Biomolecular Feedback Systems

Domitilla Del Vecchio
MIT

Richard M. Murray
Caltech

Version 1.0a, December 7, 2013

This is the electronic edition of *Biomolecular Feedback Systems* and is available from http://www.cds.caltech.edu/~murray/BFS. Hardcover editions may be purchased from Princeton Univeristy Press, `http://press.princeton.edu/titles/10285.html`.

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 dynamical models. We begin the chapter with 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.

2.1 Modeling Chemical Reactions

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. 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 to 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 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

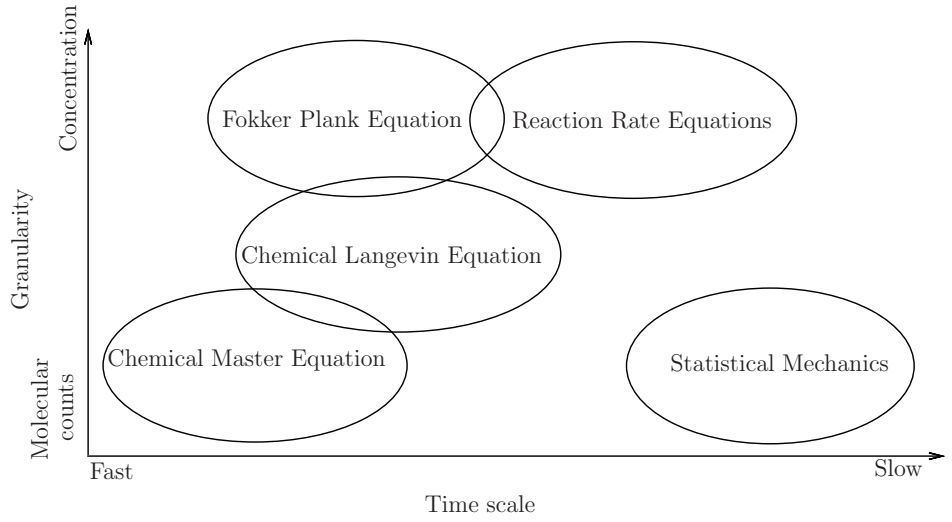


Figure 2.1: Different methods of modeling biomolecular systems.

more detailed treatment until Chapter 4.

Reaction kinetics

At the fine end of the modeling scale, 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 and we will consider the main modeling formalisms depicted in Figure 2.1. We 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 deterministic models.

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 polymerase in the cell. It is known that most RNA polymerases are bound to the DNA in a cell, either as they

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 [75]; used with permission of Garland Science.

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 the 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

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

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*,

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

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 corresponds 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.

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 type (species) in given configurations or locations. Assume we have a set of M chemical reactions `Rj`, $j = 1,\ldots,M$, in which a chemical reaction is a process that leads to the transformation of one set of chemical species to another one. We use ξ_j to represent the change in state q associated with reaction `Rj`. We describe the kinetics of the system by making use of the *propensity function* $a_j(q,t)$ associated with reaction `Rj`, which captures the instantaneous probability that at time t a system will transition between state q and state $q+\xi_j$.

More specifically, the propensity function is defined such that

$$a_j(q,t)dt \quad = \quad \text{Probability that reaction } \texttt{Rj} \text{ will occur between time } t \text{ and time } t+dt \text{ given that the microstate is } q.$$

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 $dP(q,t)/dt$ as

$$\frac{dP}{dt}(q,t) = \sum_{j=1}^{M}\Big(a_j(q-\xi_j)P(q-\xi_j,t) - a_j(q)P(q,t)\Big). \tag{2.3}$$

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 $q-\xi_j$ and the second sum represents that transitions out of the state q.

Clearly the dynamics of the distribution $P(q,t)$ depend on the form of the propensity functions $a_j(q)$. Consider a simple reversible reaction of the form

$$\mathrm{A} + \mathrm{B} \rightleftharpoons \mathrm{AB} \tag{2.4}$$

in which a molecule of A and a molecule of B come together to form the complex AB, in which A and B are bound to each other, and this complex can, in turn, dissociate back into the A and B species. In the sequel, to make notation easier we will sometime represent the complex AB as A:B. It is often useful to write reversible reactions by splitting the forward reaction from the backward reaction:

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

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 `Rf` is a bimolecular reaction and we will see in Chapter 4 that it has a propensity function

$$a_{\mathrm{f}}(q) = \frac{k_{\mathrm{f}}}{\Omega} n_{\mathrm{A}} n_{\mathrm{B}},$$

where k_{f} is a parameter that depends on the forward reaction, and n_{A} and n_{B} are the number of molecules of each species. The reverse reaction `Rr` is a unimolecular reaction and we will see that it has a propensity function

$$a_{\mathrm{r}}(q) = k_{\mathrm{r}} n_{\mathrm{AB}},$$

where k_{r} is a parameter that depends on the reverse reaction and n_{AB} is the number of molecules of AB that are present.

If we now let $q = (n_{\mathrm{A}}, n_{\mathrm{B}}, n_{\mathrm{AB}})$ represent the microstate of the system, then we can write the chemical master equation as

$$\frac{dP}{dt}(n_{\mathrm{A}}, n_{\mathrm{B}}, n_{\mathrm{AB}}) = k_{\mathrm{r}} n_{\mathrm{AB}} P(n_{\mathrm{A}} - 1, n_{\mathrm{B}} - 1, n_{\mathrm{AB}} + 1) - k_{\mathrm{f}} n_{\mathrm{A}} n_{\mathrm{B}} P(n_{\mathrm{A}}, n_{\mathrm{B}}, n_{\mathrm{AB}}).$$

The first term on the right hand side represents the transitions into the microstate $q = (n_{\mathrm{A}}, n_{\mathrm{B}}, n_{\mathrm{AB}})$ and the second term represents the transitions out of that state.

The number of differential equations depends on the number of molecules of A, B and AB that are present. For example, if we start with 1 molecule of A, 1 molecule of B, and 3 molecules of AB, then the possible states and dynamics are

$$\begin{aligned} q_0 &= (1,0,4) \qquad & dP_0/dt &= 3k_{\mathrm{r}} P_1 \\ q_1 &= (2,1,3) \qquad & dP_1/dt &= 4k_{\mathrm{r}} P_0 - 2(k_{\mathrm{f}}/\Omega) P_1 \\ q_2 &= (3,2,2) \qquad & dP_2/dt &= 3k_{\mathrm{r}} P_1 - 6(k_{\mathrm{f}}/\Omega) P_2 \\ q_3 &= (4,3,1) \qquad & dP_3/dt &= 2k_{\mathrm{r}} P_2 - 12(k_{\mathrm{f}}/\Omega) P_3 \\ q_4 &= (5,4,0) \qquad & dP_4/dt &= 1k_{\mathrm{r}} P_3 - 20(k_{\mathrm{f}}/\Omega) P_4, \end{aligned}$$

where $P_i = P(q_i, t)$. Note that the states of the chemical master equation are the probabilities that we are in a specific microstate, and the chemical master equation

is a *linear* differential equation (we see from equation (2.3) that this is true in general).

The primary difference between the statistical mechanics description given by equation (2.1) and the chemical kinetics description in equation (2.3) 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.

Reaction rate equations

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 n_{A}/Ω, 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 create a model for the dynamics of a system consisting of a set of species S_i, $i = 1,\ldots,n$, undergoing a set of reactions `Rj`, $j = 1,\ldots,m$. We write $x_i = [\mathrm{S}_i] = n_{\mathrm{S}_i}/\Omega$ 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 an ordinary differential equation (ODE)

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

where $\theta \in \mathbb{R}^p$ represents the vector of parameters that govern dynamic behavior and $f : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^n$ describes the rate of change of the concentrations as a function of the instantaneous concentrations and parameter values.

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

$$\mathrm{A} + \mathrm{B} \rightleftharpoons \mathrm{AB}.$$

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 $a_\mathrm{f}(q)dt = (k_\mathrm{f}/\Omega)n_\mathrm{A}n_\mathrm{B}dt$ and the reverse reaction has likelihood $a_\mathrm{r}(q) = k_\mathrm{r}n_\mathrm{AB}$. If we assume that n_AB is a real number instead of an integer and ignore some of the formalities of random variables, we can describe the evolution of n_AB using the equation

$$n_\mathrm{AB}(t + dt) = n_\mathrm{AB}(t) + a_f(q - \xi_f)dt - a_r(q)dt.$$

Here we let q be the state of the system with the number of molecules of AB equal to n_AB and ξ_f represents the change in state from the forward reaction (n_A and n_B are decreased by 1 and n_AB is increased by 1). Roughly speaking, this equation states that the (approximate) number of molecules of AB at time $t + dt$ compared with time t increases by the probability that the forward reaction occurs in time dt and descreases by the probability that the reverse reaction occurs in that period.

To convert this expression into an equivalent one for the concentration of the species AB, we write $[\mathrm{AB}] = n_\mathrm{AB}/\Omega$, $[\mathrm{A}] = n_\mathrm{A}/\Omega$, $[\mathrm{B}] = n_\mathrm{B}/\Omega$, and substitute the expressions for $a_f(q)$ and $a_r(q)$:

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

Taking the limit as dt approaches zero, we obtain

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

Our derivation here has skipped many important steps, including a careful derivation using random variables and some assumptions regarding the way in which dt

approaches zero. These are described in more detail when we derive the chemical Langevin equation (CLE) in Chapter 4, but the basic form of the equations are correct under the assumptions that the reactions are well-stirred and the molecular counts are sufficiently large.

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

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

where $C = [\mathrm{AB}]$, $A = [\mathrm{A}]$, and $B = [\mathrm{B}]$. These equations are known as the *mass action kinetics* or the *reaction rate equations* for the system. The parameters k_f and k_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

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

where k_f and k_r are the reaction rates. For bidirectional reactions we can also write

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

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, $i = 1,\dots,n$ and a set of reactions `Rj`, $j = 1,\dots,m$, 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. For example, for the system in equation (2.7) we have

$$x = (A,B,C,D), \qquad N = \begin{pmatrix} -1 & 1 \\ -2 & 2 \\ 2 & -2 \\ 1 & -1 \end{pmatrix}, \qquad v(x) = \begin{pmatrix} k_{\mathrm{f}} A \cdot B^2 \\ k_{\mathrm{r}} C^2 \cdot D \end{pmatrix}.$$

The conservation of species is at the basis of reaction rate models since species are usually transformed, but are not created from nothing or destroyed. Even the basic process of protein degradation transforms a protein of interest A into a product X that is not used in any other reaction. Specifically, the degradation rate of a protein is determined by the amounts of proteases present, which bind to recognition sites (degradation tags) and then degrade the protein. Degradation of a protein A by a protease P can then be modeled by the following two-step reaction

$$\mathrm{A} + \mathrm{P} \underset{d}{\overset{a}{\rightleftharpoons}} \mathrm{AP} \xrightarrow{k} \mathrm{P} + \mathrm{X}.$$

As a result of the reaction, protein A has "disappeared", so that this reaction is often simplified to $\mathrm{A} \longrightarrow \emptyset$. Similarly, the birth of a molecule is a complicated process that involves many reactions and species, as we will see later in this chapter. When the process that creates a species of interest A is not relevant for the problem under study, we will use the shorter description of a birth reaction given by

$$\emptyset \xrightarrow{k_{\mathrm{f}}} \mathrm{A}$$

and describe its dynamics using the differential equation

$$\frac{dA}{dt} = k_{\mathrm{f}}.$$

Example 2.1 (Covalent modification of a protein)**.** Consider the set of reactions involved in the phosphorylation of a protein by a kinase, as shown in Figure 2.3.

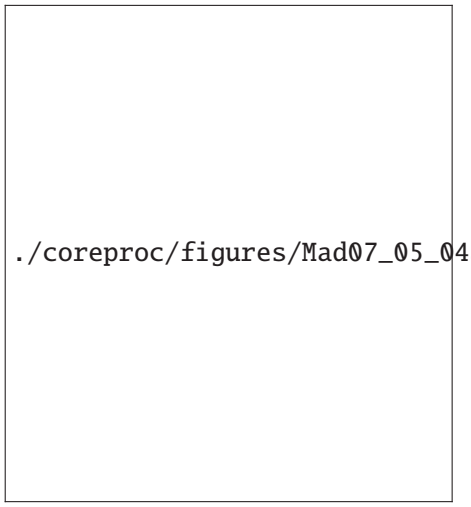


Figure 2.3: Phosphorylation of a protein via a kinase. In the process of phosphorylation, a protein called a kinase binds to ATP (adenosine triphosphate) and transfers one of the phosphate groups (P) from ATP to a substrate, hence producing a phosphorylated substrate and ADP (adenosine diphosphate). Reproduced from Madhani [61]; permission pending.

Let S represent the substrate, K represent the kinase and S^* represent the phosphorylated (activated) substrate. The sets of reactions illustrated in Figure 2.3 are

$$\begin{aligned}
&\texttt{R1}: \quad \mathrm{K} + \mathrm{ATP} \longrightarrow \mathrm{K{:}ATP} && \texttt{R5}: \quad \mathrm{S{:}K{:}ATP} \longrightarrow \mathrm{S^*{:}K{:}ADP} \\
&\texttt{R2}: \quad \mathrm{K{:}ATP} \longrightarrow \mathrm{K} + \mathrm{ATP} && \texttt{R6}: \quad \mathrm{S^*{:}K{:}ADP} \longrightarrow \mathrm{S^*} + \mathrm{K{:}ADP} \\
&\texttt{R3}: \quad \mathrm{S} + \mathrm{K{:}ATP} \longrightarrow \mathrm{S{:}K{:}ATP} && \texttt{R7}: \quad \mathrm{K{:}ADP} \longrightarrow \mathrm{K} + \mathrm{ADP} \\
&\texttt{R4}: \quad \mathrm{S{:}K{:}ATP} \longrightarrow \mathrm{S} + \mathrm{K{:}ATP} && \texttt{R8}: \quad \mathrm{K} + \mathrm{ADP} \longrightarrow \mathrm{K{:}ADP}.
\end{aligned}$$

We now write the kinetics for each reaction:

$$\begin{aligned}
v_1 &= k_1\,[\mathrm{K}][\mathrm{ATP}], & v_5 &= k_5\,[\mathrm{S{:}K{:}ATP}], \\
v_2 &= k_2\,[\mathrm{K{:}ATP}], & v_6 &= k_6\,[\mathrm{S^*{:}K{:}ADP}], \\
v_3 &= k_3\,[\mathrm{S}][\mathrm{K{:}ATP}], & v_7 &= k_7\,[\mathrm{K{:}ADP}], \\
v_4 &= k_4\,[\mathrm{S{:}K{:}ATP}], & v_8 &= k_8\,[\mathrm{K}][\mathrm{ADP}].
\end{aligned}$$

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

$$\frac{d}{dt}[\mathrm{K}] = -v_1 + v_2 + v_7 - v_8, \qquad \frac{d}{dt}[\mathrm{S}^*] = v_6,$$
$$\frac{d}{dt}[\mathrm{K{:}ATP}] = v_1 - v_2 - v_3 + v_4, \qquad \frac{d}{dt}[\mathrm{S}^*{:}\mathrm{K{:}ADP}] = v_5 - v_6,$$
$$\frac{d}{dt}[\mathrm{S}] = -v_3 + v_4, \qquad \frac{d}{dt}[\mathrm{ADP}] = v_7 - v_8,$$
$$\frac{d}{dt}[\mathrm{S{:}K{:}ATP}] = v_3 - v_4 - v_5, \qquad \frac{d}{dt}[\mathrm{K{:}ADP}] = v_6 - v_7 + v_8.$$

Collecting these equations together and writing the state as a vector, we obtain

$$\underbrace{\frac{d}{dt}\begin{pmatrix} [\mathrm{K}] \\ [\mathrm{K{:}ATP}] \\ [\mathrm{S}] \\ [\mathrm{S{:}K{:}ATP}] \\ [\mathrm{S}^*] \\ [\mathrm{S}^*{:}\mathrm{K{:}ADP}] \\ [\mathrm{ADP}] \\ [\mathrm{K{:}ADP}] \end{pmatrix}}_{x} = \underbrace{\begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 1 & -1 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 1 \end{pmatrix}}_{N} \underbrace{\begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \end{pmatrix}}_{v(x)},$$

which is in standard stoichiometric form. ∇

Reduced order mechanisms

In this section, we look at the dynamics of some common reactions that occur in biomolecular systems. Under some assumptions on the relative rates of 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 be derived using a more formal and rigorous approach.

Simple binding reaction. Consider the reaction in which two species A and B bind reversibly to form a complex C = AB:

$$\mathrm{A} + \mathrm{B} \underset{d}{\overset{a}{\rightleftharpoons}} \mathrm{C}, \tag{2.8}$$

where a is the *association rate constant* and d is the *dissociation rate constant*. 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 $A + C = [\mathrm{A}] + [\mathrm{AB}] = A_{\mathrm{tot}}$. 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 the amount of B.

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

$$\frac{dC}{dt} = aB \cdot A - dC = aB \cdot (A_{\text{tot}} - C) - dC.$$

By setting $dC/dt = 0$ and letting $K_{\text{d}} := d/a$, we obtain the expressions

$$C = \frac{A_{\text{tot}}(B/K_{\text{d}})}{1 + (B/K_{\text{d}})}, \qquad A = \frac{A_{\text{tot}}}{1 + (B/K_{\text{d}})}.$$

The constant K_{d} is called the *dissociation constant* of the reaction. Its inverse measures the affinity of A binding 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.

Note that when $B \approx K_{\text{d}}$, A and C have equal concentration. Thus the higher the value of K_{d}, the more B is required for A to form the complex C. K_{d} has the units of concentration and it can be interpreted as the concentration of B at which half of the total number of molecules of A are associated with B. Therefore a high K_{d} represents a weak affinity between A and B, while a low K_{d} represents a strong affinity.

