Biomolecular Feedback Systems

Domitilla Del Vecchio
MIT

Richard M. Murray
Caltech

Classroom Copy v0.6c, July 11, 2012

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.

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

2.1 Modeling Techniques

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

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

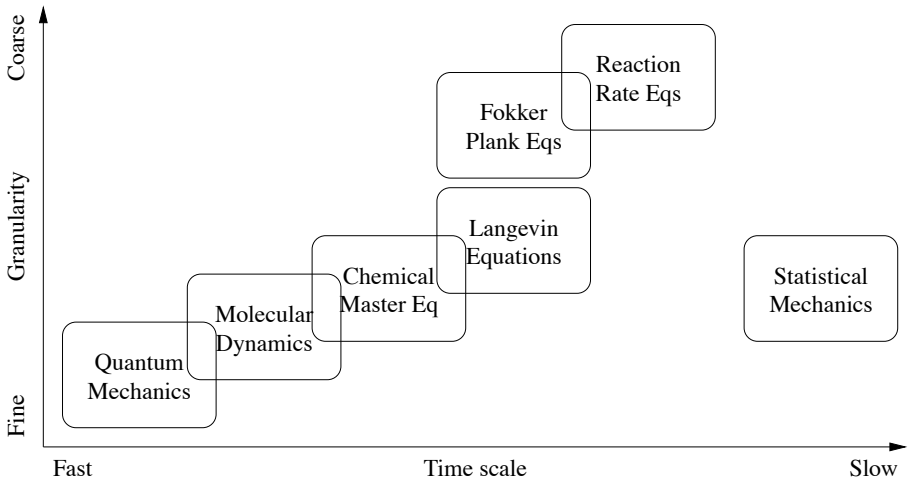


Figure 2.1: Different methods of modeling biomolecular systems.

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

Statistical mechanics and chemical kinetics

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

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

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

Figure 2.2: Microstates for RNA polymerase. Each microstate of the system corresponds to the RNA polymerase being located at some position in the cell. If we discretize the possible locations on the DNA and in the cell, the microstates corresponds to all possible non-overlapping locations of the RNA polymerases. Figure from Phillips, Kondev and Theriot [76]; used with permission of Garland Science.

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

In statistical mechanics, we model the configuration of the cell by the probability that 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*,

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

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

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

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

This can be used, for example, to compute the probability that some RNA polymerase is bound to a given promoter, averaged over many independent samples, and from this we can reason about the rate of expression of the corresponding gene. More details and several examples will be illustrated in Chapter 4.

Statistical mechanics describes the steady state distribution of microstates, but does not tell us how the microstates evolve in time. To include the dynamics, we must consider the *chemical kinetics* of the system and model the probability that we transition from one microstate to another in a given period of time. Let q represent the microstate of the system, which we shall take as a vector of integers that represents the number of molecules of a specific types in given configurations or locations. Assume we have a set of M reactions R_j, $j = 1,\ldots,M$, with ξ_j representing the change in state q associated with reaction R_j. We describe the kinetics of the system by making use of the *propensity function* $a_j(q,t)$ associated with reaction R_j, 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

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

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. The variable ξ_j in the sum ranges over all possible reactions.

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

$$\mathrm{A}+\mathrm{B} \rightleftharpoons \mathrm{AB} \quad \equiv \quad \begin{array}{l} R_\mathrm{f} : \mathrm{A}+\mathrm{B} \longrightarrow \mathrm{AB} \\ R_\mathrm{r} : \mathrm{AB} \longrightarrow \mathrm{A}+\mathrm{B}. \end{array} \tag{2.4}$$

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

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

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 R_r 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 molecules of A, 1 molecule of B, and 3 molecules of AB, then the possible states and dynamics are

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

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 proceed to write the dynamics of a system consisting of a set of species S_i, $i = 1, \ldots, n$ undergoing a set of reactions R_j, $j = 1, \ldots, m$. We write $x_i = [S_i] = n_{S_i}/\Omega$ 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 a differential equation

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

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

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

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

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 $a_\mathrm{r}(q) = k_\mathrm{r}n_\mathrm{AB}$. It follows that the concentration of the complex AB satisfies

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

in which $\mathbb{E}(x)$ denotes the expected value of x. Taking the limit as dt approaches zero (but remains large enough that we can still average across multiple reactions, as described in more detail in Chapter 4), we obtain

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

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

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

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

$$\begin{aligned} \frac{d}{dt}A &= k_\mathrm{r}C^2 \cdot D - k_\mathrm{f}A \cdot B^2, & \frac{d}{dt}C &= 2k_\mathrm{f}A \cdot B^2 - 2k_\mathrm{r}C^2 \cdot D, \\ \frac{d}{dt}B &= 2k_\mathrm{r}C^2 \cdot D - 2k_\mathrm{f}A \cdot B^2, & \frac{d}{dt}D &= k_\mathrm{f}A \cdot B^2 - k_\mathrm{r}C^2 \cdot D. \end{aligned} \tag{2.5}$$

Rearranging this equation, we can write the dynamics as

$$\frac{d}{dt}\begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ -2 & 2 \\ 2 & -2 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} k_\mathrm{f}A \cdot B^2 \\ k_\mathrm{r}C^2 \cdot D \end{pmatrix}. \tag{2.6}$$

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

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

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

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

Sometimes, the following notation will be used to denote birth and death of species

$$\emptyset \xrightarrow{k_{\mathrm{f}}} \mathrm{A}, \ \mathrm{A} \xrightarrow{k_{\mathrm{r}}} \emptyset.$$

We attach to the first reaction the differential equation

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

and to the second reaction we attach the differential equation

$$\frac{dA}{dt} = -k_{\mathrm{r}}, A.$$

From a physical point of view, these reactions simplify the representation of more complex processes, such as production of proteins or degradation of proteins due to proteases.

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 1.17. Let S represent the substrate, K represent the kinase and S^* represent the phosphorylated (activated) substrate. The sets of reactions illustrated in Figure 1.17 are

$$\begin{aligned}
R_1&: \quad \mathrm{K} + \mathrm{ATP} \longrightarrow \mathrm{K{:}ATP} & \qquad R_5&: \quad \mathrm{S{:}K{:}ATP} \longrightarrow \mathrm{S^*{:}K{:}ADP} \\
R_2&: \quad \mathrm{K{:}ATP} \longrightarrow \mathrm{K} + \mathrm{ATP} & \qquad R_6&: \quad \mathrm{S^*{:}K{:}ADP} \longrightarrow \mathrm{S^*} + \mathrm{K{:}ADP} \\
R_3&: \quad \mathrm{S} + \mathrm{K{:}ATP} \longrightarrow \mathrm{S{:}K{:}ATP} & \qquad R_7&: \quad \mathrm{K{:}ADP} \longrightarrow \mathrm{K} + \mathrm{ADP} \\
R_4&: \quad \mathrm{S{:}K{:}ATP} \longrightarrow \mathrm{S} + \mathrm{K{:}ATP} & \qquad R_8&: \quad \mathrm{K} + \mathrm{ADP} \longrightarrow \mathrm{K{:}ADP}.
\end{aligned}$$

We now write the kinetics for each reaction:

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

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

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

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

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

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:

$$\mathrm{A} + \mathrm{B} \underset{d}{\overset{a}{\rightleftharpoons}} \mathrm{C}, \tag{2.7}$$

where a is the association rate constant and b 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 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

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

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

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

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 $B \approx K_\mathrm{d}$, A and C have roughly 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 (2.7) become

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

in which B_2 denotes the dimer of B. The corresponding ODE model is given by

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

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

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

so that

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

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

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

then we have that

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

In this case, one says that the binding of B to A is *cooperative* with cooperativity n. Figure 2.3 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 as bound. In this case, the reactions are given by

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

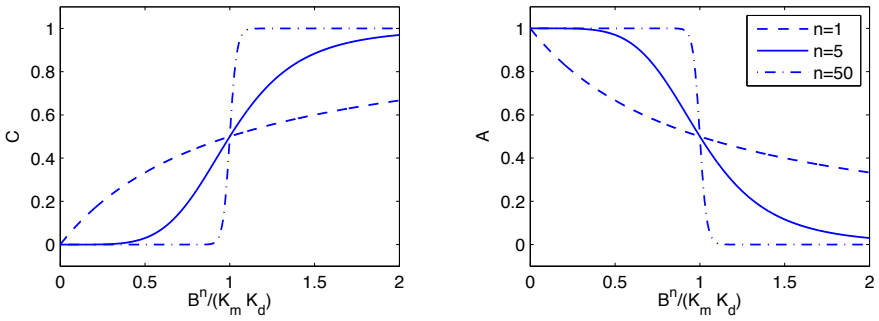


Figure 2.3: Steady state concentrations of the complex C and of A as functions of the concentration of B.

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

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

By solving this system of two equations for the unknowns C' and C, one obtains

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

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

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

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

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

for which we can write the dynamics as

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

By setting the derivatives to zero, we obtain that

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

so that

$$C_r = \frac{B_r(A_{\text{tot}} - C_a)}{B_r + K'_{\text{d}}}, \qquad C_a\left(B_a + K_{\text{d}} - \frac{B_a B_r}{B_r + K'_{\text{d}}}\right) = B_a\left(\frac{K'_{\text{d}}}{B_r + K'_{\text{d}}}\right)A_{\text{tot}},$$

from which we finally obtain that

$$C_a = \frac{(B_a/K_{\text{d}})A_{\text{tot}}}{(B_a/K_{\text{d}}) + (B_r/K'_{\text{d}}) + 1}, \qquad C_r = \frac{(B_r/K'_{\text{d}})A_{\text{tot}}}{(B_r/K'_{\text{d}}) + (B_a/K_{\text{d}}) + 1}.$$

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 as an exercise (see Exercises) that they would appear in the final expressions with a power of two.

