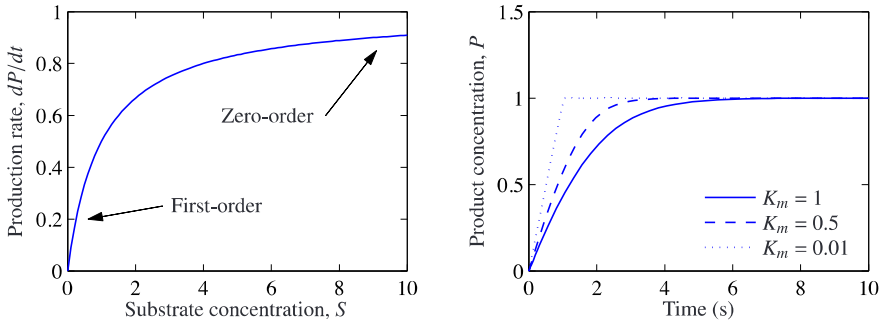

(a) Production rate (b) Time plots

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

are called *Michaelis-Menten kinetics*.

The constant V_{max} is called the maximal velocity (or maximal flux) of modification and it represents the maximal rate that can be obtained when the enzyme is completely saturated by the substrate. The value of K_m corresponds to the value of S that leads to a half-maximal value of the production rate of P. When the enzyme complex can be neglected with respect to the total substrate amount S_{tot}, we have that $S_{\text{tot}} = S + P + C \approx S + P$, so that the above equation can be also rewritten 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, $S \gg K_m$, we have that the rate of product formation becomes approximately $dP/dt \approx V_{max}$, which is the maximal speed of reaction. Since this rate is constant and does not depend on the reactant concentrations, it is usually referred to as *zero-order kinetics*. In this case, the system is said to operate in the zero-order regime. If instead $S \ll K_m$, the rate of product formation becomes $dP/dt \approx V_{max}/K_m S$, which is linear with the substrate concentration S. This production rate is referred to as *first-order kinetics* and the system is said to operate in the first-order regime (see Figure 2.5).

2.2 Transcription and translation

In this section we consider the processes of transcription and translation, using the modeling techniques described in the previous section to capture the fundamental dynamic behavior. Models of transcription and translation can be done at a variety of levels of detail and which model to use depends on the questions that one

wants to consider. We present several levels of modeling here, starting with a fairly detailed set of reactions describing transcription and translation and ending with highly simplified ordinary differential equation models that can be used when we are only interested in average production rate of mRNA and proteins at relatively long time scales.

The central dogma: Production of proteins

The genetic material inside a cell, encoded in its DNA, governs the response of a cell to various conditions. DNA is organized into collections of genes, with each gene encoding a corresponding protein that performs a set of functions in the cell. The activation and repression of genes are determined through a series of complex interactions that give rise to a remarkable set of circuits that perform the functions required for life, ranging from basic metabolism to locomotion to procreation. Genetic circuits that occur in nature are robust to external disturbances and can function in a variety of conditions. To understand how these processes occur (and some of the dynamics that govern their behavior), it will be useful to present a relatively detailed description of the underlying biochemistry involved in the production of proteins.

DNA is a double stranded molecule with the "direction" of each strand specified by looking at the geometry of the sugars that make up its backbone. The complementary strands of DNA are composed of a sequence of nucleotides that consist of a sugar molecule (deoxyribose) bound to one of four bases: adenine (A), cytocine (C), guanine (G) and thymine (T). The coding region (by convention the top row of a DNA sequence when it is written in text form) is specified from the 5′ end of the DNA to the 3′ end of the DNA. (The 5′ and 3′ refer to carbon locations on the deoxyribose backbone that are involved in linking together the nucleotides that make up DNA.) The DNA that encodes proteins consists of a promoter region, regulator regions (described in more detail below), a coding region and a termination region (see Figure 2.6). We informally refer to this entire sequence of DNA as a gene.

