Biomolecular Feedback Systems

Domitilla Del Vecchio
U. Michigan/MIT

Richard M. Murray
Caltech

DRAFT v0.3, March 11, 2010

Contents

Preface

This text serves as a supplement to *Feedback Systems* by Åström and Murray [1] (refered to throughout the text as AM08) and is intended for researchers interested in the application of feedback and control to biomolecular systems. The text has been designed so that it can be used in parallel with *Feedback Systems* as part of a course on biomolecular feedback and control systems, or as a standalone reference for readers who have had a basic course in feedback and control theory. The full text for AM08, along with additional supplemental material and a copy of these notes, is available on a companion web site:

`http://www.cds.caltech.edu/~murray/amwiki/BFS`

The text is intended to be useful to three overlapping audiences: graduate students in biology and bioengineering interested in understanding the role of feedback in natural and engineered biomolecular systems; advanced undergraduates and graduate students in engineering disciplines who are interested the use of feedback in biological circuit design; and established researchers in the the biological sciences who want to explore the potential application of principles and tools from control theory to biomolecular systems. We have written the text assuming familiarity with the material in AM08, but have tried to provide insights and motivation so that the material can be learned in parallel. We also assume some familiarity with cell biology, at the level of a first course for non-majors. The individual chapters in the text indicate the pre-requisites in more detail, most of which are covered either in AM08 or in the supplemental information available from the companion web site.

Notation

This is an internal chapter that is intended for use by the authors in fixing the notation that is used throughout the text. In the first pass of the book we are anticipating several conflicts in notation and the notes here may be useful to early users of the text.

Protein dynamics

For a gene 'genX', we write *genX* for the gene, $\mathrm{m}_{\mathrm{genX}}$ for the mRNA and GenX for the protein when they appear in text or chemical formulas. We use superscripts to differentiate between isomers, so $\mathrm{m}^{*}_{\mathrm{genX}}$ might be used to refer to mature RNA or $\mathrm{GenX}^{\mathrm{f}}$ to refer to the folded versions of a protein, if required. Mathematical formulas use italic version of the variable name, but roman font for the gene or isomeric state. The concentration of mRNA is written in text or formulas as m_{genX} (m^{*}_{genX} for mature) and the concentration of protein as p_{genX} ($p^{\mathrm{f}}_{\mathrm{genX}}$ for folded). The same naming conventions are used for common gene/protein combinations: the mRNA concentration of *tetR* is $\mathrm{m}_{\mathrm{tetR}}$, the concentration of the associated protein is p_{tetR} and parameters are α_{tetR}, δ_{tetR}, etc.

For generic genes and proteins, use X to refer to a protein, m_{x} to refer to the mRNA associated with that protein and *x* to refer to the gene that encodes X. The concentration of X can be written either as X, p_{x} or [X], with that order of preference. The concentration of m_{x} can be written either as m_{x} (preferred) or $[\mathrm{m}_{\mathrm{x}}]$. Parameters that are specific to gene *p* are written with a subscripted p: α_{p}, δ_{p}, etc. Note that although the protein is capitalized, the subscripts are lower case (so indexed by the gene, not the protein) and also in roman font (since they are not a variable).

The dynamics of protein production are given by

$$\frac{dm_{\mathrm{p}}}{dt} = \alpha_{\mathrm{p},0} - \mu m_{\mathrm{p}} - \gamma_{\mathrm{p}} m_{\mathrm{p}}, \qquad \frac{dP}{dt} = \beta_{\mathrm{p}} m_{\mathrm{p}} - \mu P - \delta_{\mathrm{p}} P,$$

where $\alpha_{\mathrm{p},0}$ is the (constitutive) rate of production, γ_{p} parameterizes the rate of dilution and degradation of the mRNA m_{p}, β_{p} is the kinetic rate of protein production, μ is the growth rate that leads to dilution of concentrations and δ_{p} parameterizes the rate of degradation of the protein P. Since dilution and degradation enter in a similar fashion, we use $\overline{\gamma} = \gamma + \mu$ and $\overline{\delta} = \delta + \mu$ to represent the aggregate degradationa

and dilution rate. If we are looking at a single gene/protein, the various subscripts can be dropped.

When we ignore the mRNA concentration, we write the simplified protein dynamics as

$$\frac{dP}{dt} = \beta_{\mathrm{p},0} - \bar{\delta}_{\mathrm{p}} P.$$

Assuming that the mRNA dynamics are fast compared to protein production, then the constant $\beta_{\mathrm{p},0}$ is given by

$$\beta_{\mathrm{p},0} = \beta_{\mathrm{p}} \frac{\bar{\gamma}_{\mathrm{p}}}{\alpha_{\mathrm{p},0}}.$$

For regulated production of proteins using Hill functions, we modify the constitutive rate of production to be $f_{\mathrm{p}}(Q)$ instead of $\alpha_{\mathrm{p},0}$ or $\beta_{\mathrm{p},0}$ as appropriate. The Hill function is written in the form

$$F_{\mathrm{p,q}}(Q) = \frac{\alpha_{\mathrm{p,q}}}{K_{\mathrm{p,q}} + Q^{n_{\mathrm{p,q}}}}.$$

The notation for F mirrors that of transfer functions: $F_{\mathrm{p,q}}$ represents the input/output relationship between input Q and output P (rate). The comma can be dropped when the genes in question are single letters:

$$F_{\mathrm{pq}}(Q) = \frac{\alpha_{\mathrm{pq}}}{K_{\mathrm{pq}} + Q^{n_{\mathrm{pq}}}}.$$

The subscripts can be dropped completely if there is only one Hill function in use.