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 included another complex comprising B_a, B_r and A. Denoting this new complex by C′, we would have added also the two additional reactions

$$\text{C}_\text{a} + \text{B}_\text{r} \underset{d}{\overset{a}{\rightleftharpoons}} \text{C}', \qquad \text{C}_\text{r} + \text{B}_\text{a} \underset{d'}{\overset{a'}{\rightleftharpoons}} \text{C}'$$

and we would have modified the conservation law for A to $A_{\text{tot}} = A + C_a + C_r + C'$. The reader can verify as an exercise (see Exercises) that in this case a mixed term $B_r B_a$ would appear in the equilibrium expressions.

Enzymatic reaction. A general enzymatic reaction can be written as

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

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

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

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

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

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

$$a, d \gg k$$

on the rate constants.

Under this assumption and assuming that $S \gg E$ (at least at time 0; see Example 3.15), C immediately reaches its steady state value (while P is still changing). 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

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

is called *Michaelis-Menten kinetics*.

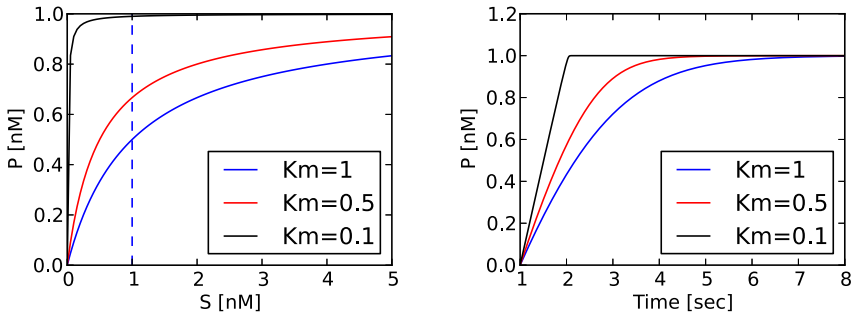
The constant V_{max} is called the maximal velocity (or maximal flux) of modification and it represents the maximal rate that can be obtained when the enzyme is completely saturated by the substrate. The value of K_m corresponds to the value of S that leads to a half-maximal value of the P production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

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

When $K_m \ll S_{\text{tot}}$ and the substrate has not yet been all converted to product, that is, $S_{\text{tot}} - P \gg K_m$, 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 *zero-order kinetics*. When $S_{\text{tot}} - P \gg K_m$, the system is said to operate in the zero-order regime (see Figure 2.4).

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



(a) Transfer curves (b) Simulations

Figure 2.4: Enzymatic reactions. (a) Transfer curve showing the production rate for P as a function of substrate concentration. (b) Time plots of product $P(t)$ for different values of the K_m. In the plots $S_{\text{tot}} = 1$ and $V_{max} = 1$. The black plot shows the behavior for a value of K_m much smaller that the total substrate amount S_{tot}. This corresponds to a constant product formation rate (at least before the substrate is almost all converted to product, that is, $S_{\text{tot}} - P \approx K_m$), which is referred to *zero-order kinetics*.

of levels of detail and which model to use depends on the questions that one wants to consider. We present several levels of modeling here, starting with a fairly detailed set of reactions and ending with highly simplified models that can be used when we are only interested in average production rate of proteins at relatively long time scales.

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

$$
\begin{aligned}
\text{Binding to DNA:} \quad & \mathrm{RNAP^c} \rightleftharpoons \mathrm{RNAP^d} \\
\text{Diffusion along DNA:} \quad & \mathrm{RNAP^d} \rightleftharpoons \mathrm{RNAP^p} \\
\text{Binding to promoter:} \quad & \mathrm{RNAP^p + DNA^p} \rightleftharpoons \mathrm{RNAP{:}DNA^p} \\
\text{Isomerization:} \quad & \mathrm{RNAP{:}DNA^p} \rightleftharpoons \mathrm{RNAP{:}DNA^o} \\
\text{Start of transcription:} \quad & \mathrm{RNAP{:}DNA^o} \longrightarrow \mathrm{RNAP{:}DNA^{g,1} + DNA^p} \\
\text{mRNA creation:} \quad & \mathrm{RNAP{:}DNA^{g,1}} \longrightarrow \mathrm{RNAP{:}DNA^{g,2} + mRNA_k^1} \\
\text{Elongation:} \quad & \mathrm{RNAP{:}DNA^{g,i+1} + mRNA_k^i} \\
& \qquad \longrightarrow \mathrm{RNAP{:}DNA^{g,i+2} + mRNA_k^{i+1}} \\
\text{Binding to terminator:} \quad & \mathrm{RNAP{:}DNA^{g,N} + mRNA_k^{N-1}} \\
& \qquad \longrightarrow \mathrm{RNAP{:}DNA^t + mRNA_k^N} \\
\text{Termination:} \quad & \mathrm{RNAP{:}DNA^t} \longrightarrow \mathrm{RNAP^c} \\
\text{Degradation:} \quad & \mathrm{mRNA_k^N} \longrightarrow \emptyset.
\end{aligned}
\tag{2.8}
$$

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

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

can then be written as

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

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

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

Given a set of reactions, the various stochastic processes that underly detailed models of transcription and translation can be specified using the stochastic modeling framework described briefly in the previous section. In particular, using either models of binding energy or measured rates, we can construct propensity functions for each of the many reactions that lead to production of proteins, including the motion of RNA polymerase and the ribosome along DNA and RNA. For many problems in which the detailed stochastic nature of the molecular dynamics of the cell are important, these models are the most relevant and they are covered in some detail in Chapter 4.

Alternatively, we can move to the reaction rate formalism and model the reactions using differential equations. To do so, we must compute the various reaction rates, which can be obtained from the propensity functions or measured experimentally. In moving to this formalism, we approximate the concentrations of various species as real numbers, which may not be accurate since some species exist at

low molecular counts in the cell. Despite all of these approximations, in many situations the reaction rate equations are perfectly sufficient, particularly if we are interested in the average behavior of a large number of cells.

In some situations, an even simpler model of the transcription, translation and folding processes can be utilized. 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

$$\frac{dm}{dt} = \alpha_{p,0} - \mu m - \gamma m, \qquad m^*(t) = e^{-\mu\tau^m} m(t-\tau^m), \tag{2.9}$$

where m is the concentration of mRNA for protein P, m^* is the concentration of active mRNA, $\alpha_{p,0}$ is the rate of production of the mRNA for protein P, μ is the growth rate of the cell (which results in dilution of the concentration) and γ 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 γ. The exponential factor accounts for dilution due to the change in volume of the cell, where μ is the cell growth rate. The constants $\alpha_{p,0}$ and γ capture the average rates of production and degradation, which in turn depend on the more detailed biochemical reactions that underlie transcription.

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

$$\frac{dP}{dt} = \beta_{p,0} m^* - \delta P, \qquad P^f(t) = e^{-\mu\tau^f} P(t-\tau^f). \tag{2.10}$$

Here P represents the concentration of the polypeptide chain for the protein, P^f represents the concentration of functional protein (after folding). The parameters that govern the dynamics are $\beta_{p,0}$, the rate of translation of mRNA; δ, 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 δ, captures both rate at which the polypeptide chain is degraded and the rate at which the concentration is diluted due to cell growth.

It will often be convenient to write the dynamics for transcription and translation in terms of the functional mRNA and functional protein. Differentiating the

expression for m^*, we see that

$$\begin{aligned}\frac{dm^*(t)}{dt} &= e^{-\mu\tau^m}\dot{m}(t-\tau^m) \\ &= e^{-\mu\tau^m}(\alpha_{p,0}-\gamma m(t-\tau^m)) = \alpha_{p,0}-\gamma m^*(t),\end{aligned} \tag{2.11}$$

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

$$\frac{dP^f(t)}{dt} = \beta_{p,0}m^*(t-\tau^f) - \delta P^f(t), \tag{2.12}$$

where $\beta_{p,0} = e^{-\mu\tau^f}\beta_{p,0}$. We shall typically use equations (2.11) and (2.12) as our (reduced) description of protein folding, dropping the superscript f and overbars when there is no risk of confusion.

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

$$\frac{dm}{dt} = \alpha_{p,0} - \gamma m, \qquad \frac{dP}{dt} = \beta_{p,0}m - \delta P. \tag{2.13}$$

Note that we here have dropped the superscripts $*$ and f since we are assuming that all mRNA is active and proteins are functional and dropped the overbar on α and β since we are assuming the time delays are negligible.

Finally, the simplest model for protein production is one in which we only keep track of the basal rate of production of the protein, without including the mRNA dynamics. This essentially amounts to assuming the mRNA dynamics reach steady state quickly and replacing the first differential equation in equation (2.13) with its equilibrium value. This is often a good assumption as mRNA degration is usually about 100–1000 times faster than protein degradation (see Table 1.1). Thus we obtain

$$\frac{dP}{dt} = \beta - \delta P, \qquad \beta := \beta_{p,0}\frac{\alpha_{p,0}}{\gamma}.$$

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

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 the 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 1.9. 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 post-transcriptional control mechanisms. We will focus on prokaryotic mechanisms.