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

$$\text{B} + \text{B} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \text{B}_2, \qquad \text{B}_2 + \text{A} \underset{d}{\overset{a}{\rightleftharpoons}} \text{C}, \qquad A + C = A_{\text{tot}},$$

in which B_2 = B:B represents the dimer of B, that is the complex of two molecules of B bound to each other. The corresponding ODE model is given by

$$\frac{dB_2}{dt} = k_1 B^2 - k_2 B_2 - aB_2 \cdot (A_{\text{tot}} - C) + dC, \qquad \frac{dC}{dt} = aB_2 \cdot (A_{\text{tot}} - C) - dC.$$

By setting $dB_2/dt = 0$, $dC/dt = 0$, and by defining $K_m := k_2/k_1$, we obtain that

$$B_2 = B^2/K_m, \qquad C = \frac{A_{\text{tot}}(B_2/K_{\text{d}})}{1 + (B_2/K_{\text{d}})}, \qquad A = \frac{A_{\text{tot}}}{1 + (B_2/K_{\text{d}})},$$

so that

$$C = \frac{A_{\text{tot}} B^2/(K_m K_{\text{d}})}{1 + B^2/(K_m K_{\text{d}})}, \qquad A = \frac{A_{\text{tot}}}{1 + B^2/(K_m K_{\text{d}})}.$$

As an exercise (Exercise 2.2), the reader can verify that if B binds to A as a complex of n copies of B, that is,

$$\text{B} + \text{B} + \cdots + \text{B} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \text{B}_n, \qquad \text{B}_n + \text{A} \underset{d}{\overset{a}{\rightleftharpoons}} \text{C}, \qquad A + C = A_{\text{tot}},$$

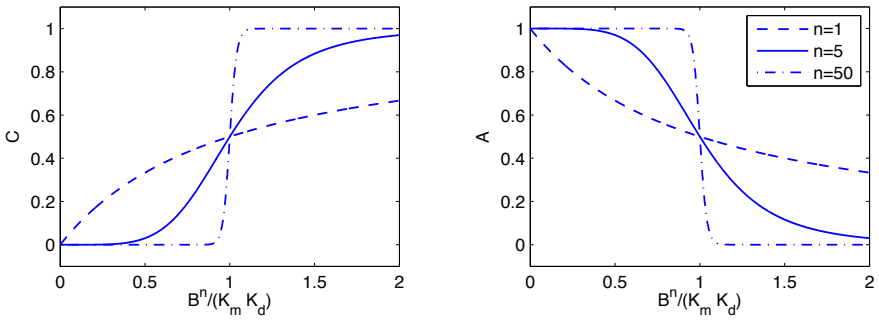


Figure 2.4: Steady state concentrations of the complex C and of A as functions of the concentration of B.

then we have that the expressions of C and A change to

$$C = \frac{A_{\text{tot}} B^n/(K_m K_{\text{d}})}{1 + B^n/(K_m K_{\text{d}})}, \qquad A = \frac{A_{\text{tot}}}{1 + B^n/(K_m K_{\text{d}})}.$$

In this case, we say that the binding of B to A is *cooperative* with cooperativity n. Figure 2.4 shows the above functions, which are often referred to as *Hill functions*.

Another type of cooperative binding is when a species R can bind A only after another species B has bound A. In this case, the reactions are given by

$$\text{B} + \text{A} \underset{d}{\overset{a}{\rightleftharpoons}} \text{C}, \qquad \text{R} + \text{C} \underset{d'}{\overset{a'}{\rightleftharpoons}} \text{C}', \qquad A + C + C' = A_{\text{tot}}.$$

Proceeding as above by writing the ODE model and equating the time derivatives to zero to obtain the equilibrium, we obtain the equilibrium relations

$$C = \frac{1}{K_{\text{d}}} B(A_{\text{tot}} - C - C'), \qquad C' = \frac{1}{K'_{\text{d}} K_{\text{d}}} R(A_{\text{tot}} - C - C').$$

By solving this system of two equations for the unknowns C' and C, we obtain

$$C' = \frac{A_{\text{tot}}(B/K_{\text{d}})(R/K'_{\text{d}})}{1 + (B/K_{\text{d}}) + (B/K_{\text{d}})(R/K'_{\text{d}})}, \qquad C = \frac{A_{\text{tot}}(B/K_{\text{d}})}{1 + (B/K_{\text{d}}) + (B/K_{\text{d}})(R/K'_{\text{d}})}.$$

In the case in which B would bind cooperatively with other copies of B with cooperativity n, the above expressions become

$$C' = \frac{A_{\text{tot}}(B^n/K_{\text{m}}K_{\text{d}})(R/K'_{\text{d}})}{1 + (B^n/K_{\text{m}}K_{\text{d}})(R/K'_{\text{d}}) + (B^n/K_{\text{m}}K_{\text{d}})},$$
$$C = \frac{A_{\text{tot}}(B^n/K_{\text{m}}K_{\text{d}})}{1 + (B^n/K_{\text{m}}K_{\text{d}})(R/K'_{\text{d}}) + (B^n/K_{\text{m}}K_{\text{d}})}.$$

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_a be the complex formed between B_a and A and let C_r be the complex formed between B_r and A. Then, we have the following reactions

$$B_a + A \underset{d}{\overset{a}{\rightleftharpoons}} C_a, \qquad B_r + A \underset{d'}{\overset{a'}{\rightleftharpoons}} C_r, \qquad A + C_a + C_r = A_{tot},$$

for which we can write the differential equation model as

$$\frac{dC_a}{dt} = aB_a \cdot (A_{tot} - C_a - C_r) - dC_a, \qquad \frac{dC_r}{dt} = a'B_r \cdot (A_{tot} - C_a - C_r) - d'C_r.$$

By setting the time derivatives to zero, we obtain

$$C_a(aB_a + d) = aB_a(A_{tot} - C_r), \qquad C_r(a'B_r + d') = a'B_r(A_{tot} - C_a),$$

so that

$$C_r = \frac{B_r(A_{tot} - C_a)}{B_r + K'_d}, \qquad C_a\left(B_a + K_d - \frac{B_aB_r}{B_r + K'_d}\right) = B_a\left(\frac{K'_d}{B_r + K'_d}\right)A_{tot},$$

from which we finally determine that

$$C_a = \frac{A_{tot}(B_a/K_d)}{1 + (B_a/K_d) + (B_r/K'_d)}, \qquad C_r = \frac{A_{tot}(B_r/K'_d)}{1 + (B_a/K_d) + (B_r/K'_d)}.$$

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 (see Exercise 2.3).

Note also that in this derivation we have assumed that the binding is competitive, that is, B_a and B_r cannot simultaneously bind to A. If they were binding simultaneously to A, we would have to include another complex comprising B_a, B_r and A. Denoting this new complex by C′, we must add the two additional reactions

$$C_a + B_r \underset{\bar{d}}{\overset{\bar{a}}{\rightleftharpoons}} C', \qquad C_r + B_a \underset{\bar{d}'}{\overset{\bar{a}'}{\rightleftharpoons}} C',$$

and we would have modified the conservation law for A to $A_{tot} = A + C_a + C_r + C'$. The reader can verify that in this case a mixed term B_rB_a would appear in the equilibrium expressions (see Exercise 2.4).

Enzymatic reaction. A general enzymatic reaction can be written as

$$E + S \underset{d}{\overset{a}{\rightleftharpoons}} C \xrightarrow{k} E + P,$$

in which E is an enzyme, S is the substrate to which the enzyme binds to form the complex C = ES, and P is the product resulting from the modification of the

substrate S due to the binding with the enzyme E. Here, a and d are the association and dissociation rate constants as before, and k is the catalytic rate constant. Enzymatic reactions are very common and include phosphorylation as we have seen in Example 2.1 and as we will see in more detail in the sequel. The corresponding ODE model is given by

$$\frac{dS}{dt} = -aE\cdot S + dC, \qquad \frac{dC}{dt} = aE\cdot S - (d+k)C,$$
$$\frac{dE}{dt} = -aE\cdot S + dC + kC, \qquad \frac{dP}{dt} = kC.$$

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

$$\frac{dS}{dt} = -a(E_{\text{tot}} - C)\cdot S + dC, \qquad \frac{dC}{dt} = a(E_{\text{tot}} - C)\cdot S - (d+k)C,$$
$$\frac{dE}{dt} = -a(E_{\text{tot}} - C)\cdot S + dC + kC, \qquad \frac{dP}{dt} = kC.$$

This system cannot be solved analytically, therefore assumptions must be 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. Under this assumption and letting the initial concentration $S(0)$ be sufficiently large (see Example 3.12), C immediately reaches its steady state value (while P is still changing). This approximation is called the *quasi-steady state assumption* and the mathematical conditions on the parameters that justify it will be dealt with in Section 3.5. The steady state value of C is given by solving $a(E_{\text{tot}} - C)S - (d+k)C = 0$ for C, which gives

$$C = \frac{E_{\text{tot}}S}{S + K_m}, \quad \text{with} \quad K_m = \frac{d+k}{a},$$

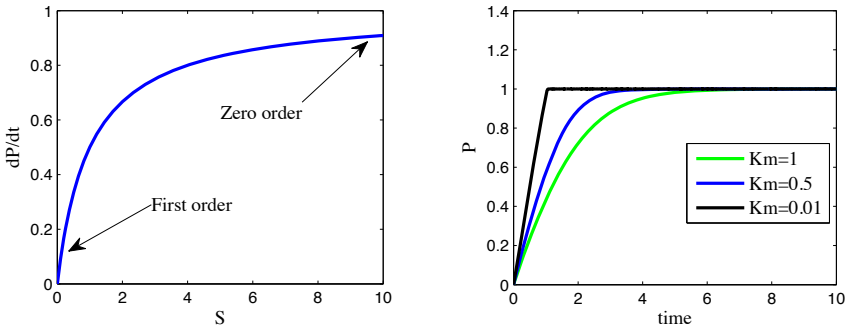
in which the constant K_m is called the *Michaelis-Menten constant*. Letting $V_{\max} = kE_{\text{tot}}$, the resulting kinetics

$$\frac{dP}{dt} = k\frac{E_{\text{tot}}S}{S + K_m} = V_{\max}\frac{S}{S + K_m} \tag{2.9}$$

are called *Michaelis-Menten kinetics*.

The constant V_{max} is called the maximal velocity (or maximal flux) of modification and it represents the maximal rate that can be obtained when the enzyme is completely saturated by the substrate. The value of K_m corresponds to the value of S that leads to a half-maximal value of the production rate of P. When the enzyme complex can be neglected with respect to the total substrate amount S_{tot}, we have that $S_{\text{tot}} = S + P + C \approx S + P$, so that the above equation can be also re-written as

$$\frac{dP}{dt} = \frac{V_{max}(S_{tot} - P)}{(S_{tot} - P) + K_m}.$$



(a) Production rate (b) Time plots

Figure 2.5: Enzymatic reactions. (a) Transfer curve showing the production rate for P as a function of substrate concentration for $K_m = 1$. (b) Time plots of product $P(t)$ for different values of the K_m. In the plots $S_{\text{tot}} = 1$ and $V_{max} = 1$.

When $K_m \ll S_{\text{tot}}$ and the substrate has not yet been all converted to product, that is, $S \gg K_m$, we have that the rate of product formation becomes approximately $dP/dt \approx V_{max}$, which is the maximal speed of reaction. Since this rate is constant and does not depend on the reactant concentrations, it is usually referred to as *zero-order kinetics*. In this case, the system is said to operate in the zero-order regime. If instead $S \ll K_m$, the rate of product formation becomes $dP/dt \approx V_{max}/K_m S$, which is linear with the substrate concentration S. This production rate is referred to as *first-order kinetics* and the system is said to operate in the first order regime (see Figure 2.5).

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 central dogma: production of proteins

The genetic material inside a cell, encoded in its DNA, governs the response of a cell to various conditions. DNA is organized into collections of genes, with each gene encoding a corresponding protein that performs a set of functions in the cell.

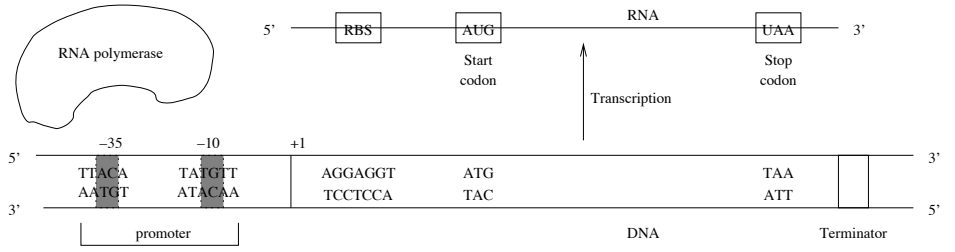


Figure 2.6: Geometric structure of DNA. The layout of the DNA is shown at the top. RNA polymerase binds to the promoter region of the DNA and transcribes the DNA starting at the +1 side and continuing to the termination site. The transcribed mRNA strand has the ribosome binding site (RBS) where the ribosomes bind, the start codon where translation starts and the stop codon where translation ends.

The activation and repression of genes are determined through a series of complex interactions that give rise to a remarkable set of circuits that perform the functions required for life, ranging from basic metabolism to locomotion to procreation. Genetic circuits that occur in nature are robust to external disturbances and can function in a variety of conditions. To understand how these processes occur (and some of the dynamics that govern their behavior), it will be useful to present a relatively detailed description of the underlying biochemistry involved in the production of proteins.

DNA is a double stranded molecule with the "direction" of each strand specified by looking at the geometry of the sugars that make up its backbone. The complementary strands of DNA are composed of a sequence of nucleotides that consist of a sugar molecule (deoxyribose) bound to one of 4 bases: adenine (A), cytocine (C), guanine (G) and thymine (T). The coding strand (by convention the top row of a DNA sequence when it is written in text form) is specified from the 5' end of the DNA to the 3' end of the DNA. (The 5' and 3' refer to carbon locations on the deoxyribose backbone that are involved in linking together the nucleotides that make up DNA.) The DNA that encodes proteins consists of a promoter region, regulator regions (described in more detail below), a coding region and a termination region (see Figure 2.6). We informally refer to this entire sequence of DNA as a gene.

Expression of a gene begins with the *transcription* of DNA into mRNA by RNA polymerase, as illustrated in Figure 2.7. RNA polymerase enzymes are present in the nucleus (for eukaryotes) or cytoplasm (for prokaryotes) and must localize and bind to the promoter region of the DNA template. Once bound, the RNA polymerase "opens" the double stranded DNA to expose the nucleotides that make up the sequence. This reaction, called *isomerization*, is said to transform the RNA polymerase and DNA from a *closed complex* to an *open complex*. After the open complex is formed, RNA polymerase begins to travel down the DNA strand and constructs an mRNA sequence that matches the 5' to 3' sequence of the DNA to which it is bound. By convention, we number the first base pair that is transcribed

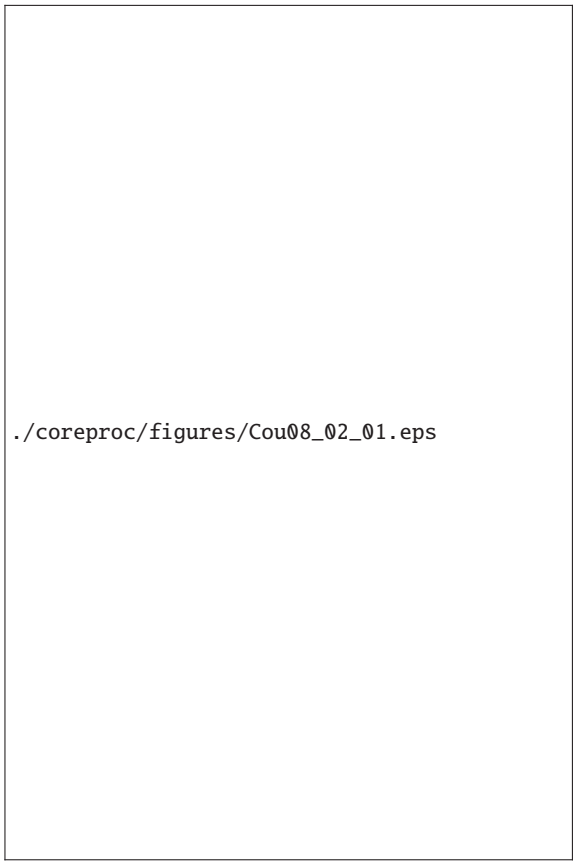


Figure 2.7: Production of messenger RNA from DNA. RNA polymerase, along with other accessory factors, binds to the promoter region of the DNA and then "opens" the DNA to begin transcription (initiation). As RNA polymerase moves down the DNA in the transcription elongation complex (TEC), it produces an RNA transcript (elongation), which is later translated into a protein. The process ends when the RNA polymerase reaches the terminator (termination). Reproduced from Courey [20]; permission pending.

as '+1' and the base pair prior to that (which is not transcribed) is labeled as '-1'. The promoter region is often shown with the -10 and -35 regions indicated, since these regions contain the nucleotide sequences to which the RNA polymerase enzyme binds (the locations vary in different cell types, but these two numbers are typically used).

The RNA strand that is produced by RNA polymerase is also a sequence of nucleotides with a sugar backbone. The sugar for RNA is ribose instead of deoxyribose and mRNA typically exists as a single stranded molecule. Another difference is that the base thymine (T) is replaced by uracil (U) in RNA sequences. RNA polymerase produces RNA one base pair at a time, as it moves from in the 5' to 3' direction along the DNA coding strand. RNA polymerase stops transcribing DNA

when it reaches a *termination region* (or *terminator*) on the DNA. This termination region consists of a sequence that causes the RNA polymerase to unbind from the DNA. The sequence is not conserved across species and in many cells the termination sequence is sometimes "leaky", so that transcription will occasionally occur across the terminator.

Once the mRNA is produced, it must be translated into a protein. This process is slightly different in prokaryotes and eukaryotes. In prokaryotes, there is a region of the mRNA in which the ribosome (a molecular complex consisting of of both proteins and RNA) binds. This region, called the *ribosome binding site (RBS)*, has some variability between different cell species and between different genes in a given cell. The Shine-Delgarno sequence, AGGAGG, is the consensus sequence for the RBS. (A consensus sequence is a pattern of nucleotides that implements a given function across multiple organisms; it is not exactly conserved, so some variations in the sequence will be present from one organism to another.)

