} \cdot X^* + k'_r\frac{Y_{tot}X^*}{X^* + K'_m},$$

which, considering that $k'_f K'_m - k'_r = k'_{cat}$, leads finally to

$$\frac{dX^*}{dt} = k_{cat}\frac{Z_{tot}X}{X + K_m} - k'_{cat}\frac{Y_{tot}X^*}{X^* + K'_m}. \tag{2.18}$$

We will come back to the modeling of this system after we have introduced singular perturbation theory, through which we will be able to perform a formal analysis of this system and mathematically characterize the assumptions needed for approximating the original system by the first order ODE model (2.18).

The full process for phosphorylation and dephosphorylation is actually a bit more complicated than we have shown here and is illustrated in circuit diagram form in Figure 2.13.

Phosphotransfer systems (TBD)

2.5 Cellular subsystems (TBD)

Intercellular Signaling

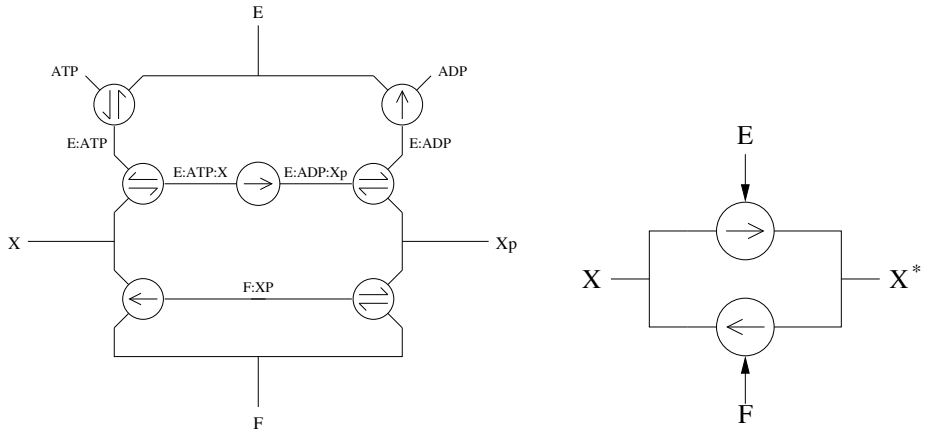
Adaptation

Logical operations

Exercises

2.1 (Hill function for a cooperative repressor) Consider a repressor that binds to an operator site as a dimer:

$$\text{R1:} \quad \mathrm{R} + \mathrm{R} \rightleftharpoons \mathrm{R}_2$$
$$\text{R2:} \quad \mathrm{R}_2 + \mathrm{DNA}^{\mathrm{p}} \rightleftharpoons \mathrm{R}_2\text{:DNA}$$
$$\text{R3:} \quad \mathrm{RNAP} + \mathrm{DNA}^{\mathrm{p}} \rightleftharpoons \mathrm{RNAP\text{:}DNA}^{\mathrm{p}}$$



(a) Full reaction (b) Reduced reaction

Figure 2.13: Circuit diagram for phosphorylation and dephoshorylation of a protein X via a kinase E and phosphatase F. The diagram on the left shows the full set of reactions. A simplified diagram is shown on the right.

Assume that the reactions are at equilibrium and that the RNA polymerase concentration is large (so that [RNAP] is roughly constant). Show that the ratio of the concentration of RNA:DNA$^{\text{p}}$ to the total amount of DNA, D_T, can be written as a Hill function

$$f(R) = \frac{[\text{RNAP:DNA}]}{D_T} = \frac{\alpha}{K + R^2}$$

and give expressions for α and K.