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

We can capture this set of molecular interactions by modifying the RNA polymerase binding reactions in equation (2.8). For a repressor (Rep), we simply have to add a reaction that represents the repressor bound to the promoter:

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

This reaction acts to "sequester" the DNA promoter site so that it is no longer available for binding by RNA polymerase (which requires $\mathrm{DNA^p}$). 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 forward rate. 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. One possible mechanism is

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

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

A simplified version of the dynamics can be obtained by assuming that transcription factors bind to the DNA rapidly, so that they are in steady state configurations. In this case, we can make use of the reduced order models described in Section 2.1. We can consider the competitive binding case to model that a strong repressor prevents RNAP to bind to the DNA. In the sequel, we remove the superscripts "p" from the DNA and RNAP for simplifying notation. The steady state amount of the complex of DNA bound to the repressor will have the expression

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

and the steady state amount of free DNA (not bound to the repressor) will be given by

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

in which K'_d is the dissociation constant of RNAP from the promoter while K_d is the dissociation constant of Rep from the promoter. The complex C, having RNAP bound, will allow transcription, while the complex [DNA:Rep] will not allow transcription as it is not bound to RNAP.

The transcription rate will be proportional to C, so that the rate of change of mRNA is described by

$$\frac{d[\text{mRNA}]}{dt} = \alpha_0 \frac{([\text{RNAP}]/K'_\text{d})[\text{DNA}]}{1 + [\text{RNAP}]/K'_\text{d} + [\text{Rep}]/K_\text{d}} - \gamma[\text{mRNA}],$$

in which the production rate is given by

$$f([\text{Rep}]) = \alpha_0 \frac{([\text{RNAP}]/K'_\text{d})\,[\text{DNA}]}{1 + [\text{RNAP}]/K'_\text{d} + [\text{Rep}]/K_\text{d}}.$$

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

$$f([\text{Rep}]) = \alpha_0 \frac{([\text{RNAP}]/K'_\text{d})[\text{DNA}]}{1 + [\text{RNAP}]/K'_\text{d} + [\text{Rep}]^n/(K_\text{d}k_\text{m})},$$

in which k_m is the dissociation constant of the reaction of n molecules of Rep binding together. The function f is usually denoted by the standard Hill function form

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

in which α and K are implicitly defined. In practice we can assume that $[\text{RNAP}]/K'_{\text{d}} \gg 1$ since there is plenty of RNAP in the cell. As a consequence, we obtain the expressions $\alpha = \alpha_0[\text{DNA}]$ and $K = (K_{\text{d}}k_{\text{m}}[\text{RNAP}]/K'_{\text{d}})^{1/n}$.

Finally, if the repressor allows RNAP to still bind to the promoter at a small rate (leaky repressor), the above expression modifies to the new form (see Section 2.1)

$$f([\text{Rep}]) = \frac{\alpha}{1 + ([\text{Rep}]/K)^n} + \alpha, \tag{2.14}$$

in which α is the basal expression level when the promoter is fully repressed, usually referred to as "leakiness".

To model the production rate of mRNA in the case in which an activator Act binds to the promoter with cooperativity n, we can consider first the case in which RNAP binds only when the activator is already bound to the promoter. This can be well modeled by a cooperative binding scenario as illustrated in Section 2.1. According to this scenario, the concentration of the complex [RNAP:DNA^o] is given by

$$[\text{RNAP:DNA}^o] = C' = \frac{([\text{RNAP}][\text{Act}]^n)/(K_{\text{d}}K'_{\text{d}}k_{\text{m}})[\text{DNA}]}{1 + ([\text{Act}]^n/K_{\text{d}}k_{\text{m}})(1 + [\text{RNAP}]/K'_{\text{d}})},$$

in which K'_{d} is the dissociation constant of RNAP with the complex of DNA bound to Act and K_{d} is the dissociation constant of Act with DNA. Since the production rate of mRNA is proportional to [RNAP:DNA^o], we have that

$$\frac{d\,[\text{mRNA}]}{dt} = f([\text{Act}]) - \gamma[\text{mRNA}]$$

with

$$f([\text{Act}]) = \alpha_0 \frac{([\text{RNAP}][\text{Act}]^n)/(K_{\text{d}}K'_{\text{d}}k_{\text{m}})[\text{DNA}]}{1 + ([\text{Act}]^n/K_{\text{d}}k_{\text{m}})(1 + [\text{RNAP}]/K'_{\text{d}})} =: \frac{\alpha([\text{Act}]/K)^n}{1 + ([\text{Act}]/K)^n},$$

in which α and K are implicitly defined. Since in practice $[\text{RNAP}]/K'_{\text{d}} \gg 1$, we have that $\alpha = \alpha_0[\text{DNA}]$ and $K = (K_{\text{d}}K'_{\text{d}}k_{\text{m}}/[\text{RNAP}])^{1/n}$.

The right-hand side expression is in the standard Hill function form. Figure 2.5 shows the shape of these Hill functions both for an activator and a repressor. If we assume that RNAP can still bind to DNA even when the activator is not bound, we have an additional basal expression rate α so that the new form of the production rate is given by

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

Example 2.2 (Repressilator)**.** As an example of how these models can be used, we consider the model of a "repressilator," originally due to Elowitz and Leibler [27] and briefly described in Section 1.5. The repressilator is a synthetic circuit in which

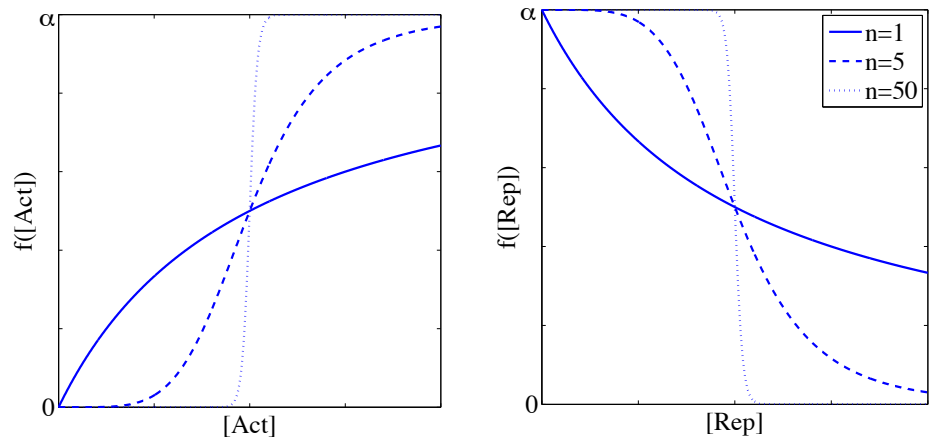


Figure 2.5: Hill function for an activator (left) and a repressor (right).

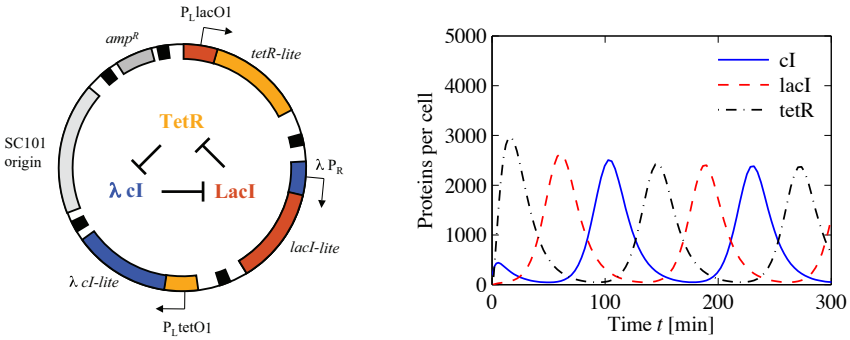
three proteins each repress another in a cycle. This is shown schematically in Figure 2.6a, where the three proteins are TetR, λ cI and LacI.

We can model this system using three copies of the repression function (2.14), with Rep replaced by the appropriate combination of TetR, cI and LacI. The state of the system is then given by $x = (m_{\text{TetR}}, p_{\text{TetR}}, m_{\text{cI}}, p_{\text{cI}}, m_{\text{LacI}}, p_{\text{LacI}})$. The full dynamics become

$$\frac{d}{dt}\begin{pmatrix} m_{\text{TetR}} \\ p_{\text{TetR}} \\ m_{\text{cI}} \\ p_{\text{cI}} \\ m_{\text{LacI}} \\ p_{\text{LacI}} \end{pmatrix} = \begin{pmatrix} \dfrac{\alpha_{\text{LacI}}}{1+(p_{\text{LacI}}/K_{\text{LacI}})^n} + \alpha_{\text{TetR}} - \gamma\, m_{\text{TetR}} \\ \beta_{\text{TetR}}\, m_{\text{TetR}} - \delta\, p_{\text{TetR}} \\ \dfrac{\alpha_{\text{TetR}}}{1+(p_{\text{TetR}}/K_{\text{TetR}})^n} + \alpha_{\text{cI}} - \gamma\, m_{\text{cI}} \\ \beta_{\text{cI}}\, m_{\text{cI}} - \delta\, p_{\text{cI}} \\ \dfrac{\alpha_{\text{cI}}}{1+(p_{\text{cI}}/K_{\text{cI}})^n} + \alpha_{\text{LacI}} - \gamma\, m_{\text{LacI}} \\ \beta_{\text{LacI}}\, m_{\text{LacI}} - \delta\, p_{\text{LacI}} \end{pmatrix}. \tag{2.15}$$

Figure 2.6b shows the traces of the three protein concentrations for (symmetric) parameters $n = 2$, $\alpha = 0.5$, $K = 6.25\times10^{-4}$, $\alpha_0 = 5\times10^{-4}$, $\gamma = 5.8\times10^{-3}$, $\beta = 0.12$ and $\delta = 1.2\times10^{-3}$ with initial conditions $x(0) = (1,200,0,0,0,0)$ (following [27]). ∇

As indicated earlier, many activators and repressors operate in the presence of inducers. To incorporate these dynamics in our description, we simply have to add the reactions that correspond to the interaction of the inducer with the relevant protein. For a negative inducer, we can simply add a reaction in which the inducer



(a) Repressilator plasmid (b) Repressilator simulation

Figure 2.6: The repressilator genetic regulatory network. (a) A schematic diagram of the repressilator, showing the layout of the genes in the plasmid that holds the circuit as well as the circuit diagram (center). (b) A simulation of a simple model for the repressilator, showing the oscillation of the individual protein concentrations. (Figure courtesy M. Elowitz.)