In eukaryotes, the RNA must undergo several additional steps before it is translated. The RNA sequence that has been created by RNA polymerase consists of *introns* that must be spliced out of the RNA (by a molecular complex called the spliceosome), leaving only the *exons*, which contain the coding sequence for the protein. The term *pre-mRNA* is often used to distinguish between the raw transcript and the spliced mRNA sequence, which is called *mature mRNA*. In addition to splicing, the mRNA is also modified to contain a *poly(A)* (polyadenine) *tail*, consisting of a long sequence of adenine (A) nucleotides on the 3' end of the mRNA. This processed sequence is then transported out of the nucleus into the cytoplasm, where the ribosomes can bind to it.

Unlike prokaryotes, eukaryotes do not have a well defined ribosome binding sequence and hence the process of the binding of the ribosome to the mRNA is more complicated. The *Kozak sequence* A/GCCACCAUGG is the rough equivalent of the ribosome binding site, where the underlined AUG is the start codon (described below). However, mRNA lacking the Kozak sequence can also be translated.

Once the ribosome is bound to the mRNA, it begins the process of *translation*. Proteins consist of a sequence of amino acids, with each amino acid specified by a codon that is used by the ribosome in the process of translation. Each codon consists of three base pairs and corresponds to one of the 20 amino acids or a "stop" codon. The ribosome translates each codon into the corresponding amino acid using transfer RNA (tRNA) to integrate the appropriate amino acid (which binds to the tRNA) into the polypeptide chain, as shown in Figure 2.8. The start codon (AUG) specifies the location at which translation begins, as well as coding for the amino acid methionine (a modified form is used in prokaryotes). All subsequent codons are translated by the ribosome into the corresponding amino acid until it reaches one of the stop codons (typically UAA, UAG and UGA).

The sequence of amino acids produced by the ribosome is a polypeptide chain that folds on itself to form a protein. The process of folding is complicated and

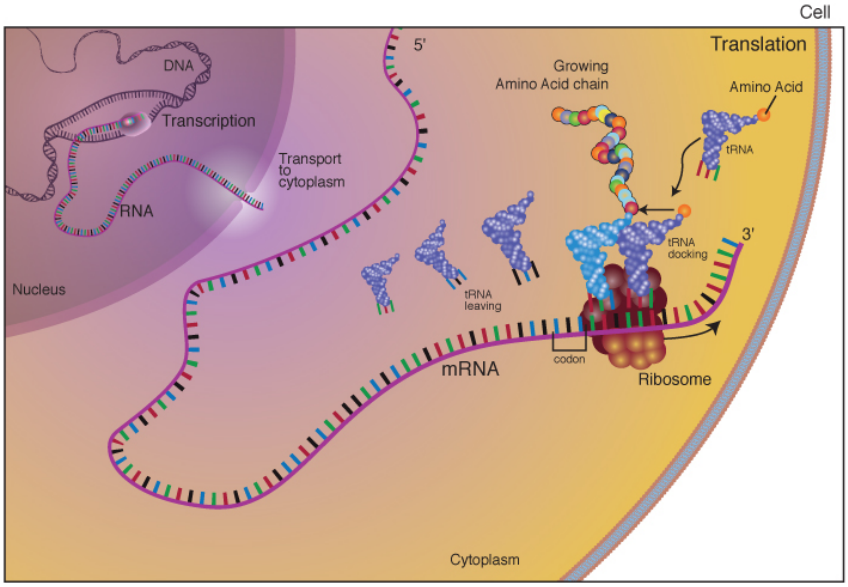


Figure 2.8: Translation is the process of translating the sequence of a messenger RNA (mRNA) molecule to a sequence of amino acids during protein synthesis. The genetic code describes the relationship between the sequence of base pairs in a gene and the corresponding amino acid sequence that it encodes. In the cell cytoplasm, the ribosome reads the sequence of the mRNA in groups of three bases to assemble the protein. Figure and caption courtesy the National Human Genome Research Institute.

involves a variety of chemical interactions that are not completely understood. Additional post-translational processing of the protein can also occur at this stage, until a folded and functional protein is produced. It is this molecule that is able to bind to other species in the cell and perform the chemical reactions that underly the behavior of the organism. The *maturation time* of a protein is the time required for the polypeptide chain to fold into a functional protein.

Each of the processes involved in transcription, translation and folding of the protein takes time and affects the dynamics of the cell. Table 2.1 shows representative rates of some of the key processes involved in the production of proteins. In particular, the dissociation constant of RNA polymerase from the DNA promoter has a wide range of values depending on whether the binding is enhanced by activators (as we will see in the sequel), in which case it can take very low values. Similarly, the dissociation constant of transcription factors with DNA can be very low in the case of specific binding and substantially larger for non-specific binding. It is important to note that each of these steps is highly stochastic, with molecules binding together based on some propensity that depends on the binding energy but also the other molecules present in the cell. In addition, although we have described everything as a sequential process, each of the steps of tran-

Table 2.1: Rates of core processes involved in the creation of proteins from DNA in *E. coli*.

Process	Characteristic rate	Source
mRNA transcription rate	24-29 bp/sec	[13]
Protein translation rate	12–21 aa/sec	[13]
Maturation time (fluorescent proteins)	6–60 min	[13]
mRNA half life	~ 100 sec	[100]
E. coli cell division time	20–40 min	[13]
Yeast cell division time	70–140 min	[13]
Protein half life	$\sim 5 \times 10^4$ sec	[100]
Protein diffusion along DNA	up to 10^4 bp/sec	[75]
RNA polymerase dissociation constant	$\sim$ 0.3–10,000 nM	[13]
Open complex formation kinetic rate	~ 0.02 sec^{-1}	[13]
Transcription factor dissociation constant	$\sim$ 0.02–10,000 nM	[13]

scription, translation and folding are happening simultaneously. In fact, there can be multiple RNA polymerases that are bound to the DNA, each producing a transcript. In prokaryotes, as soon as the ribosome binding site has been transcribed, the ribosome can bind and begin translation. It is also possible to have multiple ribosomes bound to a single piece of mRNA. Hence the overall process can be extremely stochastic and asynchronous.

Reaction models

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, $\mathrm{RNAP}^{\mathrm{p}}$ represents RNA polymerase in the promoter region, and $\mathrm{RNAP}^{\mathrm{d}}$ is RNA polymerase non-specifically bound to 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, $\mathrm{DNA}^{\mathrm{i}}$ for the ith section of the gene of interest and $\mathrm{DNA}^{\mathrm{t}}$ for the termination sequence. We write RNAP:DNA to represent RNA polymerase bound to DNA (assumed closed) and RNAP:$\mathrm{DNA}^{\mathrm{o}}$ 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

$$
\begin{aligned}
\text{Binding to DNA:} \quad & \mathrm{RNAP^c} \rightleftharpoons \mathrm{RNAP^d} \\
\text{Diffusion along DNA:} \quad & \mathrm{RNAP^d} \rightleftharpoons \mathrm{RNAP^p} \\
\text{Binding to promoter:} \quad & \mathrm{RNAP^p} + \mathrm{DNA^p} \rightleftharpoons \mathrm{RNAP : DNA^p} \\
\text{Isomerization:} \quad & \mathrm{RNAP : DNA^p} \longrightarrow \mathrm{RNAP : DNA^o} \\
\text{Start of transcription:} \quad & \mathrm{RNAP : DNA^o} \longrightarrow \mathrm{RNAP : DNA^1} + \mathrm{DNA^p} \\
\text{mRNA creation:} \quad & \mathrm{RNAP : DNA^1} \longrightarrow \mathrm{RNAP : DNA^2 : mRNA^1} \\
\text{Elongation:} \quad & \mathrm{RNAP : DNA^{i+1} : mRNA^i} \\
& \qquad \longrightarrow \mathrm{RNAP : DNA^{i+2} : mRNA^{i+1}} \\
\text{Binding to terminator:} \quad & \mathrm{RNAP{:}DNA^N : mRNA^{N-1}} \\
& \qquad \longrightarrow \mathrm{RNAP : DNA^t} + \mathrm{mRNA^N} \\
\text{Termination:} \quad & \mathrm{RNAP : DNA^t} \longrightarrow \mathrm{RNAP^c} \\
\text{Degradation:} \quad & \mathrm{mRNA^N} \longrightarrow \emptyset.
\end{aligned}
\tag{2.10}
$$

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. This allows the same DNA strand to be transcribed by multiple RNA polymerase at the same time. The species $\mathrm{RNAP : DNA^{i+1} : mRNA^i}$ represents RNA polymerases bound at the $(i+1)$th section of DNA with an elongating mRNA strand of length i attached to it. Upon binding to the terminator region, the RNA polymerase releases the full mRNA strand $\mathrm{mRNA^N}$. This mRNA has the ribosome binding site at which ribosomes can bind to start translation. The main difference between prokaryotes and eukaryotes is that in eukaryotes the RNA polymerase remains in the nucleus and the $\mathrm{mRNA^N}$ must be spliced and transported to the cytoplasm before ribosomes can start translation. As a consequence, the start of translation can occur only after $\mathrm{mRNA^N}$ has been produced. For simplicity of notation, we assume here that the entire mRNA strand should be produced before ribosomes can start translation. In the procaryotic case, instead, translation can start even for an mRNA strand that is still elongating (see Exercise 2.6).

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 ribosome binding site (RBS) of $\mathrm{mRNA^N}$, translation of the mRNA sequence into a polypeptide chain, and folding of the polypeptide chain into a functional protein. Specifically, we must keep track of the various states of the ribosome bound to different codons on the mRNA strand. We thus let $\mathrm{Ribo : mRNA^{RBS}}$ denote the ribosome bound to the ribosome binding site of $\mathrm{mRNA^N}$, $\mathrm{Ribo : mRNA^{AAi}}$ the ribosome bound to the ith codon (corresponding to an amino acid, indicated by the superscript AA),

$\text{Ribo} : \text{mRNA}^{\text{start}}$ and $\text{Ribo} : \text{mRNA}^{\text{stop}}$ the ribosome bound to the start and stop codon, respectively. We also let PPC^{i} denote the polypeptide chain consisting of i amino acids. Here, we assume that the protein of interest has M amino acids. The reactions describing translation can then be written as

$$\begin{aligned}
\text{Binding to RBS:} \quad & \text{Ribo} + \text{mRNA}^{\text{N}} \rightleftharpoons \text{Ribo} : \text{mRNA}^{\text{RBS}} \\
\text{Start of translation:} \quad & \text{Ribo} : \text{mRNA}^{\text{RBS}} \longrightarrow \text{Ribo} : \text{mRNA}^{\text{start}} + \text{mRNA}^{\text{N}} \\
\text{Polypeptide chain creation:} \quad & \text{Ribo} : \text{mRNA}^{\text{start}} \longrightarrow \text{Ribo} : \text{mRNA}^{\text{AA2}} : \text{PPC}^{1} \\
\text{Elongation, } i = 1, \ldots, M\text{:} \quad & \text{Ribo} : \text{mRNA}^{\text{AA(i+1)}} : \text{PPC}^{\text{i}} \\
& \qquad \longrightarrow \text{Ribo} : \text{mRNA}^{\text{AA(i+2)}} : \text{PPC}^{\text{i+1}} \\
\text{Stop codon:} \quad & \text{Ribo} : \text{mRNA}^{\text{AAM}} : \text{PPC}^{\text{M}-1} \\
& \qquad \longrightarrow \text{Ribo} : \text{mRNA}^{\text{stop}} + \text{PPC}^{\text{M}} \\
\text{Release of mRNA:} \quad & \text{Ribo} : \text{mRNA}^{\text{stop}} \longrightarrow \text{Ribo} \\
\text{Folding:} \quad & \text{PPC}^{\text{M}} \longrightarrow \text{protein} \\
\text{Degradation:} \quad & \text{protein} \longrightarrow \emptyset.
\end{aligned} \tag{2.11}$$

As in the case of transcription, we see that these reactions allow multiple ribosomes to translate the same piece of mRNA by freeing up mRNA^{N}. After M amino acids have been chained together, the M-long polypeptide chain PPC^{M} is released, which then folds into a protein. As complex as these reactions are, they do not directly capture a number of physical phenomena such as ribosome queuing, wherein ribosomes cannot pass other ribosomes that are ahead of them on the mRNA chain. Additionally, 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. For more detailed models of translation, the reader is referred to [3].

When the details of the isomerization, start of transcription (translation), elongation, and termination are not relevant for the phenomenon to be studied, the transcription and translation reactions are lumped into much simpler reduced reactions. For transcription, these reduced reactions take the form:

$$\begin{aligned}
& \text{RNAP} + \text{DNA}^{\text{p}} \rightleftharpoons \text{RNAP:DNA}^{\text{p}} \\
& \text{RNAP:DNA}^{\text{p}} \longrightarrow \text{mRNA} + \text{RNAP} + \text{DNA}^{\text{p}} \\
& \text{mRNA} \longrightarrow \emptyset,
\end{aligned} \tag{2.12}$$

in which the second reaction lumps together isomerization, start of transcription,

elongation, mRNA creation, and termination. Similarly, for the translation process, the reduced reactions take the form:

$$\begin{aligned} &\text{Ribo} + \text{mRNA} \rightleftharpoons \text{Ribo:mRNA} \\ &\text{Ribo:mRNA} \longrightarrow \text{protein} + \text{mRNA} + \text{Ribo} \\ &\text{Ribo:mRNA} \longrightarrow \text{Ribo} \\ &\text{protein} \longrightarrow \emptyset, \end{aligned} \tag{2.13}$$

in which the second reaction lumps the start of translation, elongation, folding, and termination. The third reaction models the fact that mRNA can also be degraded when bound to ribosomes when the ribosome binding site is left free. The process of mRNA degradation occurs through RNAse enzymes binding to the ribosome binding site and cleaving the mRNA strand. It is known that the ribosome binding site cannot be both bound to the ribosome and to the RNase [65]. However, the species Ribo:mRNA is a lumped species encompassing configurations in which ribosomes are bound on the mRNA strand but not on the ribosome binding site. Hence, we also let this species be degraded by RNase.

Reaction rate equations

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 (though this may not be accurate for some species that exist at low molecular counts in the cell). Despite these approximations, in many situations the reaction rate equations are 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. Let the "active" mRNA be the mRNA that is available for translation by the ribosome. We model its concentration through a simple time delay of length τ^m that accounts for the transcription of the ribosome binding site in prokaryotes or splicing and transport from the nucleus in eukaryotes. 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

$$\frac{dm_P}{dt} = \alpha - \mu m_P - \bar{\delta} m_P, \qquad m_P^*(t) = e^{-\mu\tau^m} m_P(t-\tau^m), \tag{2.14}$$

where m_P is the concentration of mRNA for protein P, m_P^* is the concentration of active mRNA, α 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 $\bar{\delta}$ 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 $\delta = \mu + \bar{\delta}$. The exponential factor in the second expression in equation (2.14) accounts for dilution due to the change in volume of the cell, where μ is the cell growth rate. The constants α and δ capture the average rates of production and decay, 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 that describes the production of a functional protein:

$$\frac{dP}{dt} = \kappa m_P^* - \gamma P, \qquad P^f(t) = e^{-\mu\tau^f} P(t-\tau^f). \tag{2.15}$$

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 κ, the rate of translation of mRNA; γ, the rate of degradation and dilution of P; and τ^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 γ, captures both the 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

$$\begin{aligned} \frac{dm_P^*(t)}{dt} &= e^{-\mu\tau^m} \frac{dm_P}{dt}(t-\tau^m) \\ &= e^{-\mu\tau^m}(\alpha - \delta m_P(t-\tau^m)) = \overline{\alpha} - \delta m_P^*(t), \end{aligned} \tag{2.16}$$

where $\overline{\alpha} = e^{-\mu\tau^m}\alpha$. A similar expansion for the active protein dynamics yields

$$\frac{dP^f(t)}{dt} = \overline{\kappa} m_P^*(t-\tau^f) - \gamma P^f(t), \tag{2.17}$$

where $\overline{\kappa} = e^{-\mu\tau^f}\kappa$. We shall typically use equations (2.16) and (2.17) as our (reduced) description of protein folding, dropping the superscript f and overbars

when there is no risk of confusion. Also, in the presence of different proteins, we will attach subscripts to the parameters to denote the protein to which they refer.

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 even further and write

$$\frac{dm_P}{dt} = \alpha - \delta m_P, \qquad \frac{dP}{dt} = \kappa m_P - \gamma P. \tag{2.18}$$

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 κ since we are assuming the time delays are negligible. The value of α increases with the strength of the promoter while the value of κ increases with the strength of the ribosome binding site. These strengths, in turn, can be affected by changing the specific base pair sequences that constitute the promoter RNA polymerase binding region and the ribosome binding site.

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 (2.18) with its equilibrium value. This is often a good assumption as mRNA degration is usually about 100 times faster than protein degradation (see Table 2.1). Thus we obtain

$$\frac{dP}{dt} = \beta - \gamma P, \qquad \beta := \kappa \frac{\alpha}{\delta}.$$

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 some cases it is too simple to capture the observed behavior of a biological circuit.

