Biomolecular Feedback Systems

Domitilla Del Vecchio
U. Michigan/MIT

Richard M. Murray
Caltech

DRAFT v0.3, January 10, 2010

Chapter 2
Core Processes

The goal of this chapter is to describe basic biological mechanisms in a way that can be represented by simple dynamic models. We begin the chapter with a overview of the dynamics of protein production and control, focused on the processes that determine the properties of genetic networks, followed by a discussion of the basic modeling formalisms that we will utilize. 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 is on deterministic models using ordinary differential equations; Chapter 4 describes how to model the stochastic nature of biomolecular systems.

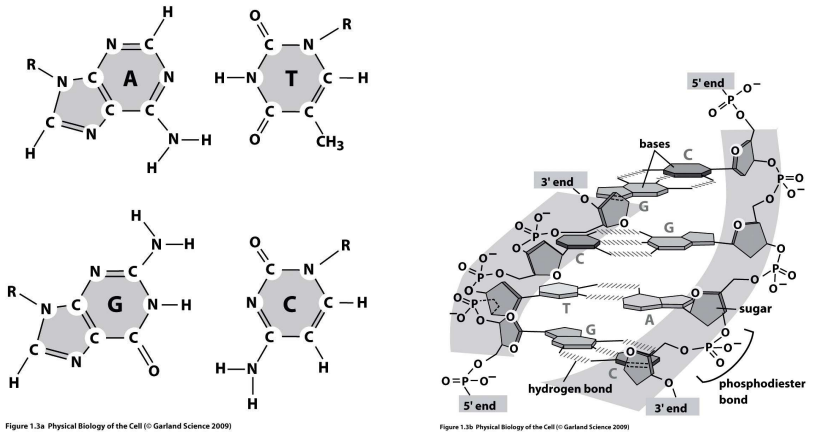
Prerequisites. Readers should have a basic understanding of ordinary differential equations, at the level of Chapter 2 of AM08, and some basic familiarity with cell biology, at the level of the description in Chapter 1.

2.1 Dynamics and Control in the Cell

The molecular processes inside a cell determine its behavior and are responsible for metabolizing nutrients, generating motion, enabling procreation and carrying out the other functions of the organism. In complex, multi-cellular organisms, different types of cells work together to enable more complex functions. In this chapter we briefly describe the role of dynamics and control within a cell and discuss the basic processes that govern its behavior and its interactions with its environment (including other cells). We build on the description of cell biology provided in Chapter 1; a much more detailed introduction to the biology of the cell and some of the processes described here can be found in standard textbooks on cell biology such as Alberts *et al*. [2] or Phillips *et al*. [28].

The central dogma: production of proteins

The genetic material inside a cell, encoded in its DNA, governs the response of a cell to various conditions. DNA is organized into collections of genes, with each gene encoding a corresponding protein that performs a set of functions in the cell. The activation and repression of genes are determined through a series of complex interactions that give rise to a remarkable set of circuits that perform the functions required for life, ranging from basic metabolism to locomotion to procreation.



(a) Base pairs (b) Double stranded

Figure 2.1: Molecular structure of DNA. (a) Individual bases (nucleotides) that make up DNA: adenine (A), cytocine (C), guanine (G) and thymine (T). (b) Double stranded DNA formed from individual nucleotides, with A binding to T and C binding to G. Each strand contains a 5' and 3' end, determined by the locations of the carbons where the next nucleotide binds. Figure from Phillips, Kondev and Theriot [28]; used with permission of Garland Science.

Genetic circuits that occur in nature are robust to external disturbances and can function in a variety of conditions. To understand how these processes occur (and some of the dynamics that govern their behavior), it will be useful to present a slightly more detailed description of the underlying biochemistry involved in the production of proteins.

DNA is double stranded molecule with the "direction" of each strand specified by looking at the geometry of the sugars that make up its backbone (see Figure 2.1). The complementary strands of DNA are composed of a sequence of nucleotides that consist of a sugar molecule (deoxyribose) bound to one of 4 bases: adenine (A), cytocine (C), guanine (G) and thymine (T). The coding strand (by convention the top row of a DNA sequence when it is written in text form) is specified from the 5' end of the DNA to the 3' end of the DNA. (As described briefly in Chapter 1, 5' and 3' refer to carbon locations on the deoxyribose backbone that are involved in linking together the nucleotides that make up DNA.) The DNA that encodes proteins consists of a promoter region, regulator regions (described in more detail below), a coding region and a termination region (see Figure 2.2).

RNA polymerase enzymes are present in the nucleus (for eukaryotes) or cytoplasm (for prokaryotes) and must localize and bind to the promoter region of the DNA template. Once bound, the RNA polymerase "opens" the double stranded DNA to expose the nucleotides that make up the sequence, as shown in Figure 2.3. This reversible reaction, called *isomerization*, is said to transform the RNA polymerase and DNA from a *closed complex* to an *open complex*. After the open complex is formed, RNA polymerase begins to travel down the DNA strand and constructs an mRNA sequence that matches the 5' to 3' sequence of the DNA to which

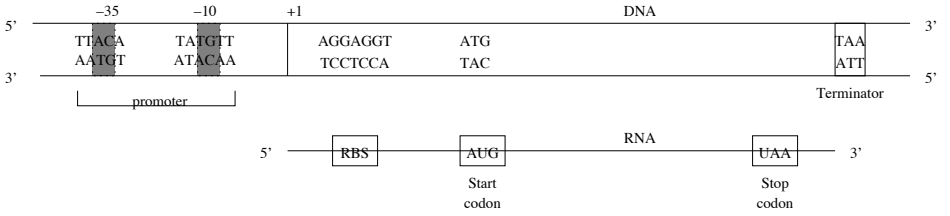


Figure 2.2: Geometric structure of DNA. The layout of the DNA is shown at the top. RNA polymerase binds to the promoter region of the DNA and transcribes the DNA starting at the +1 side and continuing to the termination site.

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

The RNA strand that is produced by RNA polymerase is also a sequence of nucleotides with a sugar backbone. The sugar for RNA is ribose instead of deoxyribose and mRNA typically exists as a single stranded molecule. Another difference is that the base thymine (T) is replaced by uracil (U) in RNA sequences. RNA polymerase produces RNA one base pair at a time, as it moves from in the 5' to 3' direction along the DNA coding strand. RNA polymerase stops transcribing DNA when it reaches a *termination region* (or *terminator*) on the DNA. This termination region consists of a sequence that causes the RNA polymerase to unbind from the DNA. The sequence is not conserved across species and in many cells the termination sequence is sometimes "leaky", so that transcription will occasionally occur across the terminator (we will see examples of this in the λ phage circuitry described in the next chapter).

Once the mRNA is produced, it must be translated into a protein. This process is slightly different in prokaryotes and eukaryotes. In prokaryotes, there is a region of the mRNA in which the ribosome (a molecular complex consisting of of both proteins and RNA) binds. This region, called the *ribosome binding site (RBS)*, has some variability between different cell species and between different genes in a given cell. The Shine-Delgarno sequence, AGGAGG, is the consensus sequence for the RBS.