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

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

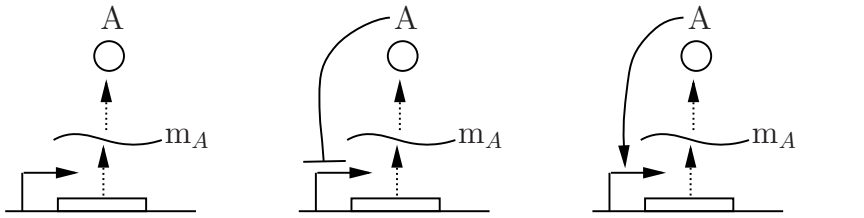
Since the above reactions are very fast compared to transcription, they can be assumed at the quasi-steady state. Hence, the free amount of repressor that can still bind to the promoter can be calculated by writing the ODE model corresponding to the above reactions and by setting the time derivatives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

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

in which $[\text{Rep}]_{\text{tot}} = [\text{Rep}] + [\text{Rep:Ind}]$ is the total amount of repressor (bound and not bound to the inducer) and K_{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

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



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

Figure 2.7: Autoregulation of gene expression. The three circuits control the expression of gene regulation using (a) unregulated, (b) negative autoregulation and (c) positive autoregulation.

Hence, in the expression of the production rate $f([\text{Act}])$, we should substitute in place of [Act] the concentration [Act:Ind]. This concentration, in turn, can be simply computed at the quasi-steady state by writing the ODE model for the inducer binding reaction and equating the time derivatives to zero. This yields

$$[\text{Act:Ind}] = \frac{[\text{Act}]_{\text{tot}}[\text{Ind}]/K_{\text{d}}}{1+[\text{Ind}]/K_{\text{d}}},$$

in which $[\text{Act}]_{\text{tot}} = [\text{Act}] + [\text{Act:Ind}]$ and K_{d} is the dissociation constant of the binding of Ind with Act.

Example 2.3 (Autoregulation of gene expression)**.** Consider the three circuits shown in Figure 2.7, 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,

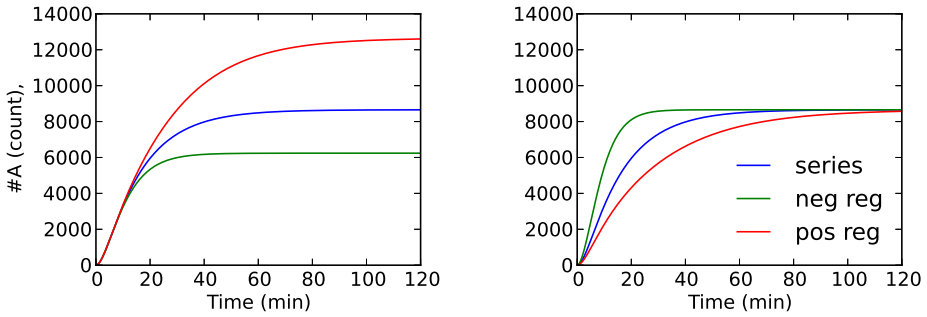
$$\frac{dm_A}{dt} = f(A) - \gamma m_A, \qquad \frac{dA}{dt} = \beta m_A - \delta A, \tag{2.16}$$

where $f(A)$ has the form

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

We choose the parameters to be

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



(a) (b)

Figure 2.8: Simulations for autoregulated gene expression. (a) Non-normalized expression levels. (b) Normalized expression.

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

Figure 2.8a shows the results of the simulation. 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 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 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 β). The equilibrium point for the negative autoregulation case is given by the solution of the equations

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

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 β' 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

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

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 2.8b 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 the Hill function can model the regulation of a gene by a single transcription factor. However, genes can also be regulated by multiple transcription factors, some of which may be activators and some may be repressors. In this case, the promoter controlling the expression of the gene is called a combinatorial promoter. The mRNA production rate can thus take several forms depending on the roles (activators versus repressors) of the various transcription factors [3]. 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 $f(p_1, ..., p_n)$.

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

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

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

For a combinatorial promoter with two input proteins, an activator $\mathrm{p_a}$ and a repressor $\mathrm{p_r}$, in which 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

$$f(p_a, p_r) = \alpha \frac{(p_a/K_a)^n}{1 + (p_a/K_a)^n + (p_r/K_r)^m}.$$

Here, we have that $K_a = (K_{m,a} K_{\mathrm{d,a}})^{(1/n)}$, $K_r = (K_{m,r} K_{\mathrm{d,r}})^{(1/m)}$, 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 the case in which the activator is "leaky", that is, some transcription still occurs even when there is no activator, the above expression will be modified to

$$f(p_a, p_r) = \alpha \frac{(p_a/K_a)^n}{1 + (p_a/K_a)^n + (p_r/K_r)^m} + \alpha,$$

in which α is the basal transcription rate when no activator is present. If such a basal rate can still be repressed by the repressor, the above expression modifies to the form

$$f(p_a, p_r) = \frac{\alpha(p_a/K_a)^n + \alpha}{1 + (p_a/K_a)^n + (p_r/K_r)^m}.$$

Example 2.4 (Activator-repressor clock)**.** As an example of where combinatorial promoters are used, we illustrate in this example an activator-repressor clock that was fabricated in *E. coli* and is shown in Figure 2.9(a) [5].

The activator A is self activated and is also repressed by the repressor R. Hence, the promoter controlling the expression of A is a combinatorial promoter. The model describing this system, assuming the mRNA dynamics have reached its quasi-steady state, is given by

$$\frac{dA}{dt} = \frac{\alpha_A (A/K_a)^n + \alpha_A}{(A/K_a)^n + (R/K_r)^m + 1} - \delta_A A, \qquad \frac{dR}{dt} = \frac{\alpha_R (A/K_a)^n + \alpha_R}{(A/K_a)^n + 1} - \delta_R R.$$

Figure 2.9 (b) shows the behavior of the activator and the repressor concentrations. We will come back to this design in Chapter 6, in which we will use the tools introduced in Chapter 3 to establish parameter conditions under which the system admits a periodic solution. ∇

Finally, a simple regulation mechanism is based on altering the half life of a protein. 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 X by a protease Y can then be modeled by the following two-step reaction

$$\mathrm{X} + \mathrm{Y} \underset{d}{\overset{a}{\rightleftharpoons}} \mathrm{C} \xrightarrow{k} \mathrm{Y},$$

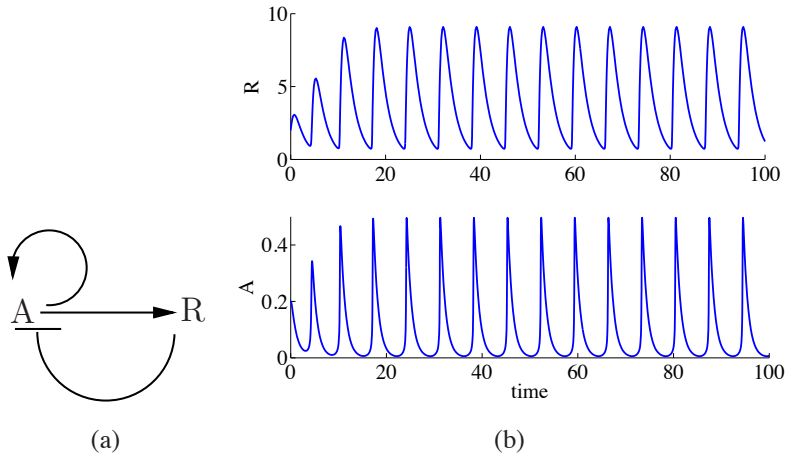


Figure 2.9: The activator-repressor clock network. (a) A schematic diagram of the circuit. (b) A simulation of a simple model for the clock, showing the oscillation of the individual protein concentrations. In the simulation, we have chosen $K_a = K_r = 1$, $\alpha_A = \alpha_R = 100$, $\alpha_A = 0.4$, $\alpha_R = 0.004$, $\delta_A = 1$, $\delta_R = 0.5$, $n = 2$, and $m = 4$.

in which C is the complex of the protease bound to the protein. By the end of the reaction, protein X has been degraded to nothing, so that sometimes this reaction is simplified to $X \longrightarrow \emptyset$.

2.4 Post-Transcriptional Regulation

In addition to regulation of expression through modifications of the process of transcription, cells can also regulate the production and activity of proteins via a collection of other post-transcriptional modifications. These include methods of modulating the translation of proteins, as well as affecting the activity of a protein via changes in its conformation, as shown in Figure 1.9.

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 three dimension conformation of the protein, turning off (or turning on) the catalytic site (Figure 2.10).