2.3 Transcriptional Regulation

The operation of a cell is governed in part 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 its many metabolic pathways, 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 discussed in Section 1.2 and illustrated in

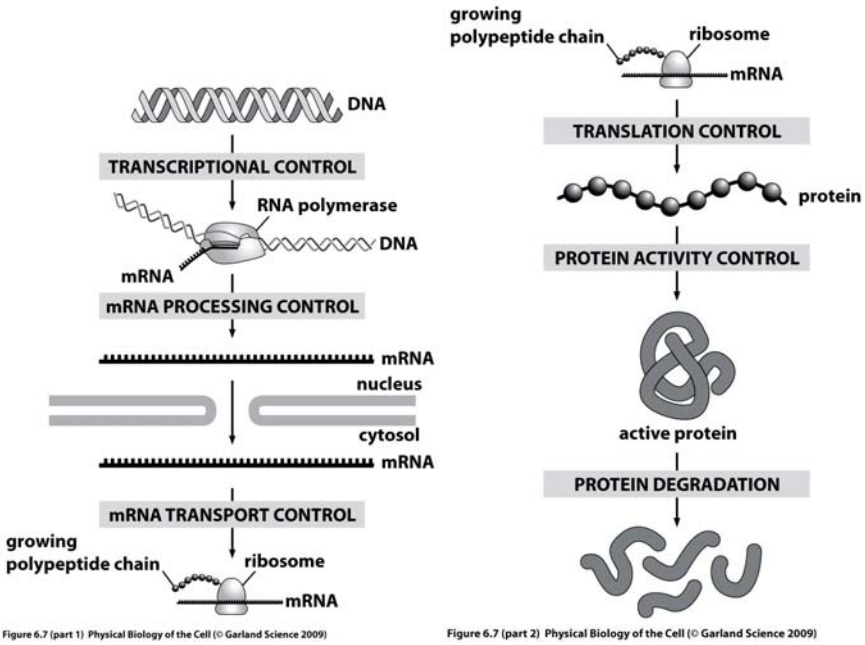


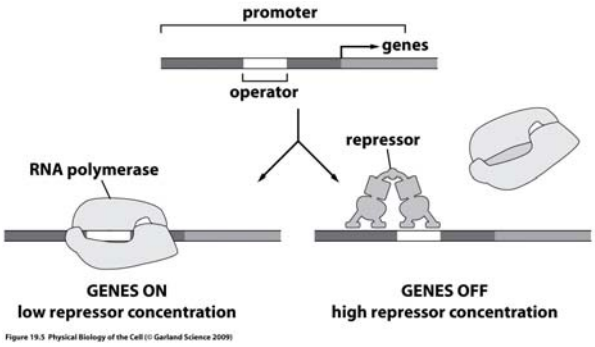
Figure 2.9: Regulation of proteins. Transcriptional control includes mechanisms to tune the rate at which mRNA is produced from DNA, while translation control includes mechanisms to tune the rate at which the protein polypeptide chain is produced from mRNA. Protein activity control encompasses many processes, such as phosphorylation, methylation, and allosteric modification. Figure from Phillips, Kondev and Theriot [75]; used with permission of Garland Science.

Figure 2.9. 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 and the next section on selected mechanisms for controlling protein activity. We will focus on prokaryotic mechanisms.

Transcriptional regulation of protein production

The simplest forms of transcriptional regulation are repression and activation, which are controlled through proteins called *transcription factors*. In the case of *repression*, the presence of a transcription factor (often a protein that binds near the promoter) turns off the transcription of the gene and this type of regulation is often called negative regulation or "down regulation". In the case of *activation* (or positive regulation), transcription is enhanced when an activator protein binds to the promoter site (facilitating binding of the RNA polymerase).

Represession. A common mechanism for repression is that a protein binds to a region of DNA near the promoter and blocks RNA polymerase from binding. The region of DNA to which the repressor protein binds is called an *operator region*



(a) Repression of gene expression

./coreproc/figures/PKT08_19_06.eps

(b) Examples of repressors

Figure 2.10: Repression of gene expression. A repressor protein binds to operator sites on the gene promoter and blocks the binding of RNA polymerase to the promoter, so that the gene is OFF. Figure from Phillips, Kondev and Theriot [75]; used with permission of Garland Science.

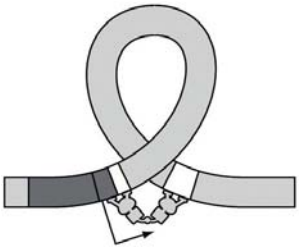
(see Figure 2.10a). If the operator region overlaps the promoter, then the presence of a protein at the promoter can "block" the DNA at that location and transcription cannot initiate. Repressor proteins often bind to DNA as dimers or pairs of dimers (effectively tetramers). Figure 2.10b shows some examples of repressors bound to DNA.

A related mechanism for repression is *DNA looping*. In this setting, two repressor complexes (often dimers) bind in different locations on the DNA and then bind to each other. This can create a loop in the DNA and block the ability of RNA polymerase to bind to the promoter, thus inhibiting transcription. Figure 2.11 shows an example of this type of repression, in the *lac* operon. (An *operon* is a set of genes that is under control of a single promoter.)

Activation. The process of activation of a gene requires that an activator protein be present in order for transcription to occur. In this case, the protein must work to either recruit or enable RNA polymerase to begin transcription.

The simplest form of activation involves a protein binding to the DNA near the promoter in such a way that the combination of the activator and the promoter sequence bind RNA polymerase. Figure 2.12 illustrates the basic concept.

Another mechanism for activation of transcription, specific to prokaryotes, is the use of *sigma factors*. Sigma factors are part of a modular set of proteins that bind to RNA polymerase and form the molecular complex that performs transcrip-



(a) DNA looping

./coreproc/figures/PKT08_08_19.eps

(b) *lac* repressor

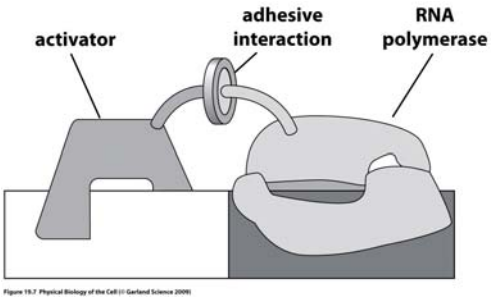
Figure 2.11: Repression via DNA looping. A repressor protein can bind simultaneously to two DNA sites downstream of the start of transcription, thus creating a loop that prevents RNA polymerase from transcribing the gene. Figure from Phillips, Kondev and Theriot [75]; used with permission of Garland Science.

tion. Different sigma factors enable RNA polymerase to bind to different promoters, so the sigma factor acts as a type of activating signal for transcription. Table 2.2 lists some of the common sigma factors in bacteria. One of the uses of sigma factors is to produce certain proteins only under special conditions, such as when the cell undergoes *heat shock*. Another use is to control the timing of the expression of certain genes, as illustrated in Figure 2.13.

Inducers. A feature that is present in some types of transcription factors is the existence of an *inducer molecule* that combines with the protein to either activate or inactivate its function. A *positive inducer* is a molecule that must be present in order for repression or activation to occur. A *negative inducer* is one in which the presence of the inducer molecule blocks repression or activation, either by changing the shape of the transcription factor protein or by blocking active sites on the protein that would normally bind to the DNA. Figure 2.14a summarizes the various possibilities. Common examples of repressor-inducer pairs include *lacI* and lactose (or IPTG), and *tetR* and aTc. Lactose/IPTG and aTc are both negative inducers, so their presence causes the otherwise repressed gene to be expressed. An example of a positive inducer is cyclic AMP (cAMP), which acts as a positive inducer for the CAP activator.

Table 2.2: Sigma factors in *E. coli* [2].

Sigma factor	Promoters recognized
σ^{70}	most genes
σ^{32}	genes associated with heat shock
σ^{38}	genes involved in stationary phase and stress response
σ^{28}	genes involved in motility and chemotaxis
σ^{24}	genes dealing with misfolded proteins in the periplasm



./coreproc/figures/PKT08_19_08.eps

(a) Activation mechanism (b) Examples of activators

Figure 2.12: Activation of gene expression. (a) Conceptual operation of an activator. The activator binds to DNA upstream of the gene and attracts RNA polymerase to the DNA strand. (b) Examples of activators: catabolite activator protein (CAP), p53 tumor suppressor, zinc finger DNA binding domain and leucine zipper DAN binding domain. Figure from Phillips, Kondev and Theriot [75]; used with permission of Garland Science.

Combinatorial promoters. In addition to promoters that can take either a repressor or an activator as the sole input transcription factor, there are *combinatorial promoters* that can take both repressors and activators as input transcription factors. This allows genes to be switched on and off based on more complex conditions, represented by the concentrations of two or more activators or repressors.

Figure 2.15 shows one of the classic examples, a promoter for the *lac* system. In the *lac* system, the expression of genes for metabolizing lactose are under the control of a single (combinatorial) promoter. CAP, which is positively induced by cAMP, acts as an activator and LacI (also called "Lac repressor"), which is negatively induced by lactose, acts as a repressor. In addition, the inducer cAMP is expressed only when glucose levels are low. The resulting behavior is that the proteins for metabolizing lactose are expressed only in conditions where there is no glucose (so CAP is active) *and* lactose is present.

More complicated combinatorial promoters can also be used to control transcription in two different directions, an example that is found in some viruses.

Antitermination. A final method of activation in prokaryotes is the use of *antitermination*. The basic mechanism involves a protein that binds to DNA and deactivates a site that would normally serve as a termination site for RNA polymerase. Additional genes are located downstream from the termination site, but without a promoter region. Thus, in the absence of the anti-terminator protein, these genes are not expressed (or expressed with low probability). However, when the antitermination protein is present, the RNA polymerase maintains (or regains) its contact

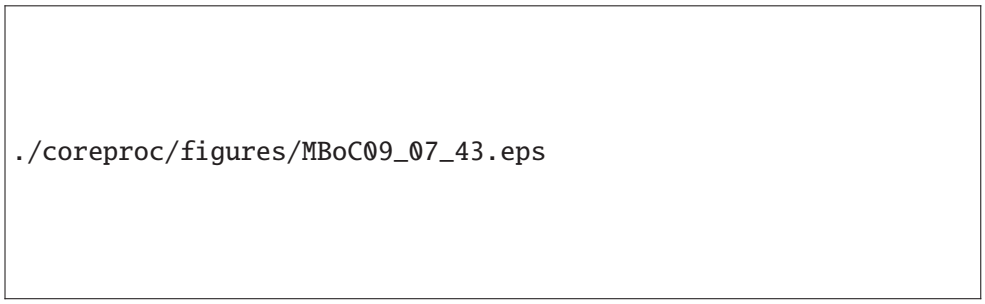


Figure 2.13: Use of sigma factors to control the timing of gene expression in a bacterial virus. Early genes are transcribed by RNA polymerase bound to bacterial sigma factors. One of the early genes, called 28, encodes a sigma-like factor that binds to RNA polymerase and allow it to transcribe middle genes, which in turn produce another sigma-like factor that allows RNA polymerase to transcribe late genes. These late genes produce proteins that form a coat for the viral DNA and lyse the cell. Reproduced from Alberts et al. [2]; permission pending.

Figure 2.14: Effects of inducers. (a) In the case of repressors, a negative inducer binds to the repressor making it unbind DNA, thus enabling transcription. A positive inducer, by contrast, activates the repressor allowing it to bind DNA. (b) In the case of activators, a negative inducer binds to the activator making it unbind DNA, thus preventing transcription. A positive inducer, instead enables the activator to bind DNA, allowing transcription. Reproduced from Alberts et al. [2]; permission pending.

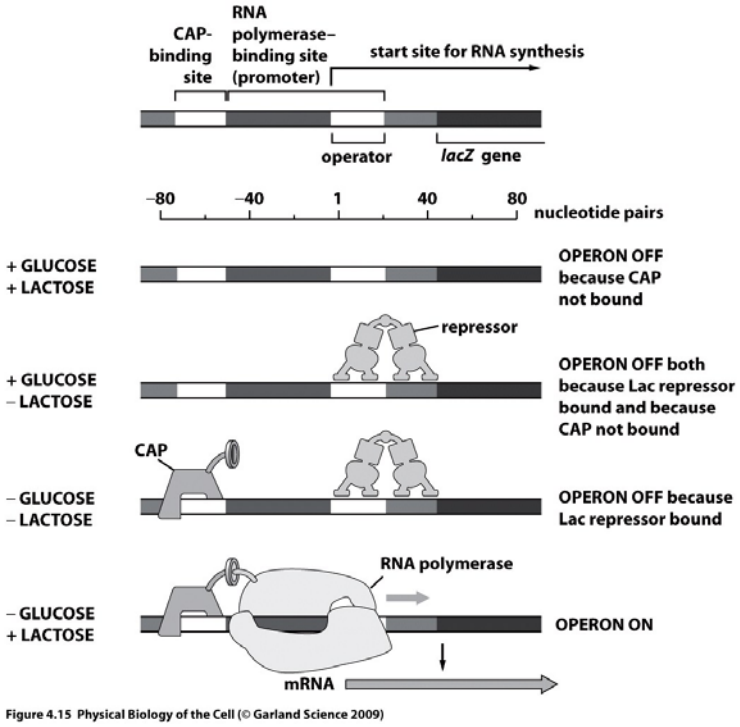


Figure 2.15: Combinatorial logic for the *lac* operator. The CAP-binding site and the operator in the promoter can be both bound by CAP (activator) and by LacI (Lac repressor), respectively. The only configuration in which RNA polymerase can bind the promoter and start transcription is where CAP is bound but LacI is not bound. Figure from Phillips, Kondev and Theriot [75]; used with permission of Garland Science.

with the DNA and expression of the downstream genes is enhanced. In this way, antitermination allows downstream genes to be regulated by repressing "premature" termination. An example of an antitermination protein is the protein N in phage λ, which binds to a region of DNA labeled nut (for N utilization), as shown in Figure 2.16 [37].

```
./coreproc/figures/GNM93-antitermination.eps
```

Figure 2.16: Antitermination. Protein N binds to DNA regions labeled nut enabling transcription of longer DNA sequences. Reproduced from [37]; permission pending.

Reaction models

We can capture the molecular interactions responsible for transcriptional regulation by modifying the RNA polymerase binding reactions in equation (2.10). For a repressor (Rep), we simply have to add a reaction that represents the repressor bound to the promoter $\mathrm{DNA}^{\mathrm{p}}$:

$$\text{Repressor binding:} \quad \mathrm{DNA}^{\mathrm{p}} + \mathrm{Rep} \rightleftharpoons \mathrm{DNA{:}Rep}.$$

This reaction acts to "sequester" the DNA promoter site so that it is no longer available for binding by RNA polymerase. The strength of the repressor is reflected in the reaction rate constants for the repressor binding reaction. Sometimes, the RNA polymerase can bind to the promoter even when the repressor is bound, usually with lower association rate constant. In this case, the repressor still allows some transcription even when bound to the promoter and the repressor is said to be "leaky".

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 the promoter. One possible mechanism, known as the *recruitment model*, is given by

$$\begin{aligned}
\text{Activator binding:} \quad & \mathrm{DNA}^{\mathrm{p}} + \mathrm{Act} \rightleftharpoons \mathrm{DNA}^{\mathrm{p}}{:}\mathrm{Act} \\
\text{RNAP binding w/ activator:} \quad & \mathrm{RNAP}^{\mathrm{p}} + \mathrm{DNA}^{\mathrm{p}}{:}\mathrm{Act} \rightleftharpoons \mathrm{RNAP{:}DNA}^{\mathrm{p}}{:}\mathrm{Act} \\
\text{Isomerization:} \quad & \mathrm{RNAP{:}DNA}^{\mathrm{p}}{:}\mathrm{Act} \longrightarrow \mathrm{RNAP{:}DNA}^{\mathrm{o}}{:}\mathrm{Act} \\
\text{Start of transcription:} \quad & \mathrm{RNAP{:}DNA}^{\mathrm{o}}{:}\mathrm{Act} \longrightarrow \mathrm{RNAP{:}DNA}^{1} + \mathrm{DNA}^{\mathrm{p}}{:}\mathrm{Act}.
\end{aligned} \tag{2.19}$$

In this model, RNA polymerase cannot bind to the promoter unless the activator is already bound to it. More generally, one can 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. This translates into the additional reaction $\mathrm{RNAP}^{\mathrm{p}} + \mathrm{DNA}^{\mathrm{p}} \rightleftharpoons \mathrm{RNAP{:}DNA}^{\mathrm{p}}$. The relative reaction rates determine how strong the activator is and the "leakiness" of transcription in the absence of the activator. A different model of activation, called *allosteric activation*, is one in which the RNA polymerase binding rate to DNA is not enhanced by the presence of the activator bound to the promoter, but the open complex (and hence start of transcription) formation can occur only (is enhanced) in the presence of the activator.