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

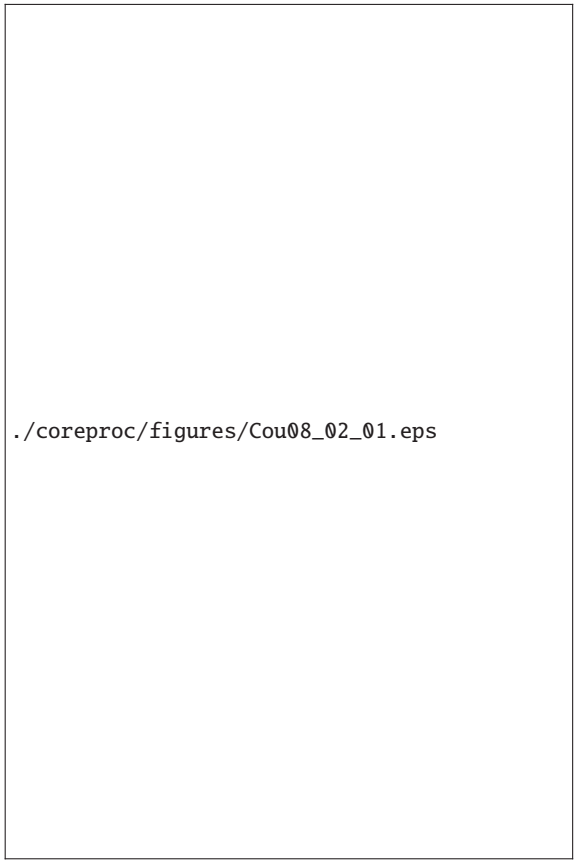


Figure 2.3: Production of messenger RNA from DNA. RNA polymerase, along with other accessory factors, binds to the promoter region of the DNA and then "opens" the DNA to begin transcription (initiation). As RNA polymerase moves down the DNA, producing an RNA transcript (elongation), which is later translated into a protein. The process ends when the RNA polymerase reaches the terminator (termination). Reproduced from Courey [9]; permission pending.

ported out of the nucleus into the cytoplasm, where the ribosomes can bind to it.

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

Once the ribosome is bound to the mRNA, it begins the process of translation. Proteins consist of a sequence of amino acids, with each amino acid specified by a codon that is used by the ribosome in the process of translation. Each codon consists of three base pairs and corresponds to one of the 20 amino acids or a "stop" codon. The genetic code mapping between codons and amino acids is shown in Table 1.1. The ribosome translates each codon into the corresponding amino acid using transfer RNA (tRNA) to integrate the appropriate amino acid (which binds

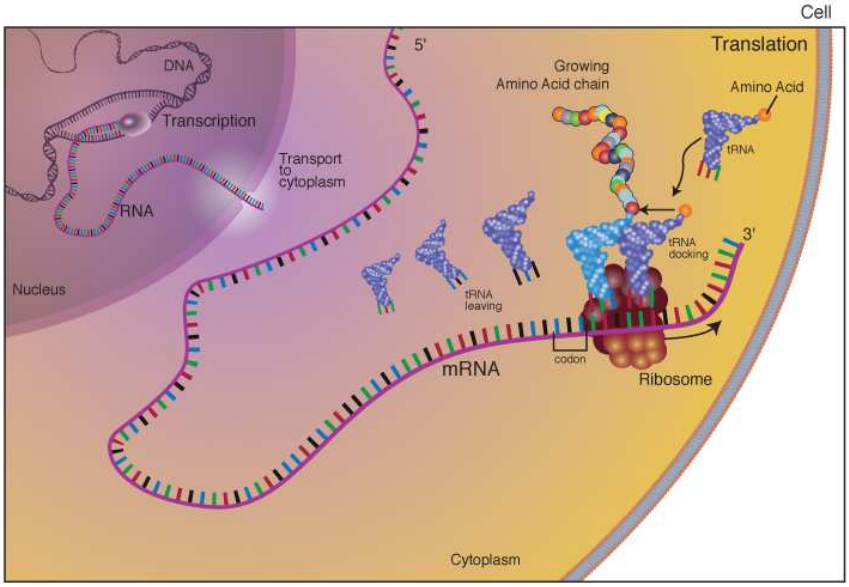


Figure 2.4: Translation is the process of translating the sequence of a messenger RNA (mRNA) molecule to a sequence of amino acids during protein synthesis. The genetic code describes the relationship between the sequence of base pairs in a gene and the corresponding amino acid sequence that it encodes. In the cell cytoplasm, the ribosome reads the sequence of the mRNA in groups of three bases to assemble the protein. Figure and caption courtesy the National Human Genome Research Institute.

to the tRNA) into the polypeptide chain, as shown in Figure 2.4. The start codon (AUG) specifies the location at which translation begins, as well as coding for the amino acid methionine (a modified form is used in prokaryotes). All subsequent codons are translated by the ribosome into the corresponding amino acid until it reaches one of the stop codons (typically UAA, UAG and UGA).

The sequence of amino acids produced by the ribosome is a polypeptide chain that folds on itself to form a protein. The process of folding is complicated and involves a variety of chemical interactions that are not completely understood. Additional post-translational processing of the protein can also occur at this stage, until a folded and functional protein is produced. It is this molecule that is able to bind to other species in the cell and perform the chemical reactions that underly the behavior of the organism.

Each of the processes involved in transcription, translation and folding of the protein takes time and affects the dynamics of the cell. Table 2.1 shows the rates of some of the key processes involved in the production of proteins. It is important to note that each of these steps is highly stochastic, with molecules binding together based on some propensity that depends on the binding energy but also the other molecules present in the cell. In addition, although we have described everything as a sequential process, each of the steps of transcription, translation and folding

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

Process	Characteristic rate	Source
mRNA production	10–30 bp/sec	Vogel and Jensen
Protein production	10–30 aa/sec	PKT08
Protein folding	???	
mRNA half life	~ 100 sec	YM03
Cell division time	~ 3000 sec	???
Protein half life	$\sim 5 \times 10^4$ sec	YM03
Protein diffusion along DNA	up to 10^4 bp/sec	

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

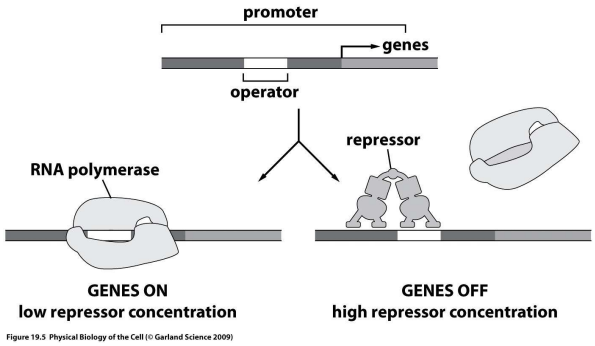
Transcriptional regulation of protein production

There are a variety of mechanisms in the cell to regulate the production of proteins. These regulatory mechanisms can occur at various points in the overall process that produces the protein. *Transcriptional regulation* refers to regulatory mechanisms that control whether or not a gene is transcribed.

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

A common mechanism for repression is that a protein binds to a region of DNA near the promoter and blocks RNA polymerase from binding. The region of DNA in which the repressor protein binds is called an *operator region* (see Figure 2.2. If the operator region overlaps the promoter, then the presence of a protein at the promoter "blocks" the DNA at that location and transcription cannot initiate, as illustrated in Figure 2.5a. Repressor proteins often bind to DNA as dimers or pairs of dimers (effectively tetramers). Figure 2.5b shows some examples of repressors bound to DNA.

A related mechanism for repression is *DNA looping*. In this setting, two repressor complexes (often dimers) bind in different locations on the DNA and then bind to each other. This can create a loop in the DNA and block the ability of RNA polymerase to bind to the promoter, thus inhibiting transcription. Figure 2.6 shows an example of this type of repression, in the *lac* operon. (An *operon* is a set



(a) Repression of gene expression

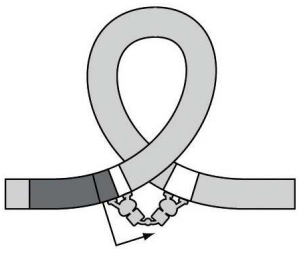
./coreproc/figures/PKT08_19_06.eps

(b) Examples of repressors

Figure 2.5: Repression of gene expression. Figure from Phillips, Kondev and Theriot [28]; used with permission of Garland Science.

of genes that is under control of a single promoter; this is discussed in more detail below.)

A feature that is present in some types of repressor proteins is the existence of an *inducer molecule* that combines with the protein to either activate or inactivate its repression function. A *positive inducer* is a molecule that must be present in order for repression to occur. A *negative inducer* is one in which the presence of the inducer molecule blocks repression, either by changing the shape of the repressor protein or by blocking active sites on the repressor protein that would normally bind to the DNA. Figure 2.7a summarizes the various possibilities. Common examples of repressor-inducer pairs include *lacI* and lactose (or IPTG), *tetR* and ATc, and tryptophan repressor and tryptophan. Lactose/IPTG and ATc are both negative



(a) DNA looping

./coreproc/figures/PKT08_08_19.eps

(b) *lac* repressor

Figure 2.6: Repression via DNA looping. Figure from Phillips, Kondev and Theriot [28]; used with permission of Garland Science.