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

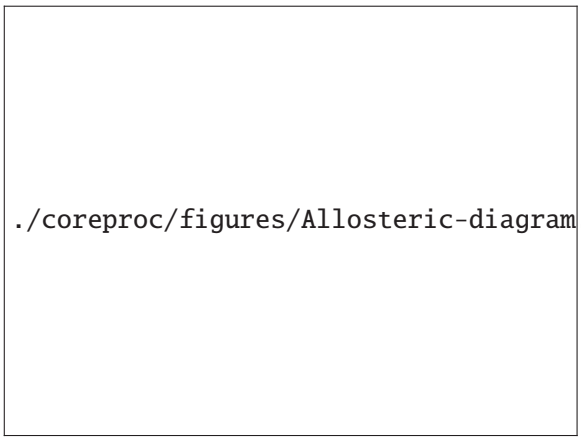


Figure 2.10: In allosteric regulation, a regulatory molecule binds to a site separate from the catalytic site (active site) of an enzyme. This binding causes a change in the three dimension conformation of the protein, turning off (or turning on) the catalytic site. Permission pending.

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 activator is absolute when the enzyme can bind to the substrate only when bound to the activator. Otherwise, the activator is not absolute. 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

$$\mathrm{E}+\mathrm{S} \underset{d}{\overset{a}{\rightleftharpoons}} \mathrm{C} \overset{k}{\rightarrow} \mathrm{S}^{*}+\mathrm{E}$$

in which enzyme E activates protein S and transforms it to the active form S^{*}. 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 inactive protein S, however, the complex EIS is non-productive, that is, it does not produce the active protein S^{*}. Then, we have the following additional reactions:

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

with the conservation laws (assuming S_{tot} is in much greater amounts than E_{tot})

$$E_{\mathrm{tot}} = E + C + EI + EIS, \qquad S_{\mathrm{tot}} = S + S^{*} + C + EIS \approx S + S^{*}.$$

Hence, the production rate of S^* is given by $dS^*/dt = kC$. Since we have that $k_+, k_-, a, b \gg k$, we can assume all the complexes to be at the quasi steady state. This gives

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

in which $K_m = (d+k)/a$ 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

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

so that

$$C = \frac{S}{S + K_m} \frac{E_{\text{tot}}}{1 + I/K_{\text{d}}}$$

and, as a consequence,

$$\frac{dS^*}{dt} = k_1 E_{\text{tot}} \left(\frac{1}{1 + I/K_{\text{d}}} \right) \left(\frac{S}{S + K_m} \right).$$

Using the conservation law for S, this is also equivalent to

$$\frac{dS^*}{dt} = k_1 E_{\text{tot}} \left(\frac{1}{1 + I/K_{\text{d}}} \right) \left(\frac{(S_{\text{tot}} - S^*)}{(S_{\text{tot}} - S^*) + K_m} \right).$$

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

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 2.11.

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 are therefore modified to be

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

and

$$\text{EA} + \text{S} \underset{d}{\overset{a}{\rightleftharpoons}} \text{EAS} \overset{k}{\rightarrow} \text{S}^* + \text{EA},$$

with conservation laws

$$E_{\text{tot}} = E + EA + EAS, \;\; S_{\text{tot}} \approx S + S^*.$$

The production rate of S^* is given by $dS^*/dt = kEAS$. Assuming as above that the complexes are at the quasi-steady state, we have that

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

which, using the conservation law for E, leads to

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

Hence, we have that

$$\frac{dS^*}{dt} = kE_{\text{tot}}\left(\frac{A}{A + K_{\text{d}}}\right)\left(\frac{S}{S + K_m}\right).$$

Using the conservation law for S, this is also equivalent to

$$\frac{dS^*}{dt} = kE_{\text{tot}}\left(\frac{A}{A + K_{\text{d}}}\right)\left(\frac{(S_{\text{tot}} - S^*)}{(S_{\text{tot}} - S^*) + K_m}\right).$$

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

Figure 2.11 shows the behavior of the enzyme activity as a function of the allosteric effector. As the dissociation constant decreases, that is, the affinity of the effector increases, a very small amount of effector will cause the enzyme activity to be completely "on" in the case of the activator and completely "off" 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 2.12).

Covalent modifications to proteins

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

At 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 2.12). We call X^* the activated protein. There are often allosteric

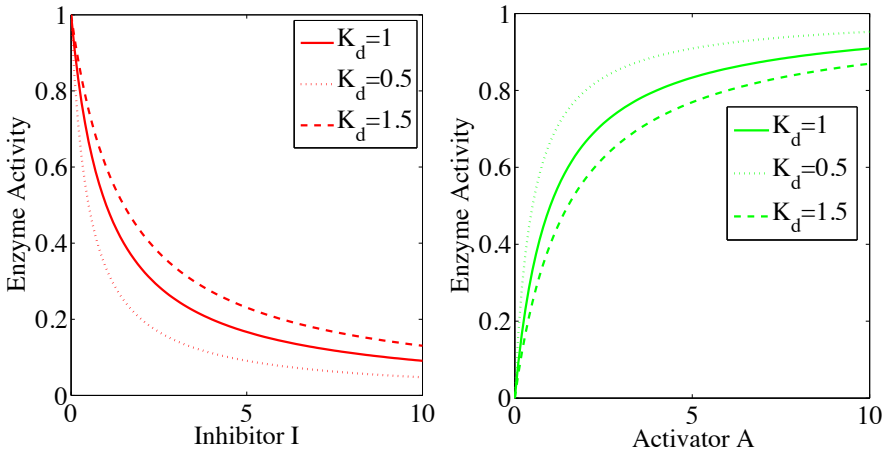


Figure 2.11: Enzyme activity in the presence of allosteric effectors (activators or inhibitors). The red plots show the enzyme activity in the presence of an inhibitor as a function of the inhibitor concentration. The green plots show the enzyme activity in the presence of an activator as a function of the activator concentration. The different plots show the effect of the dissociation constant.

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. There are several types of covalent modification, depending on the type of activation of the protein. *Phosphorylation* is a covalent modification that takes place mainly in eukaryotes and involves activation of the inactive protein X by addition of a phosphate group, PO_4. In this case, the enzyme Z is called a *kinase* while the enzyme Y is called *phosphatase*. Another type of covalent modification, which is very common in both procaryotes and eukaryotes, is *methylation*. Here, the inactive protein is activated by the addition of a methyl group, CH_3.

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

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

The corresponding ODE 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 and Y by Z_{tot}, Y_{tot}, respectively. Then, we

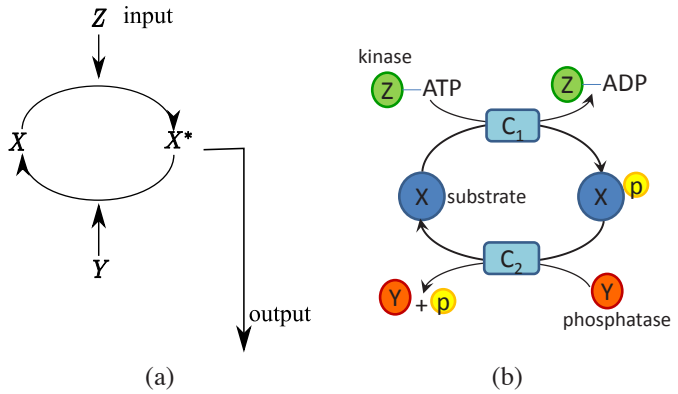


Figure 2.12: (Left) General diagram representing a covalent modification cycle. (Right) Detailed view of a phoshorylation cycle including ATP, ADP, and the exchange og the phosphate group "p".

have also the conservation laws $Z + C_1 = Z_{\text{tot}}$ and $Y + C_2 = Y_{\text{tot}}$. We can thus reduce the above system of ODE to the following one, in which we have substituted $Z = Z_{\text{tot}} - C_1$ and $Y = Y_{\text{tot}} - C_2$:

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

As for the case of the enzymatic reaction, this system cannot be analytically integrated. To simplify it, we can perform a similar approximation as done for the enzymatic reaction. In particular, the complexes C_1 and C_2 are often assumed to reach their steady state values very quickly because $a_1, d_1, a_2, d_2 \gg k_1, k_2$. 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 ODE 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

$$\frac{dX^*}{dt} = k_1 \frac{Z_{\text{tot}} X}{X + K_{m,1}} - k_2 \frac{Y_{\text{tot}} X^*}{X^* + K_{m,2}}. \tag{2.18}$$

We will come back to the modeling of this system after we have introduced singular perturbation theory, through which we will be able to perform a formal analysis and mathematically characterize the assumptions needed for approximating the original system by the first order ODE model (2.18). In the model of equation (2.18), we have that $X = X_{\text{tot}} - X^* - C_1 - C_2$ by the conservation laws. A standard assumption is that the amounts of enzymes are small compared to the amount of substrate, so that $X \approx X_{\text{tot}} - X^*$ [37].