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

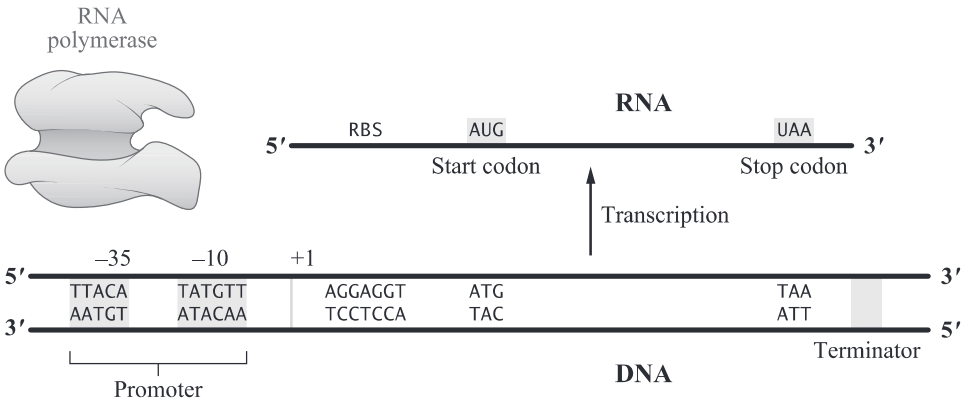

Figure 2.6: Geometric structure of DNA. The layout of the DNA is shown at the top. RNA polymerase binds to the promoter region of the DNA and transcribes the DNA starting at the +1 site and continuing to the termination site. The transcribed mRNA strand has the ribosome binding site (RBS) where the ribosomes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

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

Once the mRNA is produced, it must be translated into a protein. This process is slightly different in prokaryotes and eukaryotes. In prokaryotes, there is a region of the mRNA in which the ribosome (a molecular complex consisting of 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-Dalgarno sequence, AGGAGG, is the consensus sequence for the RBS. (A consensus sequence is a pattern of nucleotides that implements a given function across multiple organisms; it is not exactly conserved, so some variations in the sequence will be present from one organism to another.)