Figure 2.7: Effects of inducers. Reproduced from Alberts et al. [2]; permission pending.

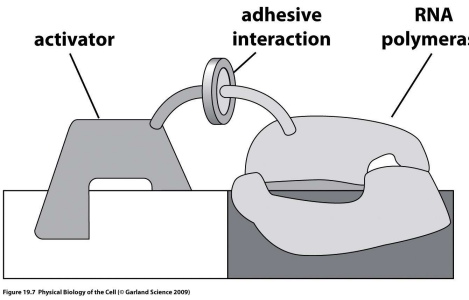
inducers, so their presence causes the otherwise repressed gene to be expressed, while tryptophan is a positive inducer.

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

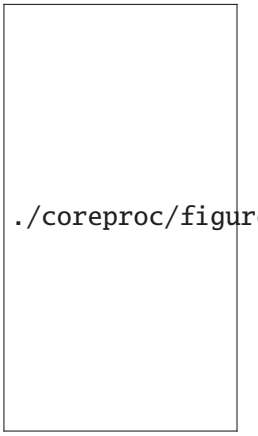
The simplest form of activation involves a protein binding to the DNA near the promoter in such a way that the combination of the activator and the promoter sequence bind RNA polymerase. One of the most well-studied examples is the *catabolite activator protein (CAP)*—also sometimes called the *cAMP receptor protein (CRP)*—shown in Figure 2.8. Like repressors, many activators have inducers, which can act in either a positive or negative fashion (see Figure 2.7b). For example, cyclic AMP (cAMP) acts as a positive inducer for CAP.

Another mechanism for activation of transcription, specific to prokaryotes, is the use of *sigma factors*. Sigma factors are part of a modular set of proteins that bind to RNA polymerase and form the molecular complex that performs transcription. Different sigma factors enable RNA polymerase to bind to different promoters, so the sigma factor acts as a type of activating signal for transcription. Table 2.2 lists some of the common sigma factors in bacteria. One of the uses of sigma factors is to produce certain proteins only under special conditions, such as when the cell undergoes *heat shock* (discussed in more detail in Chapter 5). Another use is to control the timing of the expression of certain genes, as illustrated in Figure 2.9.

In addition to repressors and activators, many genetic circuits also make use of *combinatorial promoters* that can act as either repressors or activators for genes. This allows genes to be switched on and off based on more complex conditions,



(a) Activation mechanism



(b) Examples of activators

Figure 2.8: Activation of gene expression. Figure from Phillips, Kondev and Theriot [28]; used with permission of Garland Science.

represented by the concentrations of two or more activators or repressors.

Figure 2.10 shows one of the classic examples, a promoter for the *lac* system. In the *lac* system, the expression of genes for metabolizing lactose are under the control of a single (combinatorial) promoter. CAP, which is positively induced by cAMP, acts as an activator and LacI (also called "repressor"), which is negatively induced by lactose, acts as a repressor. In addition, the inducer cAMP is expressed only when glucose levels are low. The resulting behavior is that the proteins for metabolizing lactose are expressed only in conditions where there is no glucose (so CAP is active) *and* lactose is present.

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

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

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

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

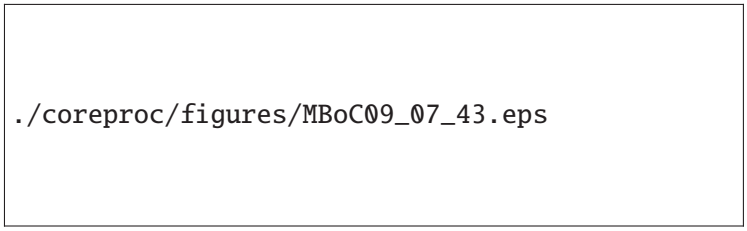


Figure 2.9: Use of sigma factors to controlling the timing of expression. Reproduced from Alberts et al. [2]; permission pending.

is present, the RNA polymerase maintains (or regains) its contact with the DNA and expression of the downstream genes is enhanced. In this way, antitermination allows downstream genes to be regulated by repressing "premature" termination. An example of an antitermination protein is the protein N in phase λ, which binds to a region of DNA labeled Nut (for N utilization) [16], as shown in Figure 2.11.

Post-transcriptional regulation of protein production

Post-translation regulation of protein activity

One of the most common types of post-transcriptional regulation is through the *phosphorylation* of proteins. Phosphorylation is an enzymatic process in which a phosphate group is added to a protein and the resulting conformation of the protein changes, usually from an inactive configuration to an active one. The enzyme that adds the phosphate group is called a *phosphotransferase* or a *kinase* and it operates by transferring a phosphate group from a bound ATP molecule to the protein, leaving behind ADP and the phosphorylated protein. *Dephosphorylation* is a complementary enzymatic process that can remove a phosphate group from a protein. The enzyme that performs dephosphorylation is called a *phosphotase*. Figure 2.12 shows the process of phosphorylation in more detail.

Phosphorylation is often used as a regulatory mechanism with the phosphorylated version of the protein being the active conformation. Since phosphorylation and dephosphorylation can occur much more quickly than protein production and degradation, it is used in my biological circuits in which a rapid response is required. One common motif is that a signaling protein will bind to a ligand and the resulting allosteric change allows the signaling protein to serve as a kinase. The newly active kinase then phosphorylates a second protein, which modulates other functions in the cell. Phosphorylation cascades can also be used to amplify the effect of the original signal; we will describe this in more detail in Section 2.6.

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

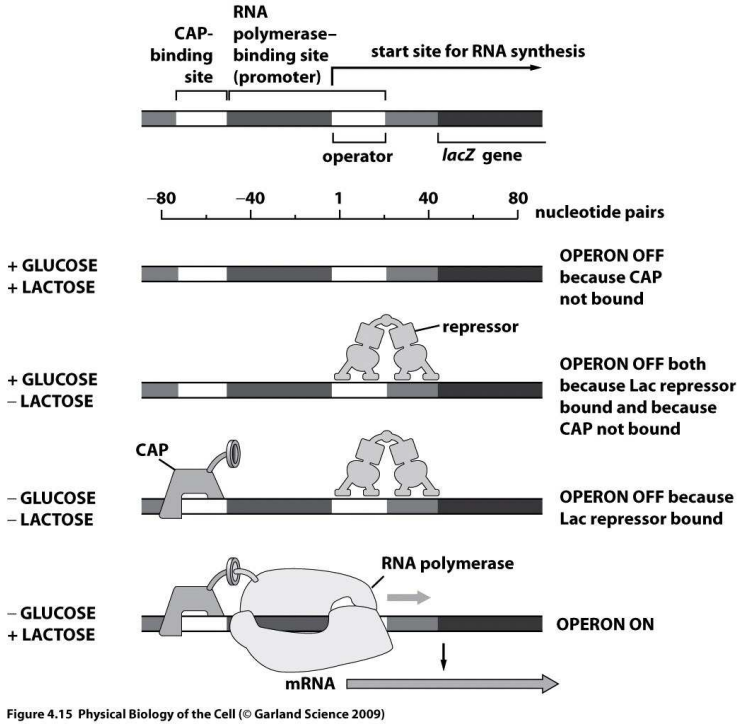


Figure 2.10: Combinatorial logic for the *lac* operator. Figure from Phillips, Kondev and Theriot [28]; used with permission of Garland Science.

proteins, and the protein is activated only when sufficient kinase activity is present.

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

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

Figure 2.11: Antitermination. Reproduced from [**?**]; permission pending.