Ultrasensitivity

One relevant aspect of the response of the covalent modification cycle to its input is the sensitivity of the steady state characteristic curve. Specifically, what parameters affect the shape of the steady state response is a crucial question. To determine the steady state characteristics, which shows how the steady state of X^* changes when the input stimulus Z_{tot} is changed, we set $dX^*/dt = 0$ in equation (2.18). Using the approximation $X \approx X_{\text{tot}} - X^*$, denoting $V_1 := k_1 Z_{\text{tot}}$, $V_2 := k_2 Y_{\text{tot}}$, $K_1 := K_{m,1}/X_{\text{tot}}$, and $K_2 := K_{m,2}/X_{\text{tot}}$, we obtain

$$y := \frac{V_1}{V_2} = \frac{X^*/X_{\text{tot}} \left(K_1 + (1 - X^*/X_{\text{tot}})\right)}{(K_2 + X^*/X_{\text{tot}})(1 - X^*/X_{\text{tot}})}. \tag{2.19}$$

We are interested in the shape of the steady state curve of X^* as function of y. This shape 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 is (Figure 2.13). 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 $X^* = y^n/(K + y^n)$, so that $y_\alpha = (\alpha/(100 - \alpha))^{(1/n)}$ and as a consequence

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

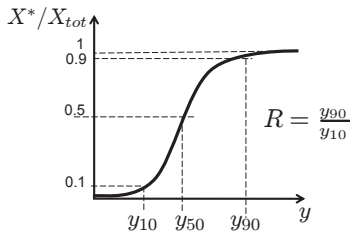


Figure 2.13: Steady state characteristic curve showing the relevance of the response coefficient for ultrasensitivity. As $R \to 1$, the points y_{10} and y_{90} tend to each other.

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*. The formula $n = \log(81)/\log(R)$ is often employed to estimate the *apparent Hill coefficient* of a dose response curve (the input/output steady state characteristic curve obtained from experimental data) since R can be calculated for any response curve directly from the data points.

In the case of the current system, from equation (2.19), we have that

$$y_{90} = \frac{(K_1 + 0.1)\, 0.9}{(K_2 + 0.9)\, 0.1} \quad \text{and} \quad y_{10} = \frac{(K_1 + 0.9)\, 0.1}{(K_2 + 0.1)\, 0.9},$$

so that

$$R = 81 \frac{(K_1 + 0.1)(K_2 + 0.1)}{(K_2 + 0.9)(K_1 + 0.9)}.$$

As a consequence, when $K_1, K_2 \gg 1$, we have that $R \to 81$, which gives a Michaelis-Menten type of response. If instead $K_1, K_2 \ll 0.1$, we have that $R \to 1$, which corresponds to a theoretic Hill coefficient $n \gg 1$, that is, a switch-like response (Figure 2.14). In particular, if we have, for example, $K_1 = K_2 = 10^{-2}$, we obtain an apparent Hill coefficient grater 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 amount of protein substrate X, 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

$$y_{50} = \frac{K_1 + 0.5}{K_2 + 0.5},$$

to find that as K_2 increases, it decreases and that as K_1 increases, it increases.

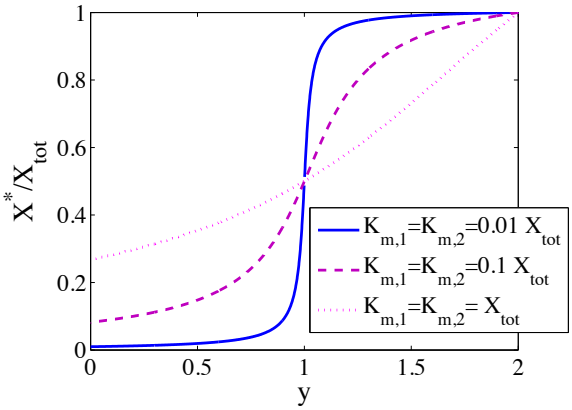


Figure 2.14: Steady state characteristics of a covalent modification cycle as a function of the Michaelis-Menten constants K_1 and K_2.

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 2.15). Each protein carrying a phosphate group can donate it to the next protein in the system through a reversible reaction. In this section, we describe a module extracted from the phosphotransferase system [91].

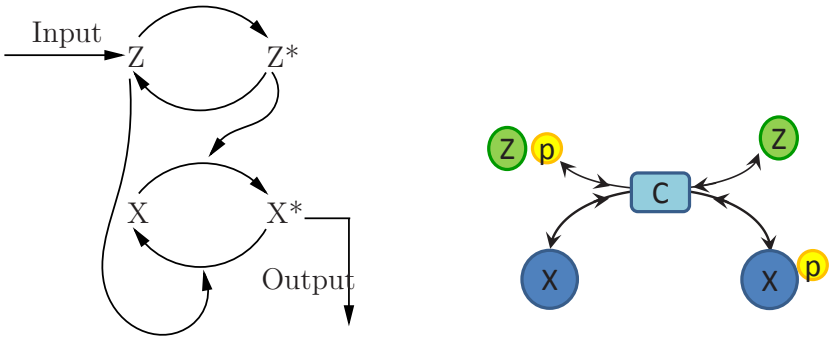
Let X be a transcription factor in its inactive form and let X^* be the same transcription factor once it has been activated by the addition of a phosphate group. Let Z^* be a phosphate donor, that is, a protein that can transfer its phosphate group to the acceptor X. The standard phosphotransfer reactions [82] 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, protein Z can be phosphorylated and protein X^* dephosphorylated by other phosphotransfer interactions. These reactions are modeled as one step reactions depending only on the concentrations of Z and X^*, that is,

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

Protein X is assumed to be conserved in the system, that is, $X_{\text{tot}} = X + C_1 + X^*$. We assume that protein Z is produced with time-varying production rate $k(t)$ and decays with rate δ. The ODE model corresponding to this system is thus given by



(a) (b)

Figure 2.15: (a) Diagram of a phosphotransfer system. (b) Proteins X and Z are transferring the phosphate group p to each other.

the equations

$$\begin{aligned}
\frac{dZ}{dt} &= k(t) - \delta Z + k_3 C_1 - k_4 X^* Z - \pi_1 Z \\
\frac{dC_1}{dt} &= k_1 X_{\text{tot}} \left(1 - \frac{X^*}{X_{\text{tot}}} - \frac{C_1}{X_{\text{tot}}}\right) Z^* - k_3 C_1 - k_2 C_1 + k_4 X^* Z \\
\frac{dZ^*}{dt} &= \pi_1 Z + k_2 C_1 - k_1 X_{\text{tot}} \left(1 - \frac{X^*}{X_{\text{tot}}} - \frac{C_1}{X_{\text{tot}}}\right) Z^* \\
\frac{dX^*}{dt} &= k_3 C_1 - k_4 X^* Z - \pi_2 X^*.
\end{aligned} \tag{2.20}$$

Sample simulation results when the input is a time-varying (periodic) stimulus are shown in Figure 2.16. The output X^* well "tracks" the input stimulus by virtue of the fast phosphotransfer reactions.

This model will be considered again in Chapter 7 when the phosphotransfer system is proposed as a possible realization of an insulation device to buffer systems from retroactivity effects.

2.5 Cellular subsystems

In the previous section we have studied how to model a variety of core processes that occure 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 2.17). The cascade consists

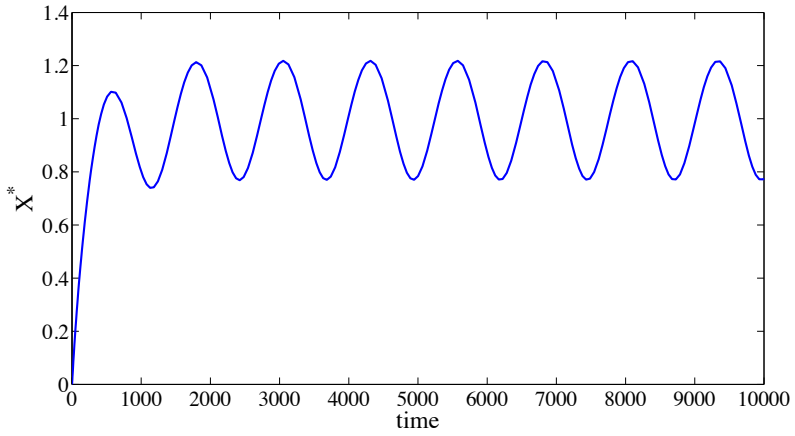


Figure 2.16: Output response of the phosphotransfer system with a step signal $k(t) = 1 + 0.5\sin(\omega t)$. The parameters are given by $\delta = 0.01, X_{\text{tot}} = 5000, k_1 = k_2 = k_3 = k_4 = \pi_1 = \pi_2 = 0.01$.

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 also phosphorylation at conserved sites. The cascade relays signals from the plasma 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 2.18. The reactions describing the system are given by

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

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

With conservation laws

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

in which we have assumed the the total amounts of enzymes are small compared to the total amount of substrate as we have explained earlier. Since $a_i, d_i \gg k_i$ and

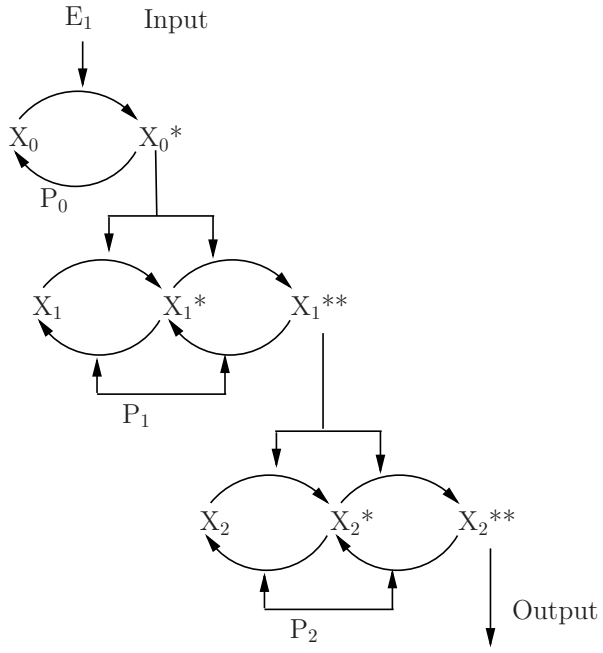


Figure 2.17: Schematic representing the MAPK cascade. It has three levels: the first one has a single phosphorylation, while the second and the third ones have a double phosphorylation.