A simplified ordinary differential equation model of transcription in the presence of activators or repressors can be obtained by accounting for the fact that transcription factors and RNAP bind to the DNA rapidly when compared to other reactions, such as isomerization and elongation. As a consequence, we can make use of the reduced order models that describe the quasi-steady state concentrations of proteins bound to DNA as described in Section 2.1. We can consider the competitive binding case to model a strong repressor that prevents RNA polymerase

from binding to the DNA. In the sequel, we remove the superscripts "p" and "d" from RNA polymerase to simplify notation. The quasi-steady state concentration of the complex of DNA promoter bound to the repressor will have the expression

$$[\text{DNA}^{\text{p}}\text{:Rep}] = \frac{[\text{DNA}]_{\text{tot}}([\text{Rep}]/K_{\text{d}})}{1 + [\text{Rep}]/K_{\text{d}} + [\text{RNAP}]/K'_{\text{d}}}$$

and the steady state amount of DNA promoter bound to the RNA polymerase will be given by

$$[\text{RNAP:DNA}^{\text{p}}] = \frac{[\text{DNA}]_{\text{tot}}([\text{RNAP}]/K'_{\text{d}})}{1 + [\text{RNAP}]/K'_{\text{d}} + [\text{Rep}]/K_{\text{d}}},$$

in which K'_{d} is the dissociation constant of RNA polymerase from the promoter, while K_{d} is the dissociation constant of Rep from the promoter, and $[\text{DNA}]_{\text{tot}}$ represents the total concentration of DNA. The free DNA promoter with RNA polymerase bound will allow transcription, while the complex DNA^{p}:Rep will not allow transcription as it is not bound to RNA polymerase. Using the lumped reactions (2.12), this can be modeled as

$$\frac{d[\text{mRNA}]}{dt} = F([\text{Rep}]) - \delta[\text{mRNA}],$$

in which the production rate is given by

$$F([\text{Rep}]) = k_{\text{f}} \frac{[\text{DNA}]_{\text{tot}}\,([\text{RNAP}]/K'_{\text{d}})}{1 + [\text{RNAP}]/K'_{\text{d}} + [\text{Rep}]/K_{\text{d}}}.$$

If the repressor binds to the promoter with cooperativity n, the above expression becomes (see Section 2.1)

$$F([\text{Rep}]) = k_{\text{f}} \frac{[\text{DNA}]_{\text{tot}}\,([\text{RNAP}]/K'_{\text{d}})}{1 + [\text{RNAP}]/K'_{\text{d}} + [\text{Rep}]^n/(K_{\text{m}}K_{\text{d}})},$$

in which K_{m} is the dissociation constant of the reaction of n molecules of Rep binding together. The function F is usually represented in the standard Hill function form

$$F([\text{Rep}]) = \frac{\alpha}{1 + ([\text{Rep}]/K)^n},$$

in which α and K are given by

$$\alpha = \frac{k_{\text{f}}[\text{DNA}]_{\text{tot}}([\text{RNAP}]/K'_{\text{d}})}{1 + ([\text{RNAP}]/K'_{\text{d}})}, \qquad K = \left(K_{\text{m}}K_{\text{d}}(1 + ([\text{RNAP}]/K'_{\text{d}})\right)^{1/n}.$$

Finally, if the repressor allows RNA polymerase to still bind to the promoter at a small rate (leaky repressor), the above expression can be modified to take the form (see Section 2.1)

$$F([\text{Rep}]) = \frac{\alpha}{1 + ([\text{Rep}]/K)^n} + \alpha_0, \tag{2.20}$$

in which α_0 is the basal expression level when the promoter is fully repressed, usually referred to as "leakiness" (see Exercise 2.8).

To model the production rate of mRNA in the case in which an activator Act is required for transcription, we can consider the case in which RNA polymerase binds only when the activator is already bound to the promoter (recruitment model). To simplify the mathematical derivation, we re-write the reactions (2.19) involving the activator with the lumped transcription reaction (2.12) into the following:

$$\begin{aligned} &\mathrm{DNA^p} + \mathrm{Act} \rightleftharpoons \mathrm{DNA^p{:}Act} \\ &\mathrm{RNAP} + \mathrm{DNA^p{:}Act} \rightleftharpoons \mathrm{RNAP{:}DNA^p{:}Act} \\ &\mathrm{RNAP{:}DNA^p{:}Act} \xrightarrow{k_f} \mathrm{mRNA} + \mathrm{RNAP} + \mathrm{DNA^p{:}Act}, \end{aligned} \tag{2.21}$$

in which the third reaction lumps together isomerization, start of transcription, elongation and termination. The first and second reactions fit the structure of the cooperative binding model illustrated in Section 2.1. Also, since the third reaction is much slower than the first two, the complex RNAP:DNA$^{\mathrm{p}}$:Act concentration can be well approximated at its quasi-steady state value. The expression of the quasi-steady state concentration was given in Section 2.1 in correspondence to the cooperative binding model and takes the form

$$[\mathrm{RNAP{:}DNA^p{:}Act}] = \frac{[\mathrm{DNA}]_{\mathrm{tot}}([\mathrm{RNAP}]/K'_{\mathrm{d}})([\mathrm{Act}])/K_{\mathrm{d}})}{1 + ([\mathrm{Act}]/K_{\mathrm{d}})(1 + [\mathrm{RNAP}]/K'_{\mathrm{d}})},$$

in which K'_{d} is the dissociation constant of RNA polymerase with the complex of DNA bound to Act and K_{d} is the dissociation constant of Act with DNA. When the activator Act binds to the promoter with cooperativity n, the above expression becomes

$$[\mathrm{RNAP{:}DNA^p{:}Act}] = \frac{[\mathrm{DNA}]_{\mathrm{tot}}([\mathrm{RNAP}][\mathrm{Act}]^n)/(K_{\mathrm{d}}K'_{\mathrm{d}}K_{\mathrm{m}})}{1 + ([\mathrm{Act}]^n/K_{\mathrm{d}}K_{\mathrm{m}})(1 + [\mathrm{RNAP}]/K'_{\mathrm{d}})},$$

in which K_{m} is the dissociation constant of the reaction of n molecules of Act binding together.

In order to write the differential equation for the mRNA concentration, we consider the third reaction in (2.21) along with the above quasi-steady state expressions of [RNAP:DNA$^{\mathrm{p}}$:Act] to obtain

$$\frac{d\,[\mathrm{mRNA}]}{dt} = F([\mathrm{Act}]) - \delta[\mathrm{mRNA}],$$

in which

$$F([\mathrm{Act}]) = k_{\mathrm{f}}\frac{[\mathrm{DNA}]_{\mathrm{tot}}([\mathrm{RNAP}][\mathrm{Act}]^n)/(K_{\mathrm{d}}K'_{\mathrm{d}}K_{\mathrm{m}})}{1 + ([\mathrm{Act}]^n/K_{\mathrm{d}}K_{\mathrm{m}})(1 + [\mathrm{RNAP}]/K'_{\mathrm{d}})} =: \frac{\alpha([\mathrm{Act}]/K)^n}{1 + ([\mathrm{Act}]/K)^n},$$

where α and K are implicitly defined. The right-hand side of this expression is in standard Hill function form. If we assume that RNA polymerase can still bind to DNA even when the activator is not bound, we have an additional basal expression rate α_0 so that the new form of the production rate is given by (see Exercise 2.9)

$$F([\text{Act}]) = \frac{\alpha([\text{Act}]/K)^n}{1 + ([\text{Act}]/K)^n} + \alpha_0. \tag{2.22}$$

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 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

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

Since the above reactions are very fast compared to transcription, they can be assumed at the quasi-steady state. Hence, the free amount of repressor that can still bind to the promoter can be calculated by writing the ODE model corresponding to the above reactions and by setting the time derivatives to zero. This yields

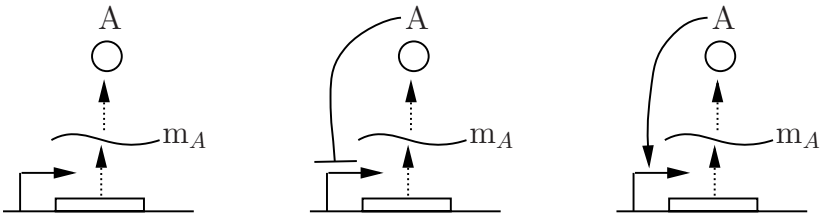
$$[\text{Rep}] = \frac{[\text{Rep}]_{\text{tot}}}{1 + [\text{Ind}]/\bar{K}_{\text{d}}},$$

in which $[\text{Rep}]_{\text{tot}} = [\text{Rep}] + [\text{Rep:Ind}]$ is the total amount of repressor (bound and unbound to the inducer) and $\bar{K}_{\text{d}}$ is the dissociation constant of Ind binding to Rep. This expression of the repressor concentration needs to be substituted in the expression of the production rate $F([\text{Rep}])$.

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

$$\begin{aligned}
\text{Inducer binding:} \quad & \text{Act} + \text{Ind} \rightleftharpoons \text{Act:Ind} \\
\text{Activator binding:} \quad & \text{DNA}^{\text{p}} + \text{Act:Ind} \rightleftharpoons \text{DNA}^{\text{p}}\text{:Act:Ind} \\
\text{RNAP binding w/ activator:} \quad & \text{RNAP} + \text{DNA}^{\text{p}}\text{:Act:Ind} \rightleftharpoons \text{RNAP:DNA}^{\text{p}}\text{:Act:Ind} \\
\text{Isomerization:} \quad & \text{RNAP:DNA}^{\text{p}}\text{:Act:Ind} \longrightarrow \text{RNAP:DNA}^{\text{o}}\text{:Act:Ind} \\
\text{Start of transcription:} \quad & \text{RNAP:DNA}^{\text{o}}\text{:Act:Ind} \longrightarrow \text{RNAP:DNA}^{1} + \\
& \qquad \text{DNA}^{\text{p}}\text{:Act:Ind}.
\end{aligned}$$

Hence, in the expression of the production rate $F([\text{Act}])$, we should substitute the concentration [Act:Ind] in place of [Act]. This concentration, in turn, is well approximated by its quasi-steady state value since binding reactions are much faster than isomerization and transcription, and can be obtained as for the negative inducer case.



(a) Unregulated (b) Negative autoregulation (c) Positive autoregulation

Figure 2.17: Autoregulation of gene expression. The three circuits control the expression of gene regulation using (a) unregulated, (b) negative autoregulation and (c) positive autoregulation.

Example 2.2 (Autoregulation of gene expression)**.** Consider the three circuits shown in Figure 2.17, representing a unregulated gene, a negatively autoregulated gene and a positively autoregulated gene. We want to model the dynamics of the protein A starting from zero initial conditions for the three different cases to understand how the three different circuit topologies affect dynamics.

The dynamics of the three circuits can be written in a common form,

$$\frac{dm_A}{dt} = F(A) - \delta m_A, \qquad \frac{dA}{dt} = \kappa m_A - \gamma A, \tag{2.23}$$

where $F(A)$ is in one of the following forms:

$$F_{\text{unreg}}(A) = \alpha_B, \qquad F_{\text{repress}}(A) = \frac{\alpha_B}{1+(A/K)^n} + \alpha_0, \qquad F_{\text{act}}(A) = \frac{\alpha_A (A/K)^n}{1+(A/K)^n} + \alpha_B$$

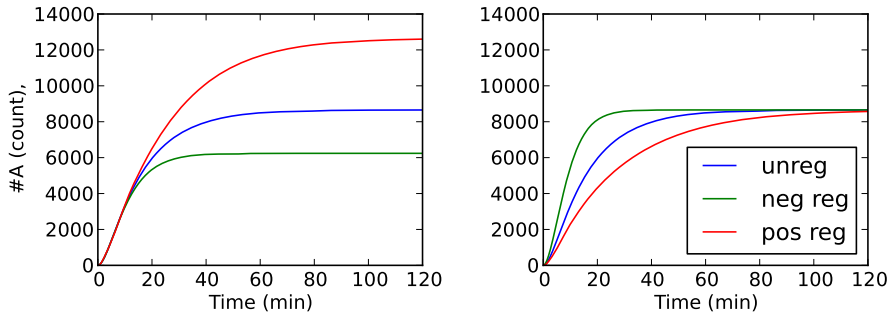
We choose the parameters to be

$$\begin{gathered} \alpha_A = 1/3, \qquad \alpha_B = 1/2, \qquad \alpha_0 = 5\times 10^{-4}, \\ \kappa = 20\log(2)/120, \qquad \delta = \log(2)/120, \qquad \gamma = \log(2)/600, \\ K = 10^4, \qquad n = 2, \end{gathered}$$

corresponding to biologically plausible values. Note that the parameters are chosen so that $F(0) \approx \alpha_B$ for each circuit.

Figure 2.18a shows the results of simulations comparing the response of the three circuits. We see that initial increase in protein concentration is identical for each circuit, consistent with our choice of Hill functions and parameters. As the expression level increases, the effects of positive and negative regulation are seen, leading to different steady state expression levels. In particular, the negative feedback circuit reaches a lower steady state expression level while the positive feedback circuit settles to a higher value.

In some situations, it makes sense to ask whether different circuit topologies have different properties that might lead us to choose one over another. In the case



(a) Default promoters (b) Modified promoters

Figure 2.18: Simulations for autoregulated gene expression. (a) Non-adjusted expression levels. (b) Equalizied expression levels.

where the circuit is going to be used as part of a more complex pathway, it may make the most sense to compare circuits that produce the same steady state concentration of the protein A. To do this, we must modify the parameters of the individual circuits, which can be done in a number of different ways: we can modify the promoter strengths, degradation rates, or other molecular mechanisms reflected in the parameters.

The steady state expression level for the negative autoregulation case can be adjusted by using a stronger promoter (modeled by α_B) or ribosome binding site (modeled by κ). The equilibrium point for the negative autoregulation case is given by the solution of the equations

$$m_{A,e} = \frac{\alpha K^n}{\delta(K^n + A_e^n)}, \qquad A_e = \frac{\kappa}{\gamma} m_{A,e}.$$

These coupled equations can be solved for $m_{A,e}$ and A_e, but in this case we simply need to find values α'_B and κ' that give the same values as the unregulated case. For example, if we equate the mRNA levels of the unregulated system with that of the negatively autoregulated system, we have

$$\frac{\alpha_B}{\delta} = \frac{1}{\delta}\left(\frac{\alpha'_B K^n}{K^n + A_e^n} + \alpha_0\right) \quad \Longrightarrow \quad \alpha'_B = (\alpha_B - \alpha_0)\frac{K^n + A_e^n}{K^n}, \quad A_e = \frac{\alpha_B \kappa}{\delta\gamma},$$

where A_e is the desired equilibrium value (which we choose using the unregulated case as a guide).

A similar calculation can be done for the case of positive autoregulation, in this case decreasing the promoter parameters α_A and α_B so that the steady state values match. A simple way to do this is to leave α_A unchanged and decrease α_B to account for the positive feedback. Solving for α'_B to give the same mRNA levels as the unregulated case yields

$$\alpha'_B = \alpha_B - \alpha_A \frac{A_e^n}{K^n + A_e^n}.$$

Figure 2.18b shows simulations of the expression levels over time for the modified circuits. We see now that the expression levels all reach the same steady state value. The negative autoregulated circuit has the property that it reaches the steady state more quickly, due to the increased rate of protein expression when A is small ($\alpha'_B > \alpha_B$). Conversely, the positive autoregulated circuit has a slower rate of expression than the constitutive case, since we have lowered the rate of protein expression when A is small. The initial higher and lower expression rates are compensated for via the autoregulation, resulting in the same expression level in steady state. ∇

We have described how a 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, as in the case of combinatorial promoters. The mRNA production rate can thus take several forms depending on the roles (activators versus repressors) of the various transcription factors. In general, the production rate resulting from a promoter that takes as input transcription factors $\mathrm{P_i}$ for $i \in \{1, ..., N\}$ will be denoted $F(P_1, ..., P_N)$.

The dynamics of a transcriptional module is often well captured by the ordinary differential equations

$$\frac{dm_{P_i}}{dt} = F(P_1, ..., P_N) - \delta_{P_i} m_{P_i}, \qquad \frac{dP_i}{dt} = \kappa_{P_i} m_{P_i} - \gamma_{P_i} P_i. \tag{2.24}$$

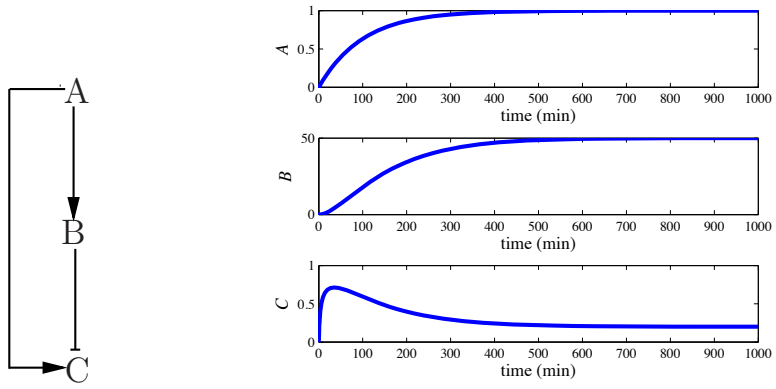
For a combinatorial promoter with two input proteins, an activator $\mathrm{P_a}$ and a repressor $\mathrm{P_r}$, in which, for example, the activator cannot bind if the repressor is bound to the promoter, the function $F(P_a, P_r)$ can be obtained by employing the competitive binding in the reduced order models of Section 2.1. In this case, assuming the activator has cooperativity n and the repressor has cooperativity m, we obtain the expression

$$F(P_a, P_r) = \alpha \frac{(P_a/K_a)^n}{1 + (P_a/K_a)^n + (P_r/K_r)^m}, \tag{2.25}$$

where $K_a = (K_{m,a} K_{\mathrm{d,a}})^{(1/n)}$, $K_r = (K_{m,r} K_{\mathrm{d,r}})^{(1/m)}$, in which $K_{\mathrm{d,a}}$ and $K_{\mathrm{d,r}}$ are the dissociation constants of the activator and repressor, respectively, from the DNA promoter site, while $K_{m,a}$ and $K_{m,r}$ are the dissociation constants for the cooperative binding reactions for the activator and repressor, respectively. In these expressions, RNA polymerase does not explicitly appear as it affects the values of the dissociation constants and of α. In the case in which the activator is "leaky", that is, some transcription still occurs even when there is no activator, the above expression should be modified to

$$F(P_a, P_r) = \alpha \frac{(P_a/K_a)^n}{1 + (P_a/K_a)^n + (P_r/K_r)^m} + \alpha_0, \tag{2.26}$$

where α_0 is the basal transcription rate when no activator is present. If the basal rate can still be repressed by the repressor, the above expression should be modified to



(a) Circuit diagram (b) Simulation

Figure 2.19: The incoherent feedforward loop (type I). (a) A schematic diagram of the circuit. (b) A simulation of a the model in equation (2.28) with $\beta_A = 0.01$, $\gamma = 0.01$, $\beta_B = 1$, $\beta_C = 100$, $K_B = 0.001$, and $K_A = 1$.