In eukaryotes, the RNA must undergo several additional steps before it is trans-

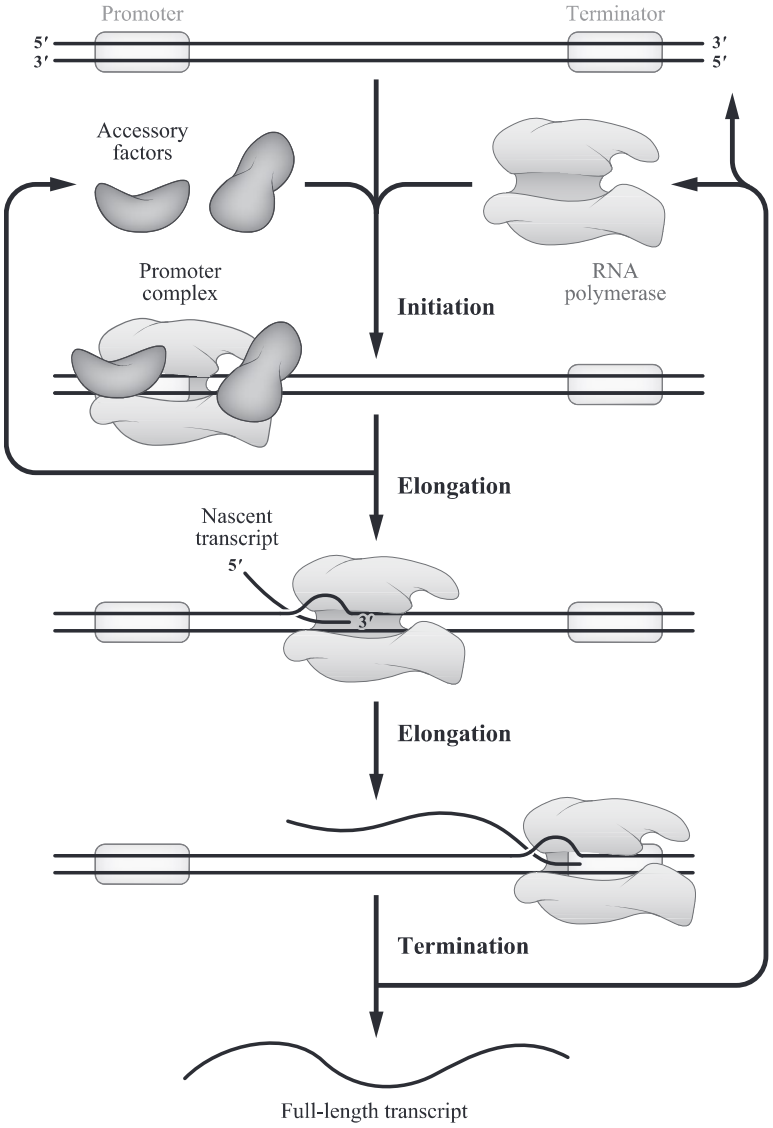

Figure 2.7: Production of messenger RNA from DNA. RNA polymerase, along with other accessory factors, binds to the promoter region of the DNA and then "opens" the DNA to begin transcription (initiation). As RNA polymerase moves down the DNA in the transcription elongation complex (TEC), it produces an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

lated. The RNA sequence that has been created by RNA polymerase consists of *introns* that must be spliced out of the RNA (by a molecular complex called the spliceosome), leaving only the *exons*, which contain the coding region for the pro-

tein. The term *pre-mRNA* is often used to distinguish between the raw transcript and the spliced mRNA sequence, which is called *mature mRNA*. In addition to splicing, the mRNA is also modified to contain a *poly(A)* (polyadenine) *tail*, consisting of a long sequence of adenine (A) nucleotides on the 3′ end of the mRNA. This processed sequence is then transported out of the nucleus into the cytoplasm, where the ribosomes can bind to it.

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

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

The sequence of amino acids produced by the ribosome is a polypeptide chain that folds on itself to form a protein. The process of folding is complicated and 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 underlie the behavior of the organism. The *maturation time* of a protein is the time required for the polypeptide chain to fold into a functional protein.

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

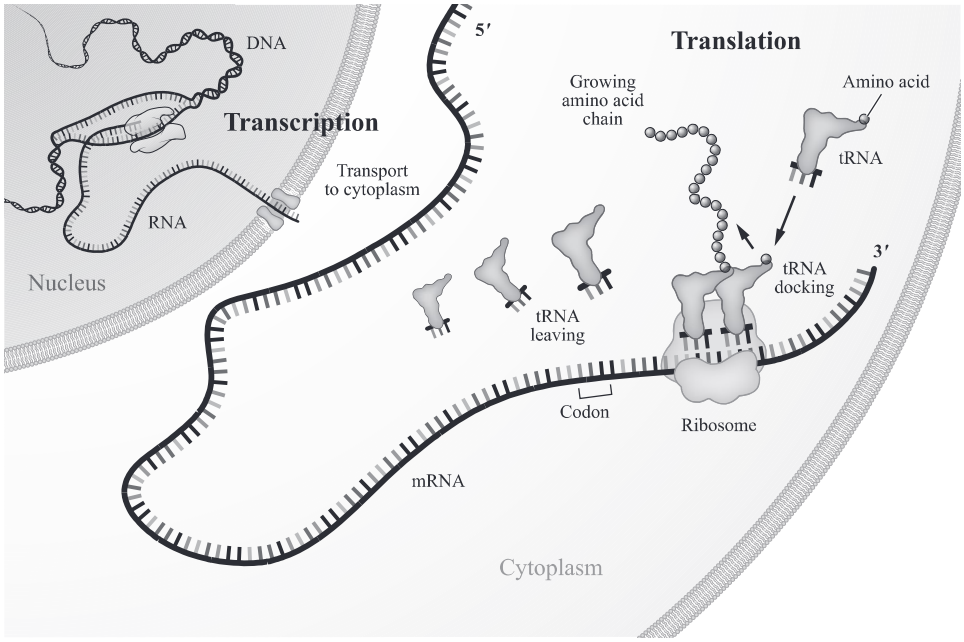

Figure 2.8: Translation is the process of translating the sequence of a messenger RNA (mRNA) molecule to a sequence of amino acids during protein synthesis. The genetic code describes the relationship between the sequence of base pairs in a gene and the corresponding amino acid sequence that it encodes.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