$a_i^*, d_i^* \gg k_i^*$, we can assume that the complexes are at the quasi-steady state (i.e., $\dot{C}_i \approx 0$), which gives the Michaelis-Menten form for the amount of formed complexes:

$$C_1 = E_{1,\text{tot}} \frac{K_1^* X}{K_1^* X + K_1 X^* + K_1 K_1^*}, \qquad C_3 = E_{1,\text{tot}} \frac{K_1 X^*}{K_1^* X + K_1 X^* + K_1 K_1^*},$$
$$C_2 = E_{2,\text{tot}} \frac{K_2^* X^*}{K_2^* X^* + K_2 X^{**} + K_2 K_2^*}, \qquad C_4 = E_{2,\text{tot}} \frac{K_2 X^{**}}{K_2^* X^* + K_2 X^{**} + K_2 K_2^*},$$

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

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

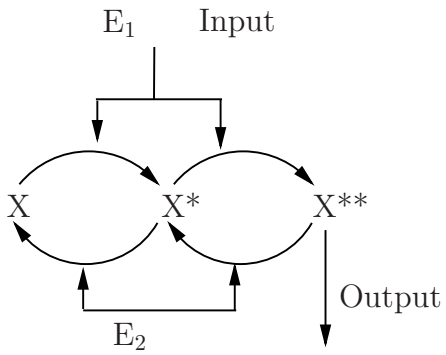


Figure 2.18: Schematic representing a double phosphorylation cycle. E_1 is the input and X^{**} is the output.

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

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

in which $X = X_{\text{tot}} - X^* - X^{**}$.

Modular model of MAPK cascades

In this section, to simplify notation, we denote "MAPK" by X_2. In a modular composition framework, the output of one stage becomes an input to the next stage downstream of it. Hence, X_0^* becomes the input enzyme that activates the phosphorylation of X_1, and X_1^{**} becomes the input enzyme that activates the phosphorylation of X_2. Let $(a_{1,i}, d_{1,i}, k_{1,i})$ and $(a_{2,i}, d_{2,i}, k_{2,i})$ be the association, dissociation, and catalytic rates for the forward and backward enzymatic reactions, respectively, for the first cycle at stage $i \in \{0,1,2\}$. Similarly, let $(a_{1,i}^*, d_{1,i}^*, k_{1,i}^*)$ and $(a_{2,i}^*, d_{2,i}^*, k_{2,i}^*)$ be the association, dissociation, and catalytic rates for the forward and backward enzymatic reactions, respectively, for the second cycle at stage $i \in \{1,2\}$. Also, denote by $K_{1,i}$ and $K_{2,i}$ for $i \in \{0,1,2\}$ the Michaelis-Menten constants of the forward and backward enzymatic reactions, respectively, of the first cycle at stage i. Similarly, denote $K_{1,i}^*$ and $K_{2,i}^*$ for $i \in \{1,2\}$ be the Michaelis-Menten constants of the forward and backward enzymatic reactions, respectively, of the second cycle at stage i. Let $P_{1,\text{tot}}$ and $P_{2,\text{tot}}$ be the total amounts of the X_1 and X_2 phosphatases,

respectively. Then, the modular ODE model of the MAPK cascade is given by

$$
\begin{aligned}
\frac{d}{dt}X_0^* &= k_{1,0}E_{1,\text{tot}}\frac{X_0}{X_0+K_{1,0}} - k_{2,0}P_{0,\text{tot}}\frac{X_0^*}{X_0^*+K_{2,0}}\\
\frac{d}{dt}X_1^* &= X_0^*\frac{k_{1,1}\ X_0\ K_{1,1}^* - k_{1,1}^*\ X_1^*\ K_{1,1}}{K_{1,1}^*\ X_1+K_{1,1}\ X_1^*+K_{1,1}K_{1,1}^*} + P_{1,\text{tot}}\frac{k_{2,1}^*\ K_{2,1}\ X_1^{**} - k_{2,1}\ X_1^*\ K_{2,1}^*}{K_{2,1}^*\ X_1^*+K_{2,1}\ X_1^{**}+K_{2,1}K_{2,1}^*}\\
\frac{d}{dt}X_1^{**} &= k_{1,1}^*\ X_0^*\frac{X_1^*\ K_{1,1}}{K_{1,1}^*\ X_1+K_{1,1}\ X_1^*+K_{1,1}K_{1,1}^*} - k_{2,1}^*\ P_{1,\text{tot}}\frac{X_1^{**}\ K_{2,1}}{K_{2,1}^*\ X_1^*+K_{2,1}\ X_1^{**}+K_{2,1}K_{2,1}^*}\\
\frac{d}{dt}X_2^* &= X_1^{**}\frac{k_{1,2}X_2\ K_{1,2}^* - k_{1,2}^*\ X_2^*\ K_{1,2}}{K_{1,2}^*\ X_2+K_{1,2}\ X_2^*+K_{1,2}^*\ K_{1,2}} + P_{2,\text{tot}}\frac{k_{2,2}^*\ K_{2,2}\ X_2^{**} - k_{2,2}\ X_2^*\ K_{2,2}^*}{K_{2,2}^*\ X_2^*+K_{2,2}\ X_2^{**}+K_{2,2}\ K_{2,2}^*}\\
\frac{d}{dt}X_2^{**} &= k_{1,2}^*\ X_1^{**}\frac{X_2^*\ K_{1,2}}{K_{1,2}^*\ X_2+K_{1,2}\ X_2^*+K_{1,2}^*\ K_{1,2}} - k_{2,2}^*\ P_{2,\text{tot}}\frac{X_2^{**}\ K_{2,2}}{K_{2,2}^*\ X_2^*+K_{2,2}\ X_2^{**}+K_{2,2}\ K_{2,2}^*}
\end{aligned}
\tag{2.21}
$$

in which, letting $X_{0,\text{tot}}, X_{1,tot}$ and $X_{2,\text{tot}}$ represent the total amounts of each stage protein, we have $X_0 = X_{0,\text{tot}} - X_0^*$, $X_1 = X_{1,\text{tot}} - X_1^* - X_1^{**}$ and $X_2 = X_{2,\text{tot}} - X_2^* - X_2^{**}$.

Mechanistic model of the MAPK cascade

We now give the entire set of reactions for the MAPK cascade of Figure 2.17 as they are found in standard references (Huang-Ferrell model [45]):

$$
\begin{aligned}
&\mathrm{E_1} + \mathrm{X_0} \underset{d_{1,0}}{\overset{a_{1,0}}{\rightleftharpoons}} \mathrm{C_1} \xrightarrow{k_{1,0}} \mathrm{X_0^* E_1} && \mathrm{P_0} + \mathrm{X_0^*} \underset{d_{2,0}}{\overset{a_{2,0}}{\rightleftharpoons}} \mathrm{C_2} \xrightarrow{k_{2,0}} \mathrm{X_0^* + P_0}\\
&\mathrm{X_0^*} + \mathrm{X_1} \underset{d_{1,1}}{\overset{a_{1,1}}{\rightleftharpoons}} \mathrm{C_3} \xrightarrow{k_{1,1}} \mathrm{X_1^* + X_0^*} && \mathrm{X_1^*} + \mathrm{P_1} \underset{d_{2,1}}{\overset{a_{2,1}}{\rightleftharpoons}} \mathrm{C_4} \xrightarrow{k_{2,1}} \mathrm{X_1 + P_1}\\
&\mathrm{X_0^*} + \mathrm{X_1^*} \underset{d_{1,1}^*}{\overset{a_{1,1}^*}{\rightleftharpoons}} \mathrm{C_5} \xrightarrow{k_{1,1}^*} \mathrm{X_1^{**} + X_0^*} && \mathrm{X_1^*} + \mathrm{P_1} \underset{d_{2,1}^*}{\overset{a_{2,1}^*}{\rightleftharpoons}} \mathrm{C_6} \xrightarrow{k_{2,1}^*} \mathrm{X_1^* + P_1}\\
&\mathrm{X_1^{**}} + \mathrm{X_2} \underset{d_{1,2}}{\overset{a_{1,2}}{\rightleftharpoons}} \mathrm{C_7} \xrightarrow{k_{1,2}} \mathrm{X_2^* + X_1^{**}} && \mathrm{X_2^*} + \mathrm{P_2} \underset{d_{2,2}}{\overset{a_{2,2}}{\rightleftharpoons}} \mathrm{C_8} \xrightarrow{k_{2,2}} \mathrm{X_2 + P_2}\\
&\mathrm{X_1^{**}} + \mathrm{X_2^*} \underset{d_{1,2}^*}{\overset{a_{1,2}^*}{\rightleftharpoons}} \mathrm{C_9} \xrightarrow{k_{1,2}^*} \mathrm{X_2^{**} + X_1^*} && \mathrm{X_2^{**}} + \mathrm{P_2} \underset{d_{2,2}^*}{\overset{a_{2,2}^*}{\rightleftharpoons}} \mathrm{C_{10}} \xrightarrow{k_{2,2}^*} \mathrm{X_2^* + P_2},
\end{aligned}
$$