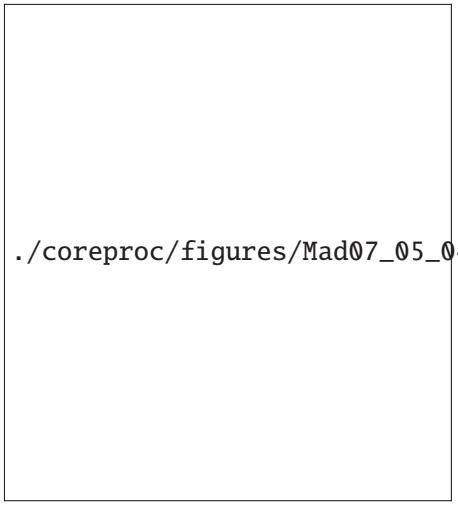


Figure 2.12: Phosphorylation of a protein via a kinase. Reproduced from Madhani [22]; permission pending.

2.2 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 2.13. The choice of which model to use depends on the questions that you want to answer, and good modeling takes practice, experience and iteration. One 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 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.

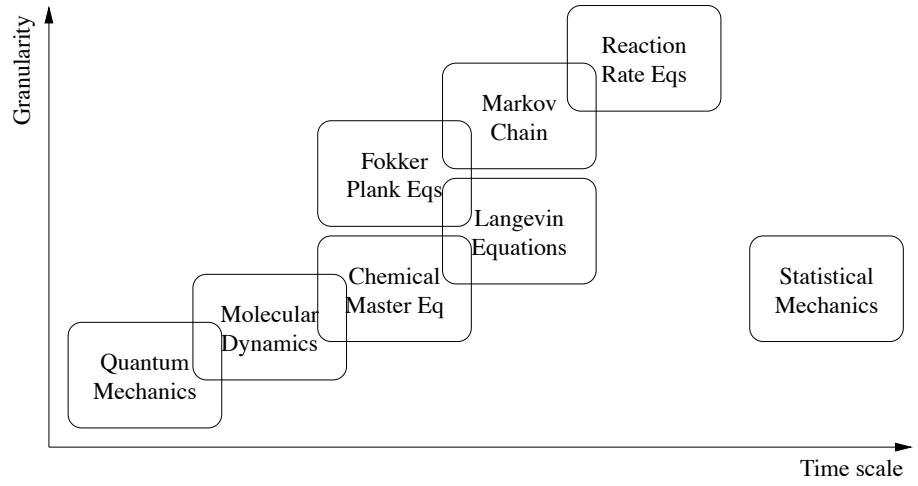


Figure 2.13: Different methods of modeling biomolecular systems.

Statistical mechanics and chemical kinetics

At the fine end of the modeling scale depicted in Figure 2.13, we can attempt to model the *molecular dynamics* of the cell, in which we attempt to model the individual proteins and other species and their interactions via molecular-scale forces and motions. At this scale, the individual interactions between protein domains, DNA and RNA are resolved, resulting in a highly detailed model of the dynamics of the cell.

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

The underlying representation for both statistical mechanics and chemical kinetics is to identify the appropriate microstates of the system. A microstate corresponds to a given configuration of the components (species) in the system relative to each other and we must enumerate all possible configurations between the molecules that are being modeled. As an example, consider the distribution of RNA 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 2.14.

Figure 2.14: 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 [28]; used with permission of Garland Science.

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

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

where E_q is the energy associated with the microstate $q \in Q$ and Z is a normalizing factor, known as the *partition function*,

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

By keeping track of those microstates that correspond to a given system state (also called a macrostate), we can compute the overall probability that a given macrostate is reached. This can be used, for example, to compute the probability that some RNA polymerase is bound to a given promoter, averaged over many independent samples, and from this we can reason about the rate of expression of the corresponding gene.

Statistical mechanics averages about 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. We describe the kinetics of the system by making use of the *propensity function* $a(\xi; q, t)$, which captures the instantaneous probability that a system will transition between state q and state $q + \xi$. More specifically, the propensity function is defined such

that

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

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

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

$$\frac{d}{dt}P(q,t) = \sum_{\xi} a(\xi;q-\xi,t)P(q-\xi,t) - \sum_{\xi} a(\xi;q,t)P(q,t). \tag{2.2}$$

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

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

$$\mathrm{A}+\mathrm{B} \rightleftharpoons \mathrm{AB} \quad \equiv \quad \begin{array}{ll} R_f: & \mathrm{A}+\mathrm{B} \longrightarrow \mathrm{AB} \\ R_r: & \mathrm{AB} \longrightarrow \mathrm{A}+\mathrm{B}. \end{array} \tag{2.3}$$

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

$$a(\xi^{\mathrm{f}};q) = c_{\xi^{\mathrm{f}}} n_{\mathrm{A}} n_{\mathrm{B}},$$

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

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

where ξ^{r} represents the reverse reaction, $c_{\xi^{\mathrm{r}}}$ is a constant coefficient and n_{AB} is the number of molecules of AB that are present.

The primary difference between the statistical mechanics description in equation (2.1) and the chemical kinetics description in equation (2.2) is that the master equation formulation describes how the probability of being in a given microstate evolves over time. Of course, if the propensity functions and energy levels are modeled properly, the steady state, average probabilities of being in a given microstate should be the same for both formulations.

Mass action kinetics

Although very general in form, the chemical master equation suffers from being a very high dimensional representation of the dynamics of the system. We shall see in Chapter 4 how to implement simulations that obey the master equation, but in many instances we will not need this level of detail in our modeling. In particular, there are many situations in which the number of molecules of a given species is such that we can reason about the behavior of a chemically reacting system by keeping track of the *concentration* of each species as a real number. This is of course an approximation, but if the number of molecules is sufficiently large, then the approximation will generally be valid and our models can be dramatically simplified.

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

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

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

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

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 μ represents the parameters that govern the dynamic behavior.

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

$$A + B \rightleftharpoons AB.$$

Each time the forward reaction occurs, we decrease the number of molecules of

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

Using the discussion from the chemical master equation, we know that the likelihood that the reaction occurs in a given interval dt is given by $a(\xi^{\mathrm{f}};x,t)dt = c_{\xi^{\mathrm{f}}} n_{\mathrm{A}} n_{\mathrm{B}} dt$ where $c_{\xi^{\mathrm{f}}}$ is a constant. Another way of viewing this equation is that the rate at which reactions occur is given by $a(\xi;x,t)$. Looking first at the species AB, we can thus write

$$\begin{aligned}\frac{d}{dt}[\mathrm{AB}] &= c_{\xi^{\mathrm{f}}} n_A n_B - c_{\xi^{\mathrm{r}}} n_{\mathrm{AB}} \\ &= (c_{\xi^{\mathrm{f}}}\Omega^2)[\mathrm{A}][\mathrm{B}] - (c_{\xi^{\mathrm{r}}}\Omega)[\mathrm{AB}] =: k_{\xi^{\mathrm{f}}}[\mathrm{A}][\mathrm{B}] - k_{\xi^{\mathrm{r}}}[\mathrm{AB}],\end{aligned}$$

where we have used the fact that $[\mathrm{A}] = n_{\mathrm{A}}/\Omega$ and similarly for B and AB. The constants $k_{\xi^{\mathrm{f}}}$ and $k_{\xi^{\mathrm{r}}}$ are the *rate constants* for the reaction and can be computed from the coefficients of the propensity functions:

$$\begin{aligned} k_{\xi^{\mathrm{f}}} &= c_{\xi^{\mathrm{f}}}\Omega^2 && \text{bimolecular reaction} \\ k_{\xi^{\mathrm{r}}} &= c_{\xi^{\mathrm{r}}}\Omega && \text{unimolecular reaction} \end{aligned} \tag{2.4}$$

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

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

where $C = [\mathrm{AB}]$. These equations are known as the *mass action kinetics* or the *reaction rate equations* for the system.