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

Process	Characteristic rate	Source
mRNA transcription rate	24-29 bp/s	[13]
Protein translation rate	12–21 aa/s	[13]
Maturation time (fluorescent proteins)	6–60 min	[13]
mRNA half-life	~ 100 s	[103]
E. coli cell division time	20–40 min	[13]
Yeast cell division time	70–140 min	[13]
Protein half-life	$\sim 5 \times 10^4$ s	[103]
Protein diffusion along DNA	up to 10^4 bp/s	[78]
RNA polymerase dissociation constant	~ 0.3–10,000 nM	[13]
Open complex formation kinetic rate	~ 0.02 s^{-1}	[13]
Transcription factor dissociation constant	~ 0.02–10,000 nM	[13]

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

Reaction models

The basic reactions that underlie 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: $\mathrm{RNAP^c}$ represents RNA polymerase in the cytoplasm, $\mathrm{RNAP^p}$ represents RNA polymerase in the promoter region, and $\mathrm{RNAP^d}$ is RNA polymerase non-specifically bound to DNA. We must similarly keep track of the state of the DNA, to ensure that multiple RNA polymerases do not bind to the same section of DNA. Thus we can write $\mathrm{DNA^p}$ for the promoter region, $\mathrm{DNA^i}$ for the ith section of the gene of interest and $\mathrm{DNA^t}$ for the termination sequence. We write $\mathrm{RNAP : DNA}$ to represent RNA polymerase bound to DNA (assumed closed) and $\mathrm{RNAP : DNA^o}$ to indicate the open complex. Finally, we must keep track of the mRNA that is produced by transcription: we write $\mathrm{mRNA^i}$ 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

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

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

A similar set of reactions can be written to model the process of translation. Here we must keep track of the binding of the ribosome to the ribosome binding site (RBS) of $\mathrm{mRNA}^{\mathrm{N}}$, translation of the mRNA sequence into a polypeptide chain, and folding of the polypeptide chain into a functional protein. Specifically, we must keep track of the various states of the ribosome bound to different codons on the mRNA strand. We thus let $\mathrm{Ribo} : \mathrm{mRNA}^{\mathrm{RBS}}$ denote the ribosome bound to the ribosome binding site of $\mathrm{mRNA}^{\mathrm{N}}$, $\mathrm{Ribo} : \mathrm{mRNA}^{\mathrm{AAi}}$ the ribosome bound to the ith codon (corresponding to an amino acid, indicated by the superscript AA), $\mathrm{Ribo} : \mathrm{mRNA}^{\mathrm{start}}$ and $\mathrm{Ribo} : \mathrm{mRNA}^{\mathrm{stop}}$ the ribosome bound to the start and stop codon, respectively. We also let $\mathrm{PPC}^{\mathrm{i}}$ denote the polypeptide chain consisting of i amino acids. Here, we assume that the protein of interest has M amino acids. The reactions describing translation can then be written as

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

As in the case of transcription, we see that these reactions allow multiple ribosomes to translate the same piece of mRNA by freeing up mRNA^{N}. After M amino acids have been chained together, the M-long polypeptide chain PPC^{M} is released, which then folds into a protein. As complex as these reactions are, they do not directly capture a number of physical phenomena such as ribosome queuing, wherein ribosomes cannot pass other ribosomes that are ahead of them on the mRNA chain. Additionally, we have not accounted for the existence and effects of the $5'$ and $3'$ untranslated regions (UTRs) of a gene and we have also left out various error correction mechanisms in which ribosomes can step back and release an incorrect amino acid that has been incorporated into the polypeptide chain. We have also left out the many chemical species that must be present in order for a variety of the reactions to happen (NTPs for mRNA production, amino acids for protein production, etc.). Incorporation of these effects requires additional reactions that track the many possible states of the molecular machinery that underlies transcription and translation. For more detailed models of translation, the reader is referred to [3].

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

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

in which the second reaction lumps together isomerization, start of transcription, elongation, mRNA creation, and termination. Similarly, for the translation process, the reduced reactions take the form:

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

in which the second reaction lumps the start of translation, elongation, folding, and termination. The third reaction models the fact that mRNA can also be degraded when bound to ribosomes when the ribosome binding site is left free. The process of mRNA degradation occurs through RNase enzymes binding to the ribosome binding site and cleaving the mRNA strand. It is known that the ribosome binding site cannot be both bound to the ribosome and to the RNase [68]. However, the species Ribo : mRNA is a lumped species encompassing configurations in which ribosomes are bound on the mRNA strand but not on the ribosome binding site. Hence, we also allow this species to be degraded by RNase.