(see Exercise 2.10)

$$F(P_a, P_r) = \frac{\alpha(P_a/K_a)^n + \alpha_0}{1 + (P_a/K_a)^n + (P_r/K_r)^m}. \tag{2.27}$$

Example 2.3 (Incoherent feedforward loops)**.** Combinatorial promoters with two inputs are often used in systems where a logical "and" is required. As an example, we illustrate here an incoherent feedforward loop (type I) [4]. Such a circuit is composed of three transcription factors A, B, and C, in which A directly activates C and B while B represses C. This is illustrated in Figure 2.19a. This is different from a coherent feedforward loop in which both A and B activate C. In the incoherent feedforward loop, if we would like C to be high only when A is high and R is low ("and" gate), we can consider a combinatorial promoter in which the activator A and the repressor B competitively bind to the promoter of C. The resulting Hill function is given by the expression in equation (2.25). Depending on the values of the constants, the expression of C is low unless A is high and B is low. The resulting ODE model, neglecting the mRNA dynamics, is given by the system

$$\begin{aligned}
\frac{dA}{dt} &= \beta_A - \gamma A, \\
\frac{dB}{dt} &= \beta_B \frac{A/K_A}{1 + (A/K_A)} - \gamma B, \\
\frac{dC}{dt} &= \beta_C \frac{A/K_A}{1 + (A/K_A) + (B/K_B)} - \gamma C,
\end{aligned} \tag{2.28}$$

in which we have assumed no cooperativity of binding for both the activator and the repressor. If we view β_A as an input to the system and C as an output, we can

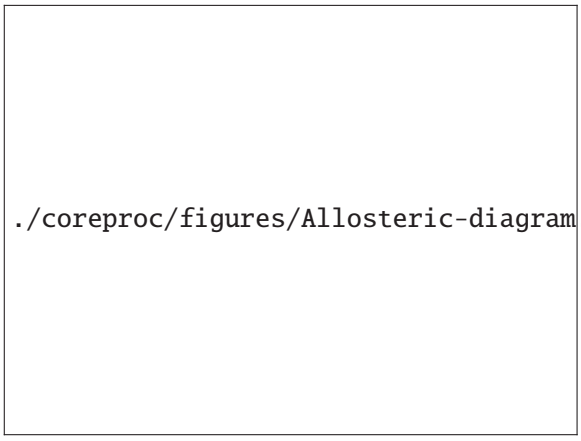


Figure 2.20: In allosteric regulation, a regulatory molecule binds to a site separate from the catalytic site (active site) of an enzyme. This binding causes a change in the three dimensional conformation of the protein, turning off (or turning on) the catalytic site. Permission pending.

investigate how this output responds to a sudden increase of β_A. Upon a sudden increase of β_A, protein A builds up and binds to the promoter of C initiating transcription, so that protein C starts getting produced. At the same time, protein B is produced and accumulates until it reaches a large enough value to repress C. Hence, we can expect a pulse of C production for suitable parameter values. This is shown in Figure 2.19b. Note that if the production rate constant β_C is very large, a little amount of A will cause C to immediately tend to a very high concentration. This explains the large initial slope of the C signal in Figure 2.19b. ∇

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.

Allosteric modifications to proteins

In allosteric regulation, a regulatory molecule, called *allosteric effector*, binds to a site separate from the catalytic site (active site) of an enzyme. This binding causes a change in the conformation of the protein, turning off (or turning on) the catalytic site (Figure 2.20).

An allosteric effector can either be an activator or an inhibitor, just like inducers

work for activation or inhibition of transcription factors. Inhibition can either be competitive or not competitive. In the case of competitive inhibition, the inhibitor competes with the substrate for binding the enzyme, that is, the substrate can bind to the enzyme only if the inhibitor is not bound. In the case of non-competitive inhibition, the substrate can be bound to the enzyme even if the latter is bound to the inhibitor. In this case, however, the product may not be able to form or may form at a lower rate, in which case, we have partial inhibition.

Activation can be absolute or not. Specifically, an allosteric effector is an *absolute activator* when the enzyme can bind to the substrate only when the enzyme is bound to the allosteric effector. Otherwise, the allosteric effector is a non-absolute activator. In this section, we derive the expressions for the production rate of the active protein in an enzymatic reaction in the two most common cases: when we have a (non-competitive) inhibitor I or an (absolute) activator A of the enzyme.

Allosteric inhibition

Consider the standard enzymatic reaction

$$\mathrm{E} + \mathrm{S} \underset{d}{\overset{a}{\rightleftharpoons}} \mathrm{ES} \overset{k}{\rightarrow} \mathrm{E} + \mathrm{P},$$

in which enzyme E binds to substrate S and transforms it into the product P. Let I be a (non-competitive) inhibitor of enzyme E so that when E is bound to I, the complex EI can still bind to substrate S, however, the complex EIS is non-productive, that is, it does not produce P. Then, we have the following additional reactions:

$$\mathrm{E} + \mathrm{I} \underset{k_-}{\overset{k_+}{\rightleftharpoons}} \mathrm{EI} \qquad \mathrm{ES} + \mathrm{I} \underset{k_-}{\overset{k_+}{\rightleftharpoons}} \mathrm{EIS} \qquad \mathrm{EI} + \mathrm{S} \underset{d}{\overset{a}{\rightleftharpoons}} \mathrm{EIS},$$

in which, for simplicity of notation, we have assumed that the dissociation constant between E and I does not depend on whether E is bound to the substrate S. Similarly, we have assumed that the dissociation constant of E from S does not depend on whether the inhibitor I is bound to E. Additionally, we have the conservation laws:

$$E_{\mathrm{tot}} = E + [\mathrm{ES}] + [\mathrm{EI}] + [\mathrm{EIS}], \qquad S_{\mathrm{tot}} = S + P + [\mathrm{ES}] + [\mathrm{EIS}].$$

The production rate of P is given by $dP/dt = k[\mathrm{ES}]$. Since binding reactions are very fast, we can assume that all the complexes concentrations are at the quasi-steady state. This gives

$$[\mathrm{EIS}] = \frac{a}{d}[\mathrm{EI}] \cdot S, \qquad [\mathrm{EI}] = \frac{k_+}{k_-} E \cdot I, \qquad [\mathrm{ES}] = \frac{1}{K_m} S \cdot E,$$

where $K_m = (d + k)/a$ is the Michaelis-Menten constant. Using these expressions, the conservation law for the enzyme, and the fact that $a/d \approx 1/K_m$, we obtain

$$E = \frac{E_{\mathrm{tot}}}{(I/K_{\mathrm{d}} + 1)(1 + S/K_m)}, \text{ with } K_{\mathrm{d}} = k_-/k_+,$$

so that

$$[\mathrm{ES}] = \frac{S}{S + K_m} \frac{E_{\mathrm{tot}}}{1 + I/K_{\mathrm{d}}}$$

and, as a consequence,

$$\frac{dP}{dt} = kE_{\mathrm{tot}} \left(\frac{1}{1 + I/K_{\mathrm{d}}} \right) \left(\frac{S}{S + K_m} \right).$$

In our earlier derivations of the Michaelis-Menten kinetics $V_{max} = k_1 E_{\mathrm{tot}}$ was called the maximal velocity, which occurs when the enzyme is completely saturated by the substrate (Section 2.1, equation (2.9)). Hence, the effect of a non-competitive inhibitor is to decrease the maximal velocity V_{max} to $V_{max}/(1 + I/K_{\mathrm{d}})$.

Another type of inhibition occurs when the inhibitor is competitive, that is, when I is bound to E, the complex EI cannot bind to protein S. Since E can either bind to I or S (not both), I competes against S for binding to E (see Exercise 2.13).

Allosteric activation

In this case, the enzyme E can transform S to its active form only when it is bound to A. Also, we assume that E cannot bind S unless E is bound to A (from here, the name absolute activator). The reactions should be modified to

$$\mathrm{E} + \mathrm{A} \underset{k_-}{\overset{k_+}{\rightleftharpoons}} \mathrm{EA}$$

and

$$\mathrm{EA} + \mathrm{S} \underset{d}{\overset{a}{\rightleftharpoons}} \mathrm{EAS} \overset{k}{\rightarrow} \mathrm{P} + \mathrm{EA},$$

with conservation laws

$$E_{\mathrm{tot}} = E + [\mathrm{EA}] + [\mathrm{EAS}], \qquad S_{\mathrm{tot}} = S + P + [\mathrm{EAS}].$$

The production rate of P is given by $dP/dt = k[\mathrm{EAS}]$. Assuming as above that the complexes are at the quasi-steady state, we have that

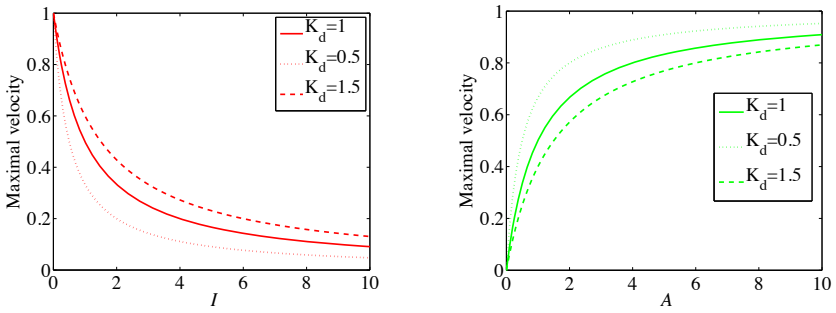
$$[\mathrm{EA}] = \frac{E \cdot A}{K_{\mathrm{d}}}, \qquad [\mathrm{EAS}] = \frac{S \cdot [\mathrm{EA}]}{K_m},$$

which, using the conservation law for E, leads to

$$E = \frac{E_{\mathrm{tot}}}{(1 + S/K_m)(1 + A/K_{\mathrm{d}})} \quad \text{and} \quad [\mathrm{EAS}] = \left(\frac{A}{A + K_{\mathrm{d}}} \right) \left(\frac{S}{S + K_m} \right) E_{\mathrm{tot}}.$$

Hence, we have that

$$\frac{dP}{dt} = kE_{\mathrm{tot}} \left(\frac{A}{A + K_{\mathrm{d}}} \right) \left(\frac{S}{S + K_m} \right).$$



(a) Allosteric inhibitor (b) Allosteric activator

Figure 2.21: Maximal velocity in the presence of allosteric effectors (inhibitors or activators). The plots in (a) show the maximal velocity $V_{max}/(1 + I/K_\mathrm{d})$ as a function of the inhibitor concentration I. The plots in (b) show the maximal velocity $V_{max}A/(A + K_\mathrm{d})$ as a function of the activator concentration A. The different plots show the effect of the dissociation constant for $V_{max} = 1$.

The effect of an absolute activator is to modulate the maximal speed of modification by a factor $A/(A + K_\mathrm{d})$.

Figure 2.21 shows the behavior of the maximal velocity as a function of the allosteric effector concentration. As the dissociation constant decreases, that is, the affinity of the effector increases, a very small amount of effector will cause the maximal velocity to reach V_{max} in the case of the activator and 0 in the case of the inhibitor.

Another type of activation occurs when the activator is not absolute, that is, when E can bind to S directly, but cannot activate S unless the complex ES first binds A (see Exercise 2.14).

Covalent modifications to proteins

In addition to regulation that controls transcription of DNA into mRNA, a variety of mechanisms are available for controlling expression after mRNA is produced. These include control of splicing and transport from the nucleus (in eukaryotes), the use of various secondary structure patterns in mRNA that can interfere with ribosomal binding or cleave the mRNA into multiple pieces, and targeted degradation of mRNA. Once the polypeptide chain is formed, additional mechanisms are available that regulate the folding of the protein as well as its shape and activity level.

One of the most common types of post-transcriptional regulation is through the covalent modification of proteins, such as through the process of *phosphorylation*. Phosphorylation is an enzymatic process in which a phosphate group is added to a

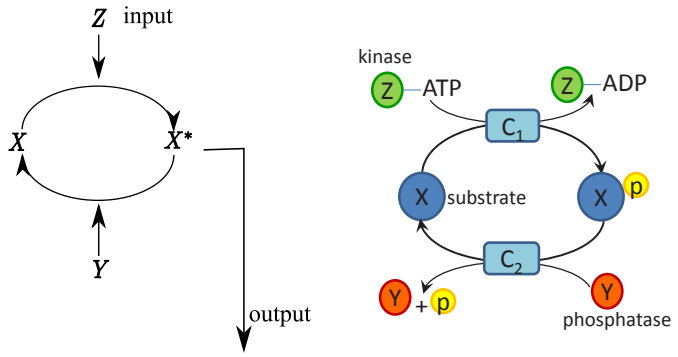
protein and the resulting conformation of the protein changes, usually from an inactive configuration to an active one. The enzyme that adds the phosphate group is called a *kinase* (or sometimes a *phosphotransferase*) and it operates by transferring a phosphate group from a bound ATP molecule to the protein, leaving behind ADP and the phosphorylated protein. *Dephosphorylation* is a complementary enzymatic process that can remove a phosphate group from a protein. The enzyme that performs dephosphorylation is called a *phosphatase*. Figure 2.3 shows the process of phosphorylation in more detail.

Phosphorylation is often used as a regulatory mechanism, with the phosphorylated version of the protein being the active conformation. Since phosphorylation and dephosphorylation can occur much more quickly than protein production and degradation, it is used in biological circuits in which a rapid response is required. One common pattern is that a signaling protein will bind to a ligand and the resulting allosteric change allows the signaling protein to serve as a kinase. The newly active kinase then phosphorylates a second protein, which modulates other functions in the cell. Phosphorylation cascades can also be used to amplify the effect of the original signal; we will describe this in more detail in Section 2.5.

Kinases in cells are usually very specific to a given protein, allowing detailed signaling networks to be constructed. Phosphatases, on the other hand, are much less specific, and a given phosphatase species may desphosphorylate many different types of proteins. The combined action of kinases and phosphatases is important in signaling since the only way to deactivate a phosphorylated protein is by removing the phosphate group. Thus phosphatases are constantly "turning off" proteins, and the protein is activated only when sufficient kinase activity is present.

Phosphorylation of a protein occurs by the addition of a charged phosphate (PO_4) group to the serine (Ser), threonine (Thr) or tyrosine (Tyr) amino acids. Similar covalent modifications can occur by the attachment of other chemical groups to select amino acids. *Methylation* occurs when a methyl group (CH_3) is added to lysine (Lys) and is used for modulation of receptor activity and in modifying histones that are used in chromatin structures. *Acetylation* occurs when an acetyl group ($COCH_3$) is added to lysine and is also used to modify histones. *Ubiquitination* refers to the addition of a small protein, ubiquitin, to lysine; the addition of a polyubiquitin chain to a protein targets it for degradation.

Here, we focus on *reversible* cycles of modification, in which a protein is interconverted between two forms that differ in activity. At a high level, a covalent modification cycle involves a target protein X, an enzyme Z for modifying it, and a second enzyme Y for reversing the modification (see Figure 2.22). 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 this added level of complexity here. The reactions describing this system are given by the following two enzymatic reactions, also called a *two-step reaction*



(a) (b)

Figure 2.22: (Left) General diagram representing a covalent modification cycle. (Right) Detailed view of a phoshorylation cycle including ATP, ADP, and the exchange og the phosphate group "p".

model:

$$\mathrm{Z}+\mathrm{X} \underset{d_1}{\overset{a_1}{\rightleftharpoons}} \mathrm{C}_1 \xrightarrow{k_1} \mathrm{X}^*+\mathrm{Z}, \qquad \mathrm{Y}+\mathrm{X}^* \underset{d_2}{\overset{a_2}{\rightleftharpoons}} \mathrm{C}_2 \xrightarrow{k_2} \mathrm{X}+\mathrm{Y},$$

in which we have let C_1=ZX be the kinase/protein complex and C_2=YX^* be the active protein/phosphatase complex. The corresponding differential equation model is given by

$$\begin{aligned}
\frac{dZ}{dt} &= -a_1 Z\cdot X + (k_1+d_1)C_1, & \frac{dX^*}{dt} &= k_1 C_1 - a_2 Y\cdot X^* + d_2 C_2,\\
\frac{dX}{dt} &= -a_1 Z\cdot X + d_1 C_1 + k_2 C_2, & \frac{dC_2}{dt} &= a_2 Y\cdot X^* - (d_2+k_2)C_2,\\
\frac{dC_1}{dt} &= a_1 Z\cdot X - (d_1+k_1)C_1, & \frac{dY}{dt} &= -a_2 Y\cdot X^* + (d_2+k_2)C_2.
\end{aligned}$$

Furthermore, we have that the total amounts of enzymes Z and Y are conserved. Denote the total concentrations of Z, Y, and X by Z_{tot}, Y_{tot}, and X_{tot}, respectively. Then, we have also the conservation laws

$$Z+C_1 = Z_{\mathrm{tot}}, \qquad Y+C_2 = Y_{\mathrm{tot}}, \qquad X_{\mathrm{tot}} = X+X^*+C_1+C_2.$$

Using the first two conservation laws, we can reduce the above system of differential equations to the following one:

$$\begin{aligned}
\frac{dC_1}{dt} &= a_1(Z_{\mathrm{tot}}-C_1)\cdot X - (d_1+k_1)C_1,\\
\frac{dX^*}{dt} &= k_1 C_1 - a_2(Y_{\mathrm{tot}}-C_2)\cdot X^* + d_2 C_2,\\
\frac{dC_2}{dt} &= a_2(Y_{\mathrm{tot}}-C_2)\cdot X^* - (d_2+k_2)C_2.
\end{aligned}$$

As in 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 concentrations C_1 and C_2 reach their steady state values very quickly under the assumption $a_1 Z_{\text{tot}}, a_2 Y_{\text{tot}}, d_1, d_2 \gg k_1, k_2$. Therefore, we can approximate the above system by substituting for C_1 and C_2 their steady state values, given by the solutions to

$$a_1(Z_{\text{tot}} - C_1)\cdot X - (d_1 + k_1)C_1 = 0$$

and

$$a_2(Y_{\text{tot}} - C_2)\cdot X^* - (d_2 + k_2)C_2 = 0.$$

By solving these equations, we obtain that

$$C_2 = \frac{Y_{\text{tot}} X^*}{X^* + K_{m,2}}, \quad \text{with} \quad K_{m,2} = \frac{d_2 + k_2}{a_2},$$

and

$$C_1 = \frac{Z_{\text{tot}} X}{X + K_{m,1}}, \quad \text{with} \quad K_{m,1} = \frac{d_1 + k_1}{a_1}.$$

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

$$\frac{dX^*}{dt} = k_1 \frac{Z_{\text{tot}} X}{X + K_{m,1}} - a_2 \frac{Y_{\text{tot}} K_{m,2}}{X^* + K_{m,2}} \cdot X^* + d_2 \frac{Y_{\text{tot}} X^*}{X^* + K_{m,2}},$$

which, considering that $a_2 K_{m,2} - d_2 = k_2$, leads finally to

$$\frac{dX^*}{dt} = k_1 \frac{Z_{\text{tot}} X}{X + K_{m,1}} - k_2 \frac{Y_{\text{tot}} X^*}{X^* + K_{m,2}}. \tag{2.29}$$

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 and mathematically characterize the assumptions needed for approximating the original system by the first order model (2.29). Also, note that X should be replaced by using the conservation law by $X = X_{\text{tot}} - X^* - C_1 - C_2$, which can be solved for X using the expressions of C_1 and C_2. Under the common assumption that the amount of enzyme is much smaller than the amount of substrate ($E_{\text{tot}} \ll X_{\text{tot}}$) [36], we have that $X \approx X_{\text{tot}} - X^*$ [36], leading to a form of the differential equation (2.29) that is simple enough to be analyzed mathematically.