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 and write the individual reactions as

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

where $k_{\xi^{\mathrm{f}}}$ and $k_{\xi^{\mathrm{r}}}$ are the reaction rates. For bidirectional reactions we can also write

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

It is easy to generalize this equation to more general reactions. For example, if we have a reversible reaction of the form

$$\mathrm{A} + 2\,\mathrm{B} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} 2\,\mathrm{C} + \mathrm{D},$$

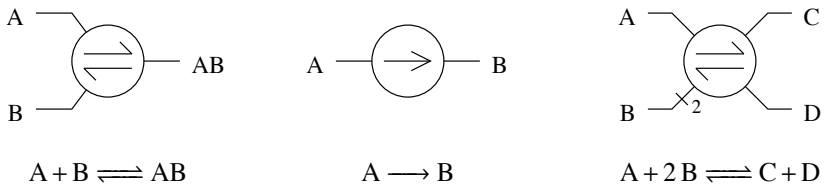


Figure 2.15: Diagrams for chemical reactions.

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

$$\begin{aligned}
\frac{d}{dt}A &= k_2 C^2 \cdot D - k_1 A \cdot B^2, \\
\frac{d}{dt}B &= 2k_2 C^2 \cdot D - 2k_1 A \cdot B^2, \\
\frac{d}{dt}C &= 2k_1 A \cdot B^2 - 2k_2 C^2 \cdot D, \\
\frac{d}{dt}D &= k_1 A \cdot B^2 - k_2 C^2 \cdot D.
\end{aligned} \tag{2.5}$$

Rearranging this equation, we can write the dynamics as

$$\frac{d}{dt}\begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ -2 & 2 \\ 2 & -2 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} k_1 A \cdot B^2 \\ k_2 C^2 \cdot D \end{pmatrix}. \tag{2.6}$$

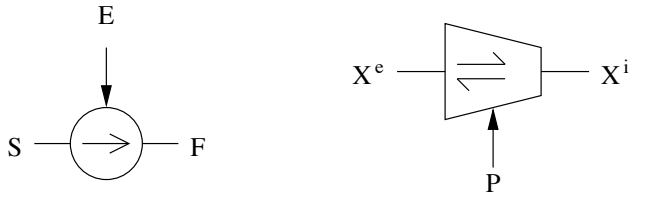
We see that in this composition, the first term on the right hand side is a matrix of integers reflecting the stoichiometry of the reactions and the second term is a vector of rates of the individual reactions.

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

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

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.

We will often find it convenient to represent collections of chemical reactions using simple diagrams, so that we can see the basic interconnection between various chemical species and properties. A standard chemical reaction diagram is shown in Figure 2.15.



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

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

Figure 2.16: Diagrams for enzymatic reactions.

Reduced order mechanisms

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

Simple binding reaction. Consider again the reaction

$$A + B \underset{k_r}{\overset{k_f}{\rightleftharpoons}} C, \tag{2.7}$$

in which we now assume that the total amount of A is conserved and we denote its total concentration by A_{tot}, so that $A + C = A_{tot}$. The corresponding rate equation for C is given by

$$\frac{dC}{dt} = k_f B \cdot (A_{tot} - C) - k_r C.$$

We are interested in determining the steady state value of the complex C concentration C and of the concentration of the free species A, i.e., A as a function of the concentration B. By setting $\dot{C} = 0$ and denoting $K_D := k_r/k_f$, we obtain the expressions:

$$C = \frac{BA_{tot}}{B + K_D}, \text{ and } A = \frac{A_{tot}K_D}{B + K_D}.$$

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

Cooperative binding reaction. Assume now that B binds to A only after a dimerization, that is, only after binding another molecule of B. Then, we have that reactions (2.7) become

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

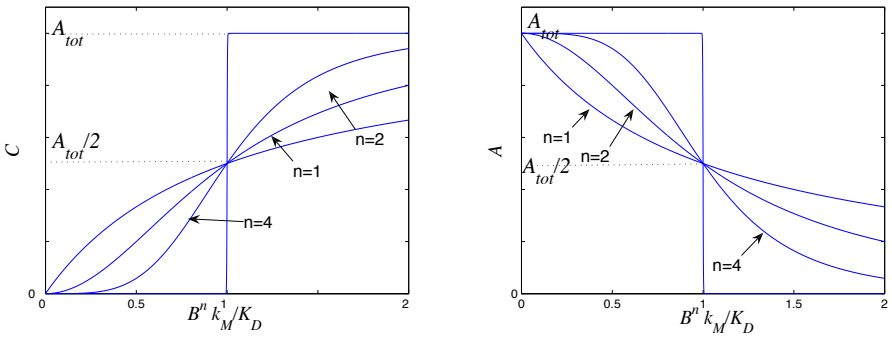


Figure 2.17: Steady state concentrations of the complex C and of A as functions of the concentration of B.

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

$$\frac{dB_d}{dt} = k_1 B^2 - k_2 B_d, \quad \frac{dC}{dt} = k_f B_d \cdot (A_{tot} - C) - k_r C.$$

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

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

so that

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

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

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

then we have that

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

In this case, one says that the binding of B to A is cooperative with cooperativity n. Figure 2.17 shows the above functions, which are often referred to as Hill functions.

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

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

for which, we can write the ODE system as

$$\frac{dC}{dt} = k_f B_a \cdot (A_{tot} - C - \bar{C}) - k_r C, \; \frac{d\bar{C}}{dt} = \bar{k}_f B_r \cdot (A_{tot} - C - \bar{C}) - k_r \bar{C}.$$

By setting the derivatives to zero, we obtain that

$$C(k_f B_a + k_r) = k_f B_a (A_{tot} - \bar{C}), \; \bar{C}(\bar{k}_f B_r + \bar{k}_r) = \bar{k}_f B_r (A_{tot} - C),$$

which, letting $\bar{K}_D := \bar{k}_r / \bar{k}_f$, leads to

$$\bar{C} = \frac{B_r (A_{tot} - C)}{B_r + \bar{K}_D}, \text{ and, } C\left(Ba + K_D - \frac{B_a B_r}{B_r + \bar{K}_D}\right) = B_a \left(\frac{\bar{K}_D}{B_r + \bar{K}_D}\right) A_{tot},$$

from which we finally obtain that

$$C = \frac{B_a A_{tot} \bar{K}_D}{\bar{K}_D B_a + K_D B_r + K_D \bar{K}_D}, \text{ and } \bar{C} = \frac{B_r A_{tot} K_D}{K_D B_r + \bar{K}_D B_a + K_D \bar{K}_D}.$$

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

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

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

add. In principle, one could consider all possible combinations of monomer, dimer, tetramer, etc. and activator, repressor, AND, different occupation states for the promoter, i.e., to consider exclusive binding or competitive binding. This should be done in a

Enzymatic reaction. A general enzymatic reaction can be written as

$$\mathrm{E} + \mathrm{S} \underset{k_r}{\overset{k_f}{\rightleftharpoons}} \mathrm{C} \xrightarrow{k_{cat}} \mathrm{E} + \mathrm{P},$$

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

system is given by

$$\begin{aligned}
\frac{dE}{dt} &= -k_f E \cdot S + k_r C + k_{cat} C \\
\frac{dS}{dt} &= -k_f E \cdot S + k_r C \\
\frac{dC}{dt} &= k_f E \cdot S - (k_r + k_{cat}) C \\
\frac{dP}{dt} &= k_{cat} C.
\end{aligned}$$

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

$$\begin{aligned}
\frac{dE}{dt} &= -k_f (E_{tot} - C) \cdot S + k_r C + k_{cat} C \\
\frac{dS}{dt} &= -k_f (E_{tot} - C) \cdot S + k_r C \\
\frac{dC}{dt} &= k_f (E_{tot} - C) \cdot S - (k_r + k_{cat}) C \\
\frac{dP}{dt} &= k_{cat} C.
\end{aligned}$$

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-equilibrium* approximation between the enzyme and the complex. This assumption can be translated in the condition

$$k_f, k_r \gg k_{cat}$$

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

$$C = \frac{E_{tot} S}{S + K_m}, \text{ with } K_m = \frac{k_r + k_{cat}}{k_f},$$