Reaction rate equations

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

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

In some situations, an even simpler model of the transcription, translation and folding processes can be utilized. Let the "active" mRNA be the mRNA that is available for translation by the ribosome. We model its concentration through a simple time delay of length τ^m that accounts for the transcription of the ribosome binding site in prokaryotes or splicing and transport from the nucleus in eukaryotes. If we assume that RNA polymerase binds to DNA at some average rate (which includes both the binding and isomerization reactions) and that transcription takes some fixed time (depending on the length of the gene), then the process of transcription can be described using the delay differential equation

$$\frac{dm_{\mathrm{P}}}{dt} = \alpha - \mu m_{\mathrm{P}} - \bar{\delta} m_{\mathrm{P}}, \qquad m^*_{\mathrm{P}}(t) = e^{-\mu\tau^m} m_{\mathrm{P}}(t-\tau^m), \tag{2.14}$$

where m_{P} is the concentration of mRNA for protein P, m^*_{P} is the concentration of active mRNA, α is the rate of production of the mRNA for protein P, μ is the growth rate of the cell (which results in dilution of the concentration) and $\bar{\delta}$ is the rate of degradation of the mRNA. Since the dilution and degradation terms are of the same form, we will often combine these terms in the mRNA dynamics and use a single coefficient $\delta = \mu + \bar{\delta}$. The exponential factor in the second expression in equation (2.14) accounts for dilution due to the change in volume of the cell, where μ is the cell growth rate. The constants α and δ capture the average rates of production and decay, which in turn depend on the more detailed biochemical reactions that underlie transcription.

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

tional protein:

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

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

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

$$\begin{aligned} \frac{dm_{\mathrm{P}}^*(t)}{dt} &= e^{-\mu\tau^m} \frac{dm_{\mathrm{P}}}{dt}(t-\tau^m) \\ &= e^{-\mu\tau^m}(\alpha - \delta m_{\mathrm{P}}(t-\tau^m)) = \bar{\alpha} - \delta m_{\mathrm{P}}^*(t), \end{aligned} \tag{2.16}$$

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

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

where $\bar{\kappa} = e^{-\mu\tau^f}\kappa$. We shall typically use equations (2.16) and (2.17) as our (reduced) description of protein folding, dropping the superscript f and overbars when there is no risk of confusion. Also, in the presence of different proteins, we will attach subscripts to the parameters to denote the protein to which they refer.

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

$$\frac{dm_{\mathrm{P}}}{dt} = \alpha - \delta m_{\mathrm{P}}, \qquad \frac{dP}{dt} = \kappa m_{\mathrm{P}} - \gamma P. \tag{2.18}$$

Note that we here have dropped the superscripts $*$ and f since we are assuming that all mRNA is active and proteins are functional and dropped the overbar on α and κ since we are assuming the time delays are negligible. The value of α increases with the strength of the promoter while the value of κ increases with the strength of the ribosome binding site. These strengths, in turn, can be affected by changing the specific base-pair sequences that constitute the promoter RNA polymerase binding region and the ribosome binding site.

Finally, the simplest model for protein production is one in which we only keep track of the basal rate of production of the protein, without including the mRNA

dynamics. This essentially amounts to assuming the mRNA dynamics reach steady state quickly and replacing the first differential equation in (2.18) with its equilibrium value. This is often a good assumption as mRNA degration is usually about 100 times faster than protein degradation (see Table 2.1). Thus we obtain

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

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

2.3 Transcriptional regulation

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

The regulation of gene expression and protein activity is accomplished through a variety of molecular mechanisms, as discussed in Section 1.2 and illustrated in Figure 2.9. At each stage of the processing from a gene to a protein, there are potential mechanisms for regulating the production processes. The remainder of this section will focus on transcriptional control and the next section on selected mechanisms for controlling protein activity. We will focus on prokaryotic mechanisms.

Transcriptional regulation of protein production

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

Repression. A common mechanism for repression is that a protein binds to a region of DNA near the promoter and blocks RNA polymerase from binding. The region of DNA to which the repressor protein binds is called an *operator region* (see Figure 2.10a). If the operator region overlaps the promoter, then the presence of a protein at the promoter can "block" the DNA at that location and transcription