Simpler models of phosphorylation cycles can be considered, which often times are instructive as a first step to study a specific question of interest. In particular, the *one-step reaction model* neglects the complex formation in the two enzymatic reactions and simply models them as a single irreversible reaction (see Exercise 2.12).

It is important to note that the speed of enzymatic reactions, such as phosphorylation and dephosphorylation, is usually much faster than the speed of protein

production and protein decay. In particular, the values of the catalytic rates k_1 and k_2, even if changing greatly from organism to organism, are typically several orders of magnitude larger than protein decay and can be on the order of 10^3 min^{-1} in bacteria where typical rates of protein decay are about 0.01 min^{-1} (`http://bionumbers.hms.harvard.edu/`).

Ultrasensitivity

One relevant aspect of the response of the covalent modification cycle to its input is the sensitivity of the steady state characteristic curve, that is, the map that determines the equilibrium value of the output X^* corresponding to a value of the input Z_{tot}. Specifically, which parameters affect the shape of the steady state characteristic is a crucial question. To study this, we set $dX^*/dt = 0$ in equation (2.29). Using the approximation $X \approx X_{\text{tot}} - X^*$, defining $\bar{K}_1 := K_{m,1}/X_{\text{tot}}$, and $\bar{K}_2 := K_{m,2}/X_{\text{tot}}$, we obtain

$$y := \frac{k_1 Z_{\text{tot}}}{k_2 Y_{\text{tot}}} = \frac{X^*/X_{\text{tot}}\left(\bar{K}_1 + (1 - X^*/X_{\text{tot}})\right)}{(\bar{K}_2 + X^*/X_{\text{tot}})(1 - X^*/X_{\text{tot}})}. \tag{2.30}$$

Since y is proportional to Z_{tot}, the input, we study the equilibrium value of X^* as a function of y. This function is usually characterized by two key parameters: the response coefficient, denoted R, and the point of half maximal induction, denoted y_{50}. Let y_α denote the value of y corresponding to having X^* equal $\alpha\%$ of the maximum value of X^* obtained for $y = \infty$, which is equal to X_{tot}. Then, the response coefficient is defined as

$$R := \frac{y_{90}}{y_{10}},$$

and measures how switch-like the response of X^* is to changes in y (Figure 2.23). When $R \to 1$ the response becomes switch-like. In the case in which the steady state characteristic is a Hill function, we have that $X^* = (y/K)^n/(1 + (y/K)^n)$, so that $y_\alpha = (\alpha/(100 - \alpha))^{(1/n)}$ and as a consequence

$$R = (81)^{(1/n)}, \text{ or equivalently } n = \frac{\log(81)}{\log(R)}.$$

Hence, when $n = 1$, that is, the characteristic is of the Michaelis-Menten type, we have that $R = 81$, while when n increases, R decreases. Usually, when $n > 1$ the response is referred to as *ultrasensitive* and the formula $n = \log(81)/\log(R)$ is often employed to estimate the *apparent Hill coefficient* of an experimentally obtained steady state characteristic since R can be calculated directly from the data points.

In the case of the current system, from equation (2.30), we have that

$$y_{90} = \frac{(\bar{K}_1 + 0.1)\, 0.9}{(\bar{K}_2 + 0.9)\, 0.1} \quad \text{and} \quad y_{10} = \frac{(\bar{K}_1 + 0.9)\, 0.1}{(\bar{K}_2 + 0.1)\, 0.9},$$

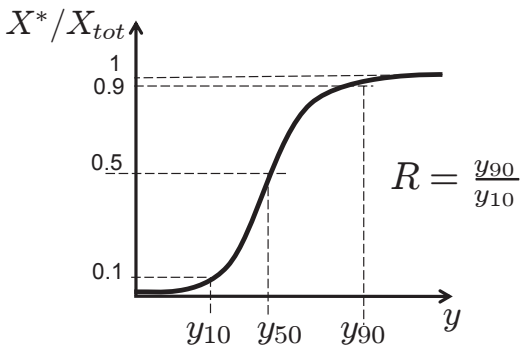


Figure 2.23: Steady state characteristic curve showing the relevance of the response coefficient for ultrasensitivity. As $R \to 1$, the points y_{10} and y_{90} tend to each other.

so that

$$R = 81\frac{(\bar{K}_1 + 0.1)(\bar{K}_2 + 0.1)}{(\bar{K}_2 + 0.9)(\bar{K}_1 + 0.9)}. \tag{2.31}$$

As a consequence, when $\bar{K}_1, \bar{K}_2 \gg 1$, we have that $R \to 81$, which gives a Michaelis-Menten type of response. If instead $\bar{K}_1, \bar{K}_2 \ll 0.1$, we have that $R \to 1$, which corresponds to a theoretical Hill coefficient $n \gg 1$, that is, a switch-like response (Figure 2.24). In particular, if we have, for example, $\bar{K}_1 = \bar{K}_2 = 10^{-2}$, we obtain an apparent Hill coefficient greater than 13. This type of ultrasensitivity is usually referred to as *zero-order ultrasensitivity*. The reason of this name is due to the fact that when $K_{m,1}$ is much smaller than the total amount of protein substrate X_{tot}, we have that $Z_{tot}X/(K_{m,1} + X) \approx Z_{tot}$. Hence, the forward modification rate is "zero order" in the substrate concentration (no free enzyme is left, all is bound to the substrate).

One can study the behavior also of the point of half maximal induction

$$y_{50} = \frac{\bar{K}_1 + 0.5}{\bar{K}_2 + 0.5},$$

to find that as $\bar{K}_2$ increases, y_{50} decreases and that as $\bar{K}_1$ increases, y_{50} increases.

Phosphotransfer systems

Phosphotransfer systems are also a common motif in cellular signal transduction. These structures are composed of proteins that can phosphorylate each other. In contrast to kinase-mediated phosphorylation, where the phosphate donor is usually ATP, in phosphotransfer the phosphate group comes from the donor protein itself (Figure 2.25). Each protein carrying a phosphate group can donate it to the next

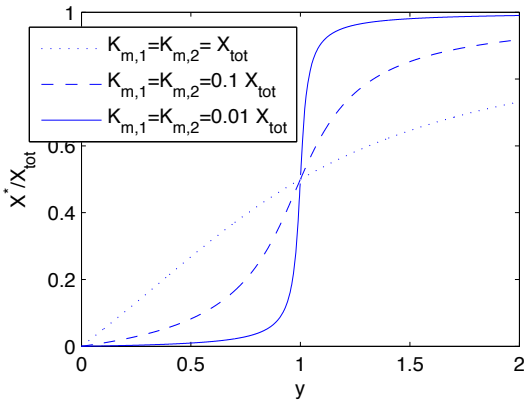


Figure 2.24: Steady state characteristic curve of a covalent modification cycle as a function of the Michaelis-Menten constants $K_{m,1}$ and $K_{m,2}$.

protein in the system through a reversible reaction.

Let X be a protein in its inactive form and let X^* be the same protein once it has been activated by the addition of a phosphate group. Let Z^* be a phosphate donor, that is, a protein that can transfer its phosphate group to the acceptor X. The standard phosphotransfer reactions can be modeled according to the two-step reaction model

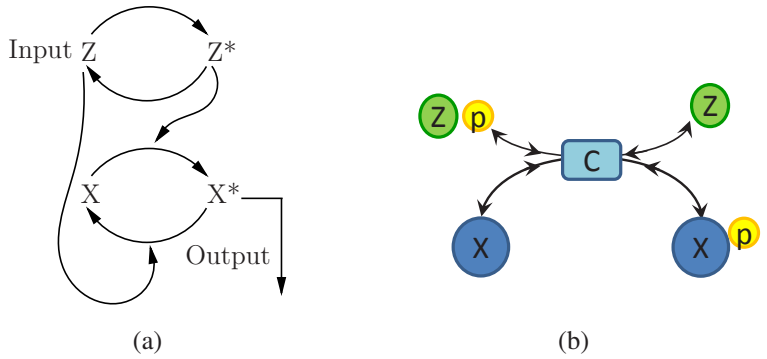
$$Z^* + X \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C_1 \underset{k_4}{\overset{k_3}{\rightleftharpoons}} X^* + Z,$$

in which C_1 is the complex of Z bound to X bound to the phosphate group. Additionally, we assume that protein Z can be phosphorylated and protein X^* dephosphorylated by other phosphorylation reactions by which the phosphate group is taken to and removed from the system. These reactions are modeled as one step reactions depending only on the concentrations of Z and X^*, that is:

$$Z \xrightarrow{\pi_1} Z^*, \qquad X^* \xrightarrow{\pi_2} X.$$

Proteins X and Z are conserved in the system, that is, $X_{\text{tot}} = X + C_1 + X^*$ and $Z_{\text{tot}} = Z + C_1 + Z^*$. We view the total amount of Z, Z_{tot}, as the input to our system and the amount of phosphorylated form of X, X^*, as the output. Note that this view is arbitrary as the system is completely symmetric with respect to X and Z. We are interested in the steady state characteristic curve describing how the steady state value of X^* depends on the value of Z_{tot}.

The differential equation model corresponding to this system is given by the



(a) (b)

Figure 2.25: (a) Diagram of a phosphotransfer system. (b) Proteins X and Z are transferring the phosphate group p to each other.

equations

$$\begin{aligned}
\frac{dC_1}{dt} &= k_1\,(X_{\text{tot}} - X^* - C_1)\cdot Z^* - k_3 C_1 - k_2 C_1 + k_4 X^*\cdot(Z_{\text{tot}} - C_1 - Z^*),\\
\frac{dZ^*}{dt} &= \pi_1(Z_{\text{tot}} - C_1 - Z^*) + k_2 C_1 - k_1\,(X_{\text{tot}} - X^* - C_1)\cdot Z^*,\\
\frac{dX^*}{dt} &= k_3 C_1 - k_4 X^*\cdot(Z_{\text{tot}} - C_1 - Z^*) - \pi_2 X^*.
\end{aligned} \tag{2.32}$$

The steady state transfer curve is shown in Figure 2.26 and it is obtained by simulating system (2.32) and recording the equilibrium values of X^* corresponding to different values of Z_{tot}. The transfer curve is linear for a large range of values of Z_{tot} and can be rendered fairly close to a linear relationship for values of Z_{tot} smaller than X_{tot} by increasing k_1. The slope of this linear relationship can be further tuned by changing the values of k_3 and k_4 (see Exercise 2.15). Hence, this system can function as an approximately linear anplifier. Its use in the realization of insulation devices that attenuate the effects of loading from interconnections will be illustrated in Chapter 6.

2.5 Cellular Subsystems

In the previous section we have studied how to model a variety of core processes that occur in cells. In this section we consider a few common "subsystems" in which these processes are combined for specific purposes.

Intercellular signaling: MAPK cascades

The mitogen activated protein kinase (MAPK) cascade is a recurrent structural motif in several signal transduction pathways (Figure 2.27). The cascade consists

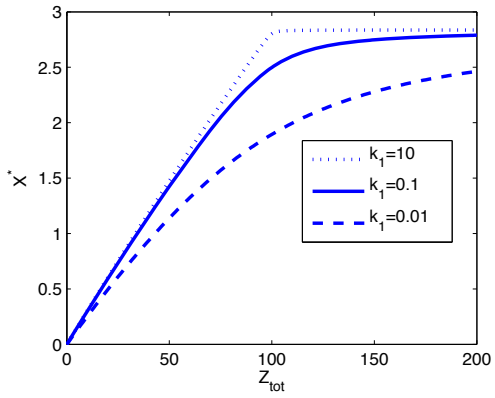


Figure 2.26: Steady state characteristic of the phosphotransfer system. Here, we have set $k_2 = k_3 = 0.1$ sec^{-1}, $k_4 = 0.1$ sec^{-1} nM^{-1}, $\pi_1 = \pi_2 = 3.1$ sec^{-1}, and $X_{\text{tot}} = 100$ nM.

of a MAPK kinase kinase (MAPKKK), denoted X_0, a MAPK kinase (MAPKK), denoted X_1, and a MAPK, denoted X_2. MAPKKKs activate MAPKKs by phosphorylation at two conserved sites and MAPKKs activate MAPKs by phosphorylation at conserved sites. The cascade relays signals from the cell membrane to targets in the cytoplasm and nucleus. It has been extensively studied and modeled. Here, we provide two different models. First, we build a modular model by viewing the system as the composition of single phosphorylation cycle modules (whose ODE model was derived earlier) and double phosphorylation cycle modules, whose ODE model we derive here. Then, we provide the full list of reactions describing the cascade and construct a mechanistic ODE model from scratch. We will then highlight the difference between the two derived models.

Double phosphorylation model. Consider the double phosphorylation motif in Figure 2.28. The reactions describing the system are given by

$$\mathrm{E}_1 + \mathrm{X} \underset{d_1}{\overset{a_1}{\rightleftharpoons}} \mathrm{C}_1 \xrightarrow{k_1} \mathrm{X}^* + \mathrm{E}_1, \qquad \mathrm{E}_2 + \mathrm{X}^* \underset{d_2}{\overset{a_2}{\rightleftharpoons}} \mathrm{C}_2 \xrightarrow{k_2} \mathrm{X} + \mathrm{E}_2,$$

$$\mathrm{X}^* + \mathrm{E}_1 \underset{d_1^*}{\overset{a_1^*}{\rightleftharpoons}} \mathrm{C}_3 \xrightarrow{k_1^*} \mathrm{X}^{**} + \mathrm{E}_1, \qquad \mathrm{E}_2 + \mathrm{X}^{**} \underset{d_2^*}{\overset{a_2^*}{\rightleftharpoons}} \mathrm{C}_4 \xrightarrow{k_2^*} \mathrm{X}^* + \mathrm{E}_2,$$

in which C_1 is the complex of E_1 with X, C_2 is the complex of E_2 with X^*, C_3 is the complex of E_1 with X^*, and C_4 is the complex of E_2 with X^{**}. The conservation laws are given by

$$E_1 + C_1 + C_3 = E_{1,\text{tot}}, \qquad E_2 + C_2 + C_4 = E_{2,\text{tot}},$$
$$X_{\text{tot}} = X + X^* + X^{**} + C_1 + C_2 + C_3 + C_4.$$

As performed earlier, we assume that the complexes are at the quasi-steady state since binding reactions are very fast compared to the catalytic rates. This gives the

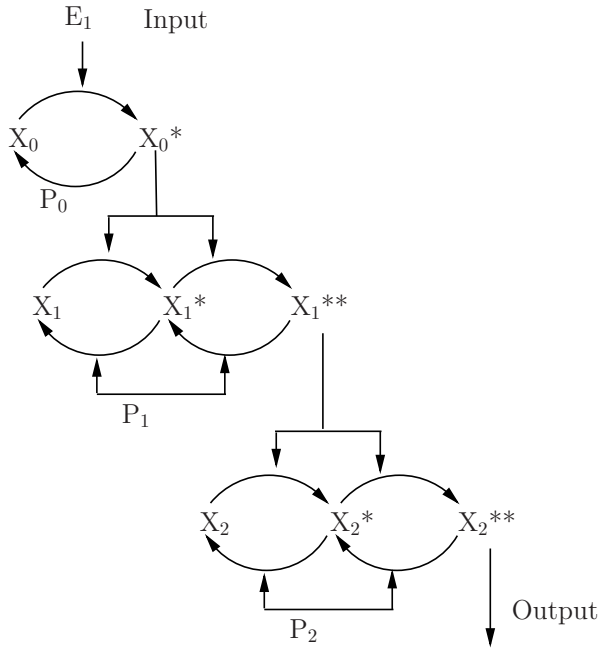


Figure 2.27: Schematic representing the MAPK cascade. It has three levels: the first one has a single phosphorylation, while the second and the third ones have a double phosphorylation.