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

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

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

Chemical reaction networks

2.3 Modeling Transcription and Translation

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

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

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

$$
\begin{aligned}
\mathrm{RNAP}^{\mathrm{c}} &\rightleftharpoons \mathrm{RNAP}^{\mathrm{d}} && \text{binding to DNA;}\\
\mathrm{RNAP}^{\mathrm{d}} &\rightleftharpoons \mathrm{RNAP}^{\mathrm{p}} && \text{diffusion along DNA;}\\
\mathrm{RNAP}^{\mathrm{p}} + \mathrm{DNA}^{\mathrm{p}} &\rightleftharpoons \mathrm{RNAP{:}DNA}^{\mathrm{p}} && \text{binding to promoter;}\\
\mathrm{RNAP{:}DNA}^{\mathrm{p}} &\rightleftharpoons \mathrm{RNAP{:}DNA}^{\mathrm{o}} && \text{isomerization;}\\
\mathrm{RNAP{:}DNA}^{\mathrm{o}} &\longrightarrow \mathrm{RNAP{:}DNA}^{\mathrm{g},1} + \mathrm{DNA}^{\mathrm{p}} && \text{start of transcription;}\\
\mathrm{RNAP{:}DNA}^{\mathrm{g},1} &\longrightarrow \mathrm{RNAP{:}DNA}^{\mathrm{g},2} + \mathrm{mRNA}^{1} && \text{creation of mRNA;}\\
\mathrm{RNAP{:}DNA}^{\mathrm{g},i+1} + \mathrm{mRNA}^{\mathrm{i}} &\longrightarrow \mathrm{RNAP{:}DNA}^{\mathrm{g},i+2} + \mathrm{mRNA}^{\mathrm{i}+1} && \text{elongation, } i = 1,\ldots,N;\\
\mathrm{RNAP{:}DNA}^{\mathrm{g,N}} + \mathrm{mRNA}^{\mathrm{N}-1} &\longrightarrow \mathrm{RNAP{:}DNA}^{\mathrm{t}} + \mathrm{mRNA}^{\mathrm{N}} && \text{binding to terminator;}\\
\mathrm{RNAP{:}DNA}^{\mathrm{t}} &\longrightarrow \mathrm{RNAP}^{\mathrm{c}} && \text{termination;}\\
\mathrm{mRNA}^{\mathrm{N}} &\longrightarrow \emptyset && \text{degradation.}
\end{aligned}
\tag{2.8}
$$

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 $\text{Ribo:mRNA}^{\text{RBS}}$ indicate the ribosome bound to the ribosome binding site, Ribo:mRNA^{i} the ribosome bound to the ith codon, $\text{Ribo:mRNA}^{\text{s}}$ for the stop codon, and PPC^{i} for a polypeptide chain consisting of i amino acids. The reactions describing translation can then be written as

$$
\begin{aligned}
\text{Ribo}^{\text{c}} &\rightleftharpoons \text{Ribo}^{\text{rna}} && \text{binding to RNA;}\\
\text{Ribo}^{\text{rna}} + \text{mRNA}^{\text{RBS}} &\rightleftharpoons \text{Ribo:mRNA}^{\text{RBS}} && \text{binding to RBS;}\\
\text{Ribo:mRNA}^{\text{RBS}} &\longrightarrow \text{Ribo:mRNA}^{1} + \text{mRNA}^{\text{RBS}} && \text{start of translation;}\\
\text{Ribo:mRNA}^{1} &\longrightarrow \text{Ribo:mRNA}^{2} + \text{ppc}^{1} && \text{creation of polypeptide chain;}\\
\text{Ribo:mRNA}^{\text{i+1}} + \text{ppc}^{\text{i}} &\longrightarrow \text{Ribo:mRNA}^{\text{i+2}} + \text{ppc}^{\text{i+1}} && \text{elongation, } i = 1,\ldots,M;\\
\text{Ribo:mRNA}^{\text{M}} + \text{ppc}^{\text{M-1}} &\longrightarrow \text{Ribo:mRNA}^{\text{s}} + \text{ppc}^{\text{M}} && \text{stop codon;}\\
\text{Ribo:mRNA}^{\text{stop}} &\longrightarrow \text{Ribo}^{\text{c}} && \text{release of mRNA;}\\
\text{ppc}^{\text{M}} &\longrightarrow \text{protein} && \text{folding;}\\
\text{protein} &\longrightarrow \emptyset && \text{degradation.}
\end{aligned}
$$

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 equation are, they are still missing many important effects. For example, we have not accounted for the possibility of multiple RNA polymerases or ribosomes interacting with each other, so it is possible in these reactions to have two or more $\text{RNAP:DNA}^{g,i}$ complexes, which would correspond to multiple RNA polymerases bound to the same spot on a single piece of DNA. 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. And we have left out the many chemical species that must be present in order for many 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 func-

tions 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 using equation (2.4) 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 (such as DNA) exist as a single molecule 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, a even simpler model of the transcription, translation and folding processes can be utilized. If we assume that RNA polymerase binds to DNA at some average rate (which includes both the binding and isomerization reactions) and that transcription takes some fixed time (depending on the length of the gene), then the process of transcription can be described using the delay differential equation

$$\frac{dm_p}{dt} = \alpha_{p,0} - \gamma_p m_p, \qquad m_p^*(t) = e^{\delta_c \tau_{m,p}} m_p(t - \tau_{m,p}), \tag{2.9}$$

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

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

$$\frac{dP}{dt} = \beta_{p,0} m_p^* - \delta_p P, \qquad \frac{dP^*}{dt} = \beta_p^* \left(e^{-\delta_c \tau_{f,p}} P(t - \tau_{f,p}) - P^* \right) - \delta_p^* P^* \tag{2.10}$$

Here P represents the concentration of the polypeptide chain for the protein, P^* 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_p and δ_p^*, the rate of degradation and dilution of P and P^* respectively; β_p^*, the rate at which unfolded protein is folded; and $\tau_{f,p}$, the time delay associated with folding and other processes required to make the protein functional. Note that the rate of production

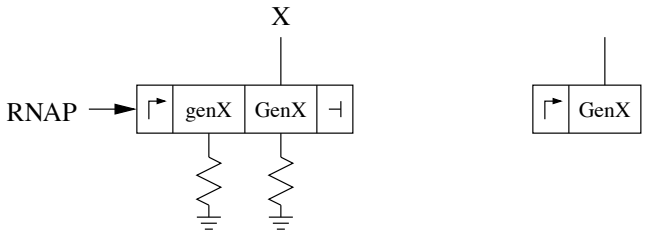


Figure 2.18: Simplified diagrams for protein production. The diagram on the left shows a section of DNA with RNA polymerase as an input, protein concentration as an output and degradation of mRNA and protein. The figure on the right is a simplified view in which only the protein output is indicated.

of the polypeptide chain P depends on the active mRNA concentration and the rate of production of the functional protein P depends on how much unfolded protein is available. We model this amount by looking at the polypeptide concentration at a time $\tau_{f,p}$ seconds ago, $P(t-\tau_{f,p})$, minus the amount of already functional protein $P(t)$. The exponential term again accounts for dilution due to cell growth. The degradation and dilution term, parameterized by δ_p and δ_p^*, captures both the rates at which the polypeptide chain and the protein are degraded and the rates at which these species are diluted due to cell growth.

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 the our model of the dynamics of protein production and write

$$\frac{dm_p}{dt} = \alpha_{p,0} - \gamma_p m_p, \qquad \frac{dP}{dt} = \beta_{p,0} m_p - \delta_p P. \tag{2.11}$$

Note that we have dropped the superscript $*$ since we are assuming that all mRNA is active and proteins are functional.