with conservation laws

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

The corresponding ODE model is given by

$$
\begin{aligned}
\frac{d}{dt}C_1 &= a_{1,0}E_1\ X_0 - (d_{1,0}+k_{1,0})\ C_1\\
\frac{d}{dt}X_0^* &= k_{1,0}\ C_1 + d_{2,0}\ C_2 - a_{2,0}\ P_0\ X_0^* + (d_{1,1}+k_{1,1})\ C_3 - a_{1,1}\ X_1\ X_0^*\\
&\quad + (d_{1,1}^*+k_{1,1}^*)\ C_5 - a_{1,1}^*\ X_0^*\ X_1^*\\
\frac{d}{dt}C_2 &= a_{2,0}\ P_0\ X_0^* - (d_{2,0}+k_{2,0})\ C_2\\
\frac{d}{dt}C_3 &= a_{1,1}\ X_1\ X_0^* - (d_{1,1}+k_{1,1})\ C_3\\
\frac{d}{dt}X_1^* &= k_{1,1}\ C_3 + d_{2,1}\ C_4 - a_{2,1}\ X_1^*\ P_1 + d_{1,1}^* C_5 - a_{1,1}^*\ X_1^*\ X_0^* + k_{2,1}^*\ C_6\\
\frac{d}{dt}C_4 &= a_{2,1}\ X_1^*\ P_1 - (d_{2,1}+k_{2,1})\ C_4\\
\frac{d}{dt}C_5 &= a_{1,1}^*\ X_0^*\ X_1^* - (d_{1,1}^*+k_{1,1}^*)\ C_5\\
\frac{d}{dt}X_1^{**} &= k_{1,1}^*\ C_5 - a_{2,1}^*\ X_1^*\ P_1 + d_{2,1}^*\ C_6 - a_{1,2}\ X_1^{**}\ X_2\\
&\quad + (d_{1,2}+k_{1,2})\ C_7 - a_{1,2}^*\ X_1^{**}\ X_2^* + (d_{1,2}^*+k_{1,2}^*)\ C_9\\
\frac{d}{dt}C_6 &= a_{2,1}^*\ X_1^{**}\ P_1 - (d_{2,1}^*+k_{2,1}^*)\ C_6\\
\frac{d}{dt}C_7 &= a_{1,2}^*\ X_1^*\ X_2 - (d_{1,2}^*+k_{1,2}^*)\ C_7\\
\frac{d}{dt}X_2^* &= -a_{2,2}\ X_2^*\ P_2 + d_{2,2}\ C_8 - a_{1,2}^*\ X_2^*\ X_2^{**} + d_{1,2}^*\ C_9 + C_{10}\ K_{10}\\
\frac{d}{dt}C_8 &= a_{2,2}^*\ X_2^*\ P_2 - (d_{2,2}+k_{2,2})\ C_8\\
\frac{d}{dt}X_2^{**} &= k_{1,2}^*\ C_9 - a_{2,2}^*\ X_2^{**}\ P_2 + d_{2,2}^*\ C_{10}\\
\frac{d}{dt}C_9 &= a_{1,2}^*\ X_1^{**}\ X_2^* - (d_{1,2}^*+k_{1,2}^*)\ C_9\\
\frac{d}{dt}C_{10} &= a_{2,2}^*\ X_2^{**}\ P_2 - (d_{2,2}^*+k_{2,2}^*)\ C_{10}.
\end{aligned}
$$

Assuming as before that the total amounts of enzymes are much smaller than the total amounts of substrates ($E_{1,\mathrm{tot}}, P_{0,\mathrm{tot}}, P_{1,\mathrm{tot}}, P_{2,\mathrm{tot}} \ll X_{0,\mathrm{tot}}, X_{1,\mathrm{tot}}, X_{2,\mathrm{tot}}$), we can approximate the conservation laws as

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

Using these and assuming that the complexes are at the quasi-steady state, we obtain 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 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 impedance in electrical systems and to back-effect in mechanical systems. It will be studied at length in Chapter 7.

This fact is in clear contrast with the ODE model obtained by modular composition, in which each stage dynamics depended upon the variables of the upstream stages and not upon those of the downstream stages. That is, from equations (2.21), it is apparent that the dynamics of X_0^* (first stage) do not depend on the variables of the second stage (X_1, X_1^*, X_1^{**}). In turn, the dynamics of X_1^* and X_1^{**} (second stage) do not depend on the variables of the third stage (X_2^* and X_2^{**}). Indeed modular composition does not consider the fact that the proteins of each stage are "used-up" in the process of transmitting information to the downstream stages. This backward effect has been theoretically shown to lead to sustained oscillations in the MAPK cascade [80]. By contrast, the modular ODE model of MAPK cascades does not give rise to sustained oscillations.

Properties of the MAPK Cascade

The stimulus-response 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 increases from the MAPKK response to the MAPK response. That is, an increase ultrasensitivity is observed moving down in the cascade (Figure 2.19). These model observations persist when key parameters, such as the Michaelis-Menten constants are changed [45]. Furthermore, zero-order ultrasensitivity effects can be observed. Specifically, 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

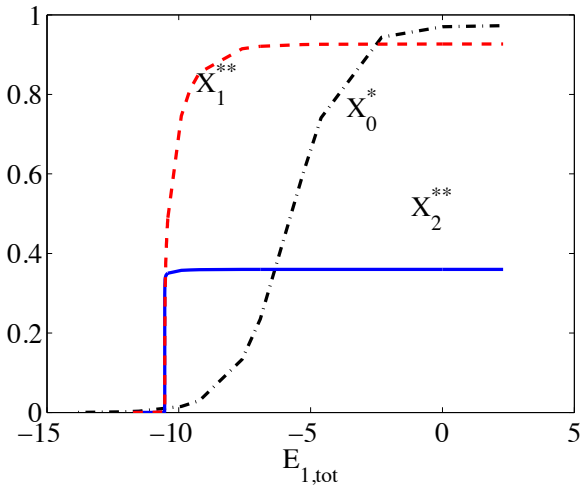


Figure 2.19: Dose response of the MAPK cascade for every stage. Simulations from the model of [80].

(similar to single phosphorylation) predicted substantially lower apparent Hill coefficients.

Additional topics to be added later: Review

1. Transport across the membrane
2. Membrane receptors, ligand binding, G-proteins

Exercises

2.1 (BE 150, Winter 2011) 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 γ. All proteins have the same degradation/dilution rate α.

(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 response time of protein Z with respect to the time of addition of S_x?

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

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

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

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

and give expressions for α and K.

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

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

the steady state values of C and A are

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

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

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

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

2.4 Consider the following modification of the competitive binding reactions:

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

and

$$\mathrm{C} + \mathrm{B_r} \underset{k_r'}{\overset{k_f'}{\rightleftharpoons}} \mathrm{C}', \text{ and } C + \mathrm{B_a} \underset{k_r'}{\overset{k_f'}{\rightleftharpoons}} \mathrm{C}'$$

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

2.5 Consider the case of a competitive binding of an activator A and a repressor R with D and assume that before they can bind D they have to cooperatively bind according to the following reactions:

$$\mathrm{A} + \mathrm{A} + ... + \mathrm{A} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{A_n}, \qquad \mathrm{R} + \mathrm{R} + ... + \mathrm{R} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{R_m},$$

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

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

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

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

2.7 (BE 150, Winter 2011) Consider a positive transcriptional feedback loop composed of two negative interactions $X \dashv Y$ and $Y \dashv X$.

(a) Write the ODEs for the system above. Assume that the two transcrption/repression mechanisms have the same dynamics and both genes are degraded at the same rate 0.2. Let the basal transcription rate be 1, $K = 2, n = 2$.

(b) To solve for the steady states, plot the *nullclines* by solving $\frac{dX}{dt} = 0$ and $\frac{dY}{dt} = 0$ (i.e. solve for $Y = g_1(X)$ where $\frac{dX}{dt} = 0$ and $Y = g_2(X)$ where $\frac{dY}{dt} = 0$ and plot both solutions). The steady states are given by the intersections of the two nullclines.

(c) Plot the time response of X and Y using the following two initial conditions:

$$(X(0), Y(0)) = (1, 4) \quad \text{and} \quad (4, 1).$$

Next, plot the phase plane of the system using *pplane* in MATLAB. How do the responses change with initial conditions? Describe a situation where this type of interaction would be useful.

2.8 Consider the phosphorylation reactions described in Section 2.4, but suppose that the kinase concentration Z is not constant, but is produced and decays according to the reaction $\mathrm{Z} \underset{k(t)}{\overset{\delta}{\rightleftharpoons}} \emptyset$. How should the system in equation (2.18) be modified?

Use a MATLAB simulation to apply a periodic input stimulus $k(t)$ using parameter values: $k_{\text{cat}} = k'_{\text{cat}} = 1$, $k_f = k'_f = k_r = k'_r = 10$, $\delta = 0.01$. Is the cycle capable of "tracking" the input stimulus? If yes, to what extent? What are the tracking properties depending on?

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

2.10 (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

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

2.11 (Competitive Inhibition) Derive the expression of the production rate of W^* in the presence of a competitive inhibitor I.

2.12 (Non-absolute activator) Derive the expression of the production rate of W^* in the presence of a non-absolute activator A.

2.13 (BE 150, Winter 2011) Consider the following network $X \to Y$ and $X \to X$.

(a) Write the ODEs for the system above. Use basal expression $\beta_X = \beta_Y = 2$ and activation coefficients $K_X = 1$, $K_Y = 2$, $n_1 = n_2 = 2$. The degradation coefficients for X and Y are both 0.5.

(b) Plot the vector field using pplane. How many steady states do you observe?

(c) Solve for the steady states of the system using the derived ODEs, linearize the system and do a stability analysis.