2.2 (Switch-like behavior in cooperative binding) For a cooperative binding reaction

$$\text{B} + \text{B} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \text{B}_\text{d}, \qquad \text{B}_\text{d} + \text{A} \underset{k_r}{\overset{k_f}{\rightleftharpoons}} \text{C}, \quad \text{and} \quad A + C = A_\text{tot},$$

the steady state values of C and A are

$$C = \frac{k_M A_{tot} B^2}{k_M B^2 + K_D}, \quad \text{and} \quad A = \frac{A_\text{tot} K_D}{k_M B^2 + K_D}.$$

Derive the expressions of C and A at the steady state when you modify these reactions to

$$\text{B} + \text{B} + \ldots + \text{B} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \text{B}_\text{n}, \quad \text{B}_\text{n} + \text{A} \underset{k_r}{\overset{k_f}{\rightleftharpoons}} \text{C}, \quad \text{and} \quad A + C = A_\text{tot}.$$

Make MATLAB plots of the expressions that you obtain and verify that as n increases the functions become more switch-like.

2.3 Consider the following modification of the competitive binding reactions:

$$\mathrm{B_a} + \mathrm{A} \underset{k_r}{\overset{k_f}{\rightleftharpoons}} \mathrm{C},\ \mathrm{B_r} + \mathrm{A} \underset{\bar{k}_r}{\overset{\bar{k}_f}{\rightleftharpoons}} \bar{C},$$

and

$$\mathrm{C} + \mathrm{B_r} \underset{k'_r}{\overset{k'_f}{\rightleftharpoons}} \mathrm{C}',\ \text{and}\ \bar{C} + \mathrm{B_a} \underset{\bar{k}'_r}{\overset{\bar{k}'_f}{\rightleftharpoons}} \mathrm{C}'$$

with $A_{tot} = A + C + \bar{C} + C'$. What are the steady state expressions for A and C? What information do you deduce from these expressions if A is a promoter, Ba is an activator protein, and C is the activator/DNA complex that makes the gene transcriptionally active?

2.4 Assume that we have an activator $\mathrm{B_a}$ and a repressor protein $\mathrm{B_r}$. We want to obtain an input function such that when a lot of $\mathrm{B_a}$ is present, the gene is transcriptionally active only if there is no $\mathrm{B_r}$, when low amounts of $\mathrm{B_a}$ are present, the gene is transcriptionally inactive (with or without $\mathrm{B_r}$). Write down the reactions among $\mathrm{B_a}$, $\mathrm{B_r}$, and complexes with the DNA (A) that lead to such an input function. Demonstrate that indeed the set of reactions you picked leads to the desired input function.

2.5 Consider the phosphorylation reactions described in Section 2.4, but suppose that the kinase concentration Z is not constant, but is produced and decays according to the reaction $\mathrm{Z} \underset{k(t)}{\overset{\delta}{\rightleftharpoons}} \emptyset$. How should the system in equation (2.18) be modified? Use a MATLAB simulation to apply a periodic input stimulus $k(t)$ using parameter values: $k_{cat} = k'_{cat} = 10$, $k_f = k'_f = k_r = k'_r = 1$, $\delta = 0.01$. Is the cycle capable of "tracking" the input stimulus? If yes, to what extent? What are the tracking properties depending on?

2.6 Another model for the phosphorylation reactions, referred to as one step reaction model, is given by $\mathrm{Z} + \mathrm{X} \rightleftharpoons \mathrm{X}^* + \mathrm{Z}$ and $\mathrm{Y} + \mathrm{X}^* \rightleftharpoons \mathrm{X} + \mathrm{Y}$, in which the complex formations are neglected. Write down the ODE model and comparing the differential equation of X^* to that of equation (2.18), list the assumptions under which the one step reaction model is a good approximation of the two step reaction model.

2.7 (Transcriptional regulation with delay) Consider a repressor or activator B^* modeled by a Hill function $F_(B)$. Show that in the presence of transcriptional delay τ^m, the dynamics of the active mRNA can be written as

$$\frac{dm^*(t)}{dt} = e^{-\tau^m} F(B(t-\tau^m)) - \bar{\gamma} m^*.$$