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 (2.11) with its equilibrium value. Thus we obtain

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

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

We will often find it convenient to represent protein production using a simple diagram that hides the details of the particular model that we decide to use. Figure 2.18 shows the symbol that we will use through the text. The diagram is intended to resemble a section of double stranded DNA, with a promoter and ter-

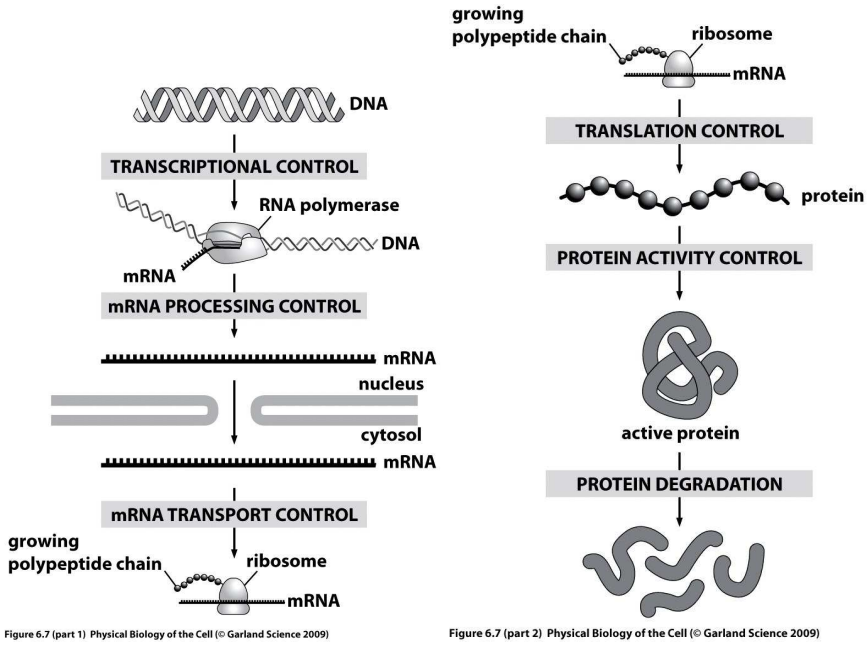


Figure 2.19: Regulation of proteins. Figure from Phillips, Kondev and Theriot [28]; used with permission of Garland Science.

minator at the ends, and then a list of the gene and protein in the middle. The boxes labeled by the gene and protein schematically represent the mRNA and protein concentration, with the line at the left of the DNA represent the input of RNA polymerase and the line on the top representing the the (folded) protein. The symbols at the bottom represent the degradation and dilution of mRNA and protein.

2.4 Transcriptional Regulation

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

The regulation of gene expression and protein activity is accomplished through a variety of molecular mechanisms, as illustrated in Figure 2.19. We see that at each stage of the processing from a gene to a protein, there are potential mechanisms for regulating the production processes. The remainder of this section will focus on transcriptional control, the next section on control between transcription and translation, and the third section on post-translational control mechanisms. We

begin with a description of regulation mechanisms in prokaryotes (bacterial) and then describe the additional mechanisms that are specific to eukaryotes.

Prokaryotic mechanisms

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

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

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

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 and the equilibrium concentrations of $\mathrm{DNA^p}$ versus DNA:Rep model the "leakiness" of the repressor.

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

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

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

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

modeled by adding the reaction

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

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

$$\begin{aligned}
\text{Act} + \text{Ind} &\rightleftharpoons \text{Act:Ind} && \text{inducer binding;}\\
\text{DNA}^{\text{p}} + \text{Act:Ind} &\rightleftharpoons \text{DNA:Act:Ind} && \text{activator binding;}\\
\text{RNAP}^{\text{d}} &\rightleftharpoons \text{RNAP}^{\text{p}} && \text{diffusion along DNA;}\\
\text{RNAP}^{\text{p}} + \text{DNA:Act:Ind} &\rightleftharpoons \text{RNAP:DNA}^{\text{o}} + \text{DNA:Act:Ind} && \text{binding to promoter w/ activator.}
\end{aligned}$$

A simplified version of the dynamics can be obtained by assuming that transcription factors bind to the DNA rapidly, so that they are in steady state configurations. In this case, we can make use of the steady state statistical mechanics techniques described in Section 2.2 and relate the expression of the gene to the probability that the activator or repressor is bound to the DNA (P_{bound}). This could be done at the level of the reaction rate equation by replacing the differential equations for activator or repressor binding with their steady state values. Here instead we demonstrate how to account for this rapid binding in the simplified differential equation models presented at the end of Section 2.3.

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

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

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

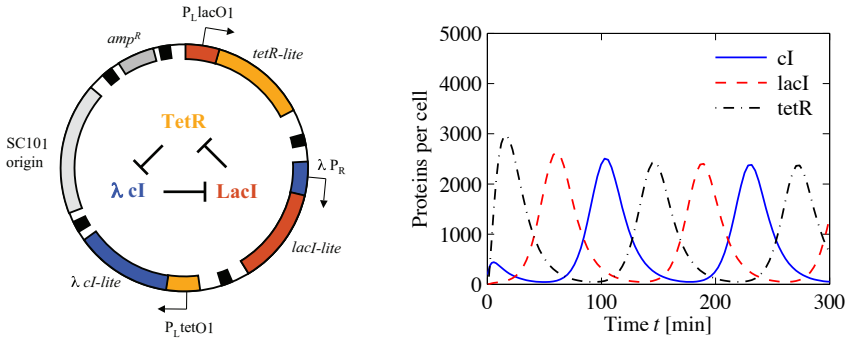
$$\frac{dm_a}{dt} = (1 - P_{\text{bound}}(p_{\text{b}}))\alpha_{a0} - \gamma_a m_a. \tag{2.12}$$

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

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

$$\frac{dm_a}{dt} = P_{\text{bound}}(p_{\text{b}})\alpha_{a0} - \gamma_a m_a, \tag{2.13}$$

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



(a) Repressilator plasmid (b) Repressilator simulation

Figure 2.20: 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.)

As we shall see in Chapter 4 (see also Exercise 2.1, the functional form of P_{bound} can be nicely approximated by a monotonic rational function, called a *Hill function* [10, 24]. For a repressor, the Hill function is given by

$$f_a^r(p_b) = \frac{\alpha_{ab}}{k_{ab} + p_{\text{b}}^{n_{ab}}} + \alpha_a,$$

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

Example 2.1 (Repressilator)**.** As an example of how these models can be used, we consider the model of a "repressilator," originally due to Elowitz and Leibler [12]. The repressilator is a synthetic circuit in which three proteins each repress another in a cycle. This is shown schematically in Figure 2.20a, where the three proteins are TetR, λ cI and LacI.

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

We can model this system using three copies of equation (2.12), with A and B replaced by the appropriate combination of TetR, cI and LacI. The state of the system is then given by $x = (m_{\text{TetR}}, p_{\text{TetR}}, m_{\text{cI}}, p_{\text{cI}}, m_{\text{LacI}}, p_{\text{LacI}})$. Figure 2.20b shows the traces of the three protein concentrations for parameters $n = 2$, $\alpha = 0.5$, $k =$

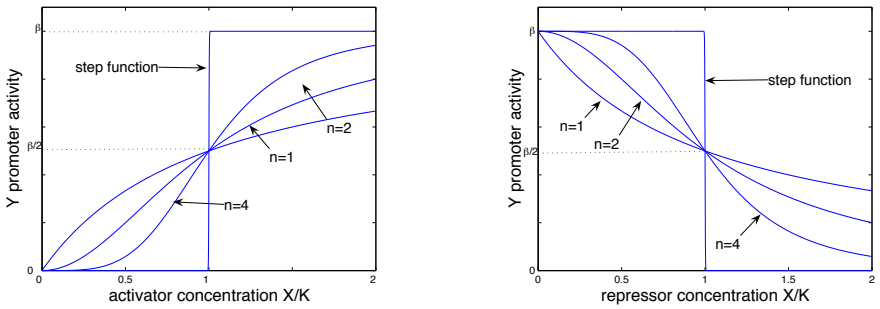


Figure 2.21: Hill function for an activator (left) and for a repressor (right).