Michaelis-Menten form for the amount of formed complexes:

$$C_1 = E_{1,\mathrm{tot}} \frac{K_1^* X}{K_1^* X + K_1 X^* + K_1 K_1^*}, \qquad C_3 = E_{1,\mathrm{tot}} \frac{K_1 X^*}{K_1^* X + K_1 X^* + K_1 K_1^*},$$

$$C_2 = E_{2,\mathrm{tot}} \frac{K_2^* X^*}{K_2^* X^* + K_2 X^{**} + K_2 K_2^*}, \qquad C_4 = E_{2,\mathrm{tot}} \frac{K_2 X^{**}}{K_2^* X^* + K_2 X^{**} + K_2 K_2^*},$$

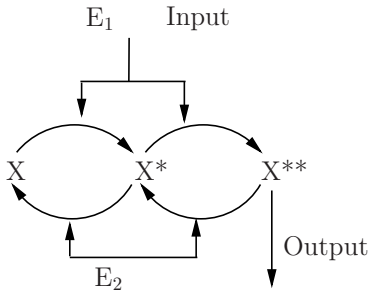


Figure 2.28: Schematic representing a double phosphorylation cycle. E_1 is the input and X^{**} is the output.

in which $K_i = (d_i + k_i)/a_i$ and $K_i^* = (d_i^* + k_i^*)/a_i^*$ are the Michaelis-Menten constants for the enzymatic reactions. Since the complexes are at the quasi-steady state, it follows that

$$\frac{d}{dt}X^* = k_1 C_1 - k_2 C_2 - k_1^* C_3 + k_2^* C_4,$$
$$\frac{d}{dt}X^{**} = k_1^* C_3 - k_2^* C_4,$$

from which, substituting the expressions of the complexes, we obtain that

$$\frac{d}{dt}X^* = E_{1,\text{tot}}\frac{k_1 X K_1^* - k_1^* X^* K_1}{K_1^* X + K_1 X^* + K_1^* K_1} + E_{2,\text{tot}}\frac{k_2^* X^{**} K_2 - k_2 X^* K_2^*}{K_2^* X^* + K_2 X^{**} + K_2 K_2^*}$$
$$\frac{d}{dt}X^{**} = k_1^* E_{1,\text{tot}}\frac{K_1 X^*}{K_1^* X + K_1 X^* + K_1 K_1^*} - k_2^* E_{2,\text{tot}}\frac{K_2\ X^{**}}{K_2^* X^* + K_2 X^{**} + K_2 K_2^*}.$$

Mechanistic model of the MAPK cascade. We now give the entire set of reactions for the MAPK cascade of Figure 2.27 as they are found in standard references (Huang-Ferrell model [44]):

$$\mathrm{E}_1 + \mathrm{X}_0 \underset{d_{1,0}}{\overset{a_{1,0}}{\rightleftharpoons}} \mathrm{C}_1 \xrightarrow{k_{1,0}} \mathrm{X}_0^* + \mathrm{E}_1 \qquad \mathrm{P}_0 + \mathrm{X}_0^* \underset{d_{2,0}}{\overset{a_{2,0}}{\rightleftharpoons}} \mathrm{C}_2 \xrightarrow{k_{2,0}} \mathrm{X}_0 + \mathrm{P}_0$$
$$\mathrm{X}_0^* + \mathrm{X}_1 \underset{d_{1,1}}{\overset{a_{1,1}}{\rightleftharpoons}} \mathrm{C}_3 \xrightarrow{k_{1,1}} \mathrm{X}_1^* + \mathrm{X}_0^* \qquad \mathrm{X}_1^* + \mathrm{P}_1 \underset{d_{2,1}}{\overset{a_{2,1}}{\rightleftharpoons}} \mathrm{C}_4 \xrightarrow{k_{2,1}} \mathrm{X}_1 + \mathrm{P}_1$$
$$\mathrm{X}_0^* + \mathrm{X}_1^* \underset{d_{1,1}^*}{\overset{a_{1,1}^*}{\rightleftharpoons}} \mathrm{C}_5 \xrightarrow{k_{1,1}^*} \mathrm{X}_1^{**} + \mathrm{X}_0^* \qquad \mathrm{X}_1^{**} + \mathrm{P}_1 \underset{d_{2,1}^*}{\overset{a_{2,1}^*}{\rightleftharpoons}} \mathrm{C}_6 \xrightarrow{k_{2,1}^*} \mathrm{X}_1^* + \mathrm{P}_1$$
$$\mathrm{X}_1^{**} + \mathrm{X}_2 \underset{d_{1,2}}{\overset{a_{1,2}}{\rightleftharpoons}} \mathrm{C}_7 \xrightarrow{k_{1,2}} \mathrm{X}_2^* + \mathrm{X}_1^{**} \qquad \mathrm{X}_2^* + \mathrm{P}_2 \underset{d_{2,2}}{\overset{a_{2,2}}{\rightleftharpoons}} \mathrm{C}_8 \xrightarrow{k_{2,2}} \mathrm{X}_2 + \mathrm{P}_2$$
$$\mathrm{X}_1^{**} + \mathrm{X}_2^* \underset{d_{1,2}^*}{\overset{a_{1,2}^*}{\rightleftharpoons}} \mathrm{C}_9 \xrightarrow{k_{1,2}^*} \mathrm{X}_2^{**} + \mathrm{X}_1^{**} \qquad \mathrm{X}_2^{**} + \mathrm{P}_2 \underset{d_{2,2}^*}{\overset{a_{2,2}^*}{\rightleftharpoons}} \mathrm{C}_{10} \xrightarrow{k_{2,2}^*} \mathrm{X}_2^* + \mathrm{P}_2,$$

with conservation laws

$$X_{0,\text{tot}} = X_0 + X_0^* + C_1 + C_2 + C_3 + C_5$$
$$X_{1,\text{tot}} = X_1 + X_1^* + C_3 + X_1^{**} + C_4 + C_5 + C_6 + C_7 + C_9$$
$$X_{2,\text{tot}} = X_2 + X_2^* + X_2^{**} + C_7 + C_8 + C_9 + C_{10}$$
$$E_{1,\text{tot}} = E_1 + C_1,\ \ P_{0,\text{tot}} = P_0 + C_2$$
$$P_{1,\text{tot}} = P_1 + C_4 + C_6$$
$$P_{2,\text{tot}} = P_2 + C_8 + C_{10}.$$

The corresponding ODE model is given by

$$
\begin{aligned}
\frac{dC_1}{dt} &= a_{1,0}E_1\ X_0 - (d_{1,0} + k_{1,0})\ C_1 \\
\frac{dX_0^*}{dt} &= k_{1,0}\ C_1 + d_{2,0}\ C_2 - a_{2,0}\ P_0\ X_0^* + (d_{1,1} + k_{1,1})\ C_3 - a_{1,1}\ X_1\ X_0^* \\
&\quad + (d_{1,1}^* + k_{1,1}^*)\ C_5 - a_{1,1}^*\ X_0^*\ X_1^* \\
\frac{dC_2}{dt} &= a_{2,0}\ P_0\ X_0^* - (d_{2,0} + k_{2,0})\ C_2 \\
\frac{dC_3}{dt} &= a_{1,1}\ X_1\ X_0^* - (d_{1,1} + k_{1,1})\ C_3 \\
\frac{dX_1^*}{dt} &= k_{1,1}\ C_3 + d_{2,1}\ C_4 - a_{2,1}\ X_1^*\ P_1 + d_{1,1}^* C_5 - a_{1,1}^*\ X_1^*\ X_0^* + k_{2,1}^*\ C_6 \\
\frac{dC_4}{dt} &= a_{2,1}\ X_1^*\ P_1 - (d_{2,1} + k_{2,1})\ C_4 \\
\frac{dC_5}{dt} &= a_{1,1}^*\ X_0^*\ X_1^* - (d_{1,1}^* + k_{1,1}^*)\ C_5 \\
\frac{dX_1^{**}}{dt} &= k_{1,1}^*\ C_5 - a_{2,1}^*\ X_1^*\ P_1 + d_{2,1}^*\ C_6 - a_{1,2}\ X_1^{**}\ X_2 \\
&\quad + (d_{1,2} + k_{1,2})\ C_7 - a_{1,2}^*\ X_1^{**}\ X_2^* + (d_{1,2}^* + k_{1,2}^*)\ C_9 \\
\frac{dC_6}{dt} &= a_{2,1}^*\ X_1^{**}\ P_1 - (d_{2,1}^* + k_{2,1}^*)\ C_6 \\
\frac{dC_7}{dt} &= a_{1,2}^*\ X_1^*\ X_2 - (d_{1,2}^* + k_{1,2}^*)\ C_7 \\
\frac{dX_2^*}{dt} &= -a_{2,2}\ X_2^*\ P_2 + d_{2,2}\ C_8 - a_{1,2}^*\ X_2^*\ X_2^{**} + d_{1,2}^*\ C_9 + C_{10}\ K_{10} \\
\frac{dC_8}{dt} &= a_{2,2}^*\ X_2^*\ P_2 - (d_{2,2} + k_{2,2})\ C_8 \\
\frac{dX_2^{**}}{dt} &= k_{1,2}^*\ C_9 - a_{2,2}^*\ X_2^{**}\ P_2 + d_{2,2}^*\ C_{10} \\
\frac{dC_9}{dt} &= a_{1,2}^*\ X_1^{**}\ X_2^* - (d_{1,2}^* + k_{1,2}^*)\ C_9 \\
\frac{dC_{10}}{dt} &= a_{2,2}^*\ X_2^{**}\ P_2 - (d_{2,2}^* + k_{2,2}^*)\ C_{10}.
\end{aligned}
$$

The steady state characteristic curve obtained with the mechanistic model predicts that the response of the MAPKKK to the stimulus $E_{1,\text{tot}}$ is of the Michaelis-Menten type. By contrast, the stimulus-response curve obtained for the MAPKK and MAPK are sigmoidal and show high Hill coefficients, which increase from the MAPKK response to the MAPK response. That is, an increased ultrasensitivity is observed moving down in the cascade (Figure 2.29). These model observations persist when key parameters, such as the Michaelis-Menten constants are changed [44]. Furthermore, zero-order ultrasensitivity effects can be observed. Specifically,

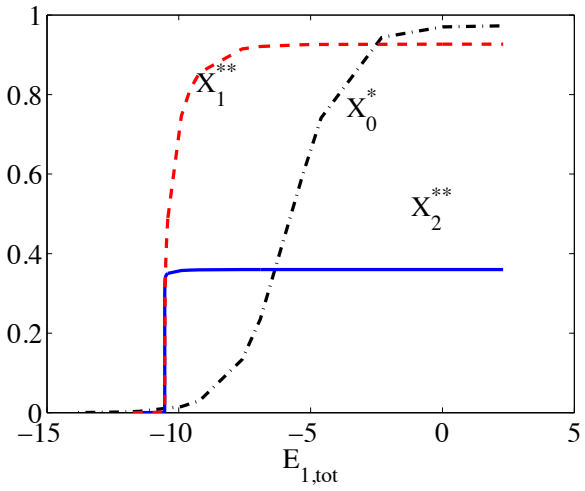


Figure 2.29: Dose response of the MAPK cascade for every stage. The x-axis shows concentration of $E_{1,\text{tot}}$ between 10^{-15} and 10^{5} as indicated. Simulations from the model of [79].

if the amounts of MAPKK were increased, one would observe a higher apparent Hill coefficient for the response of MAPK. Similarly, if the values of the K_m for the reactions in which the MAPKK takes place were decreased, one would also observe a higher apparent Hill coefficient for the response of MAPK. Double phosphorylation is also key to obtain a high apparent Hill coefficient. In fact, a cascade in which the double phosphorylation was assumed to occur through a one-step model (similar to single phosphorylation) predicted substantially lower apparent Hill coefficients.

Notice that even if we are used to think phosphorylation cascades such as the MAPK cascade as unidirectional signal transmission systems, it is actually possible for information to travel backward (from downstream to upstream). This can be qualitatively seen as follows. Assume as before that the total amounts of enzymes are much smaller than the total amounts of substrates ($E_{1,\text{tot}}, P_{0,\text{tot}}, P_{1,\text{tot}}, P_{2,\text{tot}} \ll X_{0,\text{tot}}, X_{1,\text{tot}}, X_{2,\text{tot}}$), we can approximate the conservation laws as

$$
\begin{aligned}
X_{0,\text{tot}} &\approx X_0 + X_0^* + C_3 + C_5, \\
X_{1,\text{tot}} &\approx X_1 + X_1^* + C_3 + X_1^{**} + C_5 + C_7 + C_9, \\
X_{2,\text{tot}} &\approx X_2 + X_2^* + X_2^{**} + C_7 + C_9.
\end{aligned}
$$

Using these and assuming that the complexes are at the quasi-steady state, we ob-

tain the following functional dependencies:

$$C_1 = f_1(X_0^*, X_1^*, X_1^{**}, X_2^*, X_2^{**}), \qquad C_2 = f_2(X_0^*),$$
$$C_3 = f_3(X_0^*, X_1^*, X_1^{**}, X_2^*, X_2^{**}), \qquad C_5 = f_5(X_0^*, X_1^*),$$
$$C_7 = f_7(X_1^*, X_1^{**}, X_2^*, X_2^{**}), \qquad C_9 = f_9(X_1^{**}, X_2^*).$$

The fact that C_7 depends on X_2^* and X_2^{**} illustrates the counter-intuitive fact that the dynamics of the second stage are influenced by those of the third stage. Similarly, the fact that C_3 depends on $X_1^*, X_1^{**}, X_2^*, X_2^{**}$ indicates that the dynamics of the first stage are influenced by those of the second stage and by that of the third stage. The phenomenon by which the behavior of a "module" is influenced by that of its downstream clients is called *retroactivity*, which is a phenomenon similar to loading in electrical and mechanical systems, studied at length in Chapter 6. This phenomenon in signaling cascades can allow perturbations to travel from downstream to upstream [74] and can lead to interesting dynamic behaviors such as sustained oscillations in the MAPK cascade [79].

Exercises

2.1 Consider a cascade of three activators X→Y→ Z. Protein X is initially present in the cell in its inactive form. The input signal of X, S_x, appears at time $t = 0$. As a result, X rapidly becomes active and binds the promoter of gene Y, so that protein Y starts to be produced at rate β. When Y levels exceed a threshold K, gene Z begins to be transcribed and translated at rate β. All proteins have the same degradation/dilution rate γ.

(a) What are the concentrations of proteins Y and Z as a function of time?

(b) What is the minimum duration of the pulse S_x such that Z will be produced?

(c) What is the response time of protein Z with respect to the time of addition of S_x?

2.2 (Switch-like behavior in cooperative binding) Derive the expressions of C and A as a function of B at the steady state when you have the cooperative binding reactions

$$\mathrm{B} + \mathrm{B} + \ldots + \mathrm{B} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{B_n}, \quad \mathrm{B_n} + \mathrm{A} \underset{d}{\overset{a}{\rightleftharpoons}} \mathrm{C}, \quad \text{and} \quad A + C = A_{\mathrm{tot}}.$$

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 case of a competitive binding of an activator A and a repressor R with D and assume that before they can bind to D they have to cooperatively bind

according to the following reactions:

$$\mathrm{A}+\mathrm{A}+\ldots+\mathrm{A} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{A_n}, \qquad \mathrm{R}+\mathrm{R}+\ldots+\mathrm{R} \underset{\bar{k}_2}{\overset{\bar{k}_1}{\rightleftharpoons}} \mathrm{R_m},$$

in which the complex $\mathrm{A_n}$ contains n molecules of A and the complex $\mathrm{R_m}$ contains m molecules of R. The competitive binding reactions with A are given by

$$\mathrm{A_n}+\mathrm{D} \underset{d}{\overset{a}{\rightleftharpoons}} \mathrm{C}, \qquad \mathrm{R_m}+\mathrm{D} \underset{d'}{\overset{a'}{\rightleftharpoons}} \mathrm{C}',$$

and $D_{\mathrm{tot}} = D + C + C'$. What are the steady state expressions for C and D as functions of A and R?

2.4 Consider the following modification of the competitive binding reactions:

$$\mathrm{B_a}+\mathrm{A} \underset{d}{\overset{a}{\rightleftharpoons}} \mathrm{C}, \qquad \mathrm{B_r}+\mathrm{A} \underset{\bar{d}}{\overset{\bar{a}}{\rightleftharpoons}} \bar{C},$$

and

$$\mathrm{C}+\mathrm{B_r} \underset{d'}{\overset{a'}{\rightleftharpoons}} \mathrm{C}', \qquad \bar{C}+\mathrm{B_a} \underset{\bar{d}'}{\overset{\bar{a}'}{\rightleftharpoons}} \mathrm{C}',$$

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

2.5 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 large quantity 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 the complexes formed with DNA (D) that lead to such an input function. Demonstrate that indeed the set of reactions you picked leads to the desired input function.

2.6 Consider the transcription and translation reactions incorporating the elongation process as considered in this chapter in equations (2.10)–(2.11). Modify them to the case in which an mRNA molecule can be translated to a polypeptide chain even while it is still elongating.

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

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

2.8 Derive the expression of the parameters α, α_0 and K for the Hill function given in equation (2.20), which is the form obtained for transcriptional repression with a leaky repressor.

2.9 Consider the form of the Hill function in the presence of an activator with some basal level of expression given in equation (2.22). Derive the expressions of α, K and α_0,

2.10 Derive the form of the Hill functions for combinatorial promoters with leakiness given in expressions (2.26)–(2.27).

2.11 Consider the phosphorylation reactions described in Section 2.4, but suppose that the kinase concentration Z is not constant, but is instead produced and decays according to the reaction $\mathrm{Z} \underset{k(t)}{\overset{\gamma}{\rightleftharpoons}} \emptyset$. How should the system in equation (2.29) be modified? Use a MATLAB simulation to apply a periodic input stimulus $k(t)$ using parameter values: $k_1 = k_2 = 1$, $a_1 = a_2 = d_1 = d_2 = 10$, $\gamma = 0.01$. Is the cycle capable of "tracking" the input stimulus? If yes, to what extent when the frequency of $k(t)$ is increased? What are the tracking properties depending on?

2.12 Another model for the phosphorylation reactions, referred to as *one- step reaction model*, is given by $\mathrm{Z} + \mathrm{X} \longrightarrow \mathrm{X}^* + \mathrm{Z}$ and $\mathrm{Y} + \mathrm{X}^* \longrightarrow \mathrm{X} + \mathrm{Y}$, in which the complex formations are neglected. Write down the differential equation model and compare the differential equation of X^* to that of equation (2.29). List the assumptions under which the one step reaction model is a good approximation of the two step reaction model.

2.13 (Competitive inhibition) Derive the expression of the production rate of X^* in a phosphorylation cycle in the presence of a competitive inhibitor I for Z.

2.14 (Non-absolute activator) Derive the expression of the production rate of X^* in a phosphorylation cycle in the presence of a non-absolute activator A for Z.

2.15 Consider the model of phosphotransfer systems of equation (2.32) and determine how the steady state transfer curve changes when the values of k_3 and k_4 are changed.