6.25×10^{-4}, $\alpha_0 = 5\times10^{-4}$, $\gamma = 5.8\times10^{-3}$, $\beta = 0.12$ and $\delta = 1.2\times10^{-3}$ with initial conditions $x(0) = (1,0,0,200,0,0)$ (following [12]). ∇

For an activator the Hill function is given by

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

where the variables are the same as described previously. Note that in the case of the activator, if p_{b} is zero, then the production rate is α_{a0} (versus $\alpha_{ab} + \alpha_{a0}$ for the repressor). As p_{b} gets large, the first term in the Hill function approaches α_{ab} and the transcription rate becomes $\alpha_{ab} + \alpha_{a0}$ (versus α_{a0} for the repressor). Thus we see that the activator and repressor act in opposite fashion from each other. Figure 2.21 shows the standard Hill functions for activation and repression.

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

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

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

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

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

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

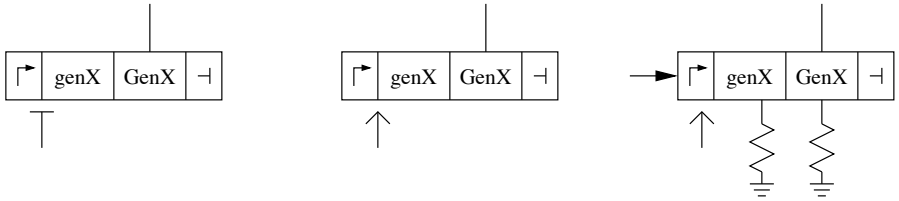


Figure 2.22: Circuit diagrams for transcriptional regulation of a gene. The first two figures represent repression and activation. If desired, additional mechanisms can also be indicated, as shown in the diagram on the right.

present and assume this is constant, we can write

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

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

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

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

We will often find it convenient to represent the process of regulation in a graphical fashion that hides the specific details of the model that we choose to use. Figure 2.22 shows the notation that we will use in this text to represent the process of transcription, translation and regulation.

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

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

$$\frac{dm_y}{dt} = f(p_1,...,p_n) - \gamma_{\text{y}} m_y, \qquad \frac{dp_y}{dt} = \beta_{\text{y}} m_y - \delta_{\text{y}} p_y, \tag{2.14}$$

where m_{y} denotes the concentration of mRNA translated by gene y, the constants γ_{y} and δ_{y} incorporate the dilution and degradation processes, and β_{y} is a constant that establishes the rate at which the mRNA is translated.

Several other methods of transcriptional regulation can exist in cells.

Antitermination. Antitermination can also be used as a transcriptional regulatory mechanism. To model its effects, assume that we have a coding region labeled h

that occurs after an antitermination site. We modify the termination reactions from equation (2.8):

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

Regulation in eukaryotes

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

2.5 Post-Transcriptional and Post-Translational 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.

RNA-based regulation

Allosteric modifications to proteins

Covalent modifications to proteins

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

At high level, any covalent modification cycle involves a target protein, say X, an enzyme for modifying it, say Z, and one for reversing the modification, say Y (see Figure 2.23). We call X^* the activated protein. There are often allosteric effectors or further covalent modification systems that regulate the activity of the modifying enzymes, but we do not consider here this added level of complexity. There are several types of covalent modification, depending on the type of activation of the protein. *Phosphorylation* is a covalent modification that takes place mainly in eukaryotes and involves activation of the inactive protein X by addition of a phosphate group. In this case, the enzyme Z is called a kinase while the enzyme Y is called phosphatase. Another type of covalent modification, which is

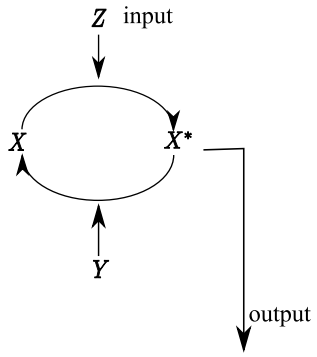


Figure 2.23: Diagram representing a covalent modification cycle.

very common in both procaryotes and eukaryotes, is *methylation*. Here, the inactive protein is activated by the addition of a methyl group.

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

$$\mathrm{Z} + \mathrm{X} \underset{k_r}{\overset{k_f}{\rightleftharpoons}} \mathrm{C} \xrightarrow{k_{cat}} \mathrm{X}^* + \mathrm{Z}$$

$$\mathrm{Y} + \mathrm{X}^* \underset{k'_r}{\overset{k'_f}{\rightleftharpoons}} \mathrm{C}' \xrightarrow{k'_{cat}} \mathrm{X} + \mathrm{Y}.$$

The corresponding ODE model is given by

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

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

$Z_{tot} - C$ and $Y = Y_{tot} - C'$.

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

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

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

and

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

By solving these equations, we obtain that

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

and that

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

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

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

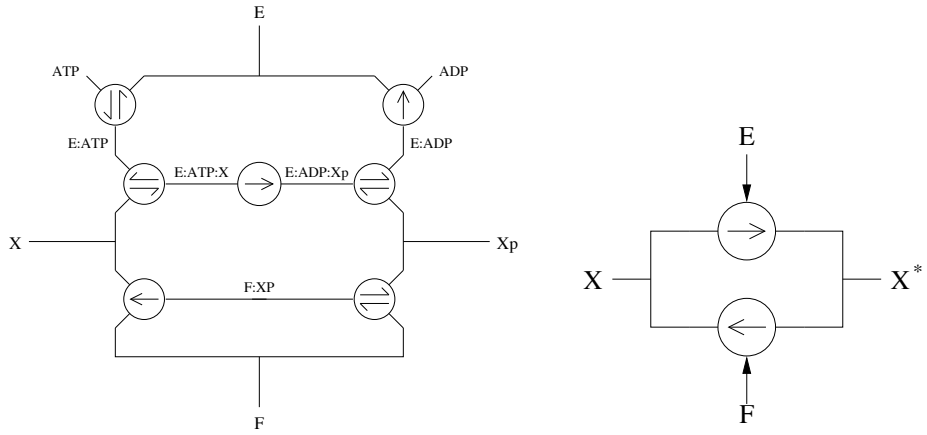
which, considering that $k'_f K'_m - k'_r = k'_{cat}$, leads finally to

$$\frac{dX^*}{dt} = k_{cat}\frac{Z_{tot}X}{X + K_m} - k'_{cat}\frac{Y_{tot}X^*}{X^* + K'_m}. \tag{2.15}$$

We will come back to the modeling of this system after we have introduced singular perturbation theory, through which we will be able to perform a formal analysis of this system and mathematically characterize the assumptions needed for approximating the original system by the first order ODE model (2.15).

Exercise. As an exercise, the reader can consider the case in which the kinase Z is not constant, but it is produced and decays according to the reaction $0 \underset{\delta}{\overset{k(t)}{\rightleftharpoons}} \mathrm{Z}$. How should the system in equation (2.15) be modified?

Exercise. There is another model for the phosphorylation reactions, referred to as one step reaction model, given by $\mathrm{Z} + \mathrm{X} \rightleftharpoons X^* + \mathrm{Z}$ and $\mathrm{Y} + \mathrm{X}^* \rightleftharpoons X + \mathrm{Y}$,



(a) Full reaction (b) Reduced reaction

Figure 2.24: Circuit diagram for phosphorylation and dephoshorylation of a protein X via a kinase E and phosphotase F. The diagram on the left shows the full set of reactions. A simplified diagram is shown on the right.

in which the complex formations are neglected. Write down the ODE model and comparing the differential equation of X^* to that of equation (2.15), list the assumptions under which the one step reaction model is a good approximation of the two step reaction model.

The phosphorylation/dephosophorylation process is illustrated in circuit diagram form in Figure 2.24.

Phosphotransfer systems

2.6 Cellular subsystems

Intercellular Signalling

Adaptation

Logical operations

Exercises

2.1 Consider a repressor that binds to an operator site as a dimer:

$$\begin{aligned} &\text{R1:} \quad R + R \rightleftharpoons R_2 \\ &\text{R2:} \quad R_2 + DNA^p \rightleftharpoons R_2\text{:DNA} \\ &\text{R3:} \quad RNAP + DNA^p \rightleftharpoons RNAP\text{:}DNA^p \end{aligned}$$

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^p to the total amount of DNA, D_T, can be written as a

Hill function

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

and give expressions for α and K.