Chemical reactions

We write the symbol for a chemical species A using roman type. The number of molecules of a species A is written as n_a. The concentration of the species is occasionally written as [A], but we more often use the notation A, as in the case of proteins, or x_a. For a reaction $\mathrm{A} + \mathrm{B} \longleftrightarrow \mathrm{C}$, we use the notation

$$\mathrm{R}_1 : \mathrm{A} + \mathrm{B} \underset{k^{\mathrm{r}}_{r1}}{\overset{k^{f}_{r1}}{\rightleftharpoons}} \mathrm{C} \qquad \frac{dC}{dt} = k^{f}_{r1} AB - k^{\mathrm{r}}_{r1} C$$

This notation is primarily intended for situations where we have multiple reactions and need to distinguish between many different constants. For a small number of reactions, the reaction number can be dropped or replaced with a single digit (k^{f}_1, k^{r}_2, etc).

It will often be the case that two species A and B will form a covalent bond, in which case we write the resulting species as AB. We will distinguish covalent bonds from much weaker hydrogen bonding by writing the latter as A:B. Finally, in some situations we will have labeled section of DNA that are connected together,

which we write as A−B, where here A represents the first portion of the DNA strand and B represents the second portion. When describing (single) strands of DNA, we write A′ to represent the Watson-Crick complement of the strand A. Thus A−B:B′−A′ would represent a double stranded length of DNA with domains A and B.

The choice of representing covalent molecules using the convential chemical notation AB can lead to some confusion when writing the reaction dynamics using A and B to represent the concentrations of those species. Namely, the symbol AB could represent either the concentration of A times the concentration of B or the concentration of AB. To remove this ambiguity, when using this notation we will write [A][B] as $A \cdot B$.

When working with a system of chemical reactions, we write S_i, $i = 1, \ldots, n$ for the species and R_j, $j = 1, \ldots, m$ for the reactions. We write n_i to refer to the molecular count for species i and $x_i = [S_i]$ to refer to the concentration of the species. The individual equations for a given species are written

Missing. Figure out notation here. BST?

The collection of reactions are written as

$$\dot{x} = Nv(x, \mu), \qquad \dot{x}_i = N_{ij} v_j(x, \mu)$$

where x_i is the concentration of species S_i, $N \in \mathbb{R}^{n \times m}$ is the stochiometry matrix, v_j is the reaction flux vector for reaction j, and μ is the collection of parameters that the define the reaction rates.

Figures

In the public version of the text, certain copyrighted figures are missing. These filenames for these figures are listed and the figures can be looked up in the following references:

- `Cou08` - *Mechanisms in Transcriptional Regulation* by A. J. Courey [11]
- `GNM93` - J. Greenblatt, J. R. Nodwell and S. W. Mason [**?**]
- `Mad07` - *From a to alpha: Yeast as a Model for Cellular Differentiation* by H. Madhani [28]
- `MBoC` - *The Molecular Biology of the Cell* by Alberts et al. [2]
- `PKT08` - *Physical Biology of the Cell* [35]

The remainder of the filename lists the chapter and figure number.

Chapter 1
Cell Biology Primer

Note: The text and figures in this chapter are based on *A Science Primer* by the National Center for Biotechnology Information (NCBI) of the National Library of Medicine (NLM) at the National Institutes of Health (NIH) [33]. The text in this chapter is not subject to copyright and may be used freely for any purpose, as described by the NLM:

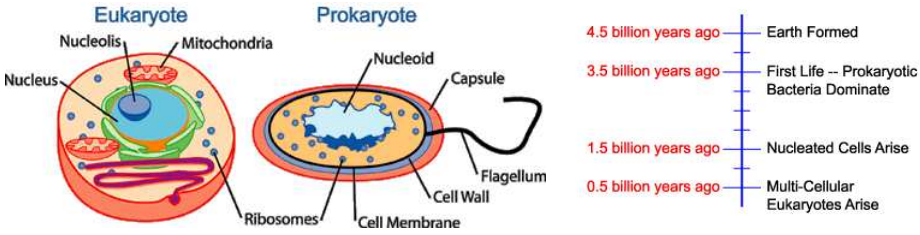
> *Information that is created by or for the US government on this site is within the public domain. Public domain information on the National Library of Medicine (NLM) Web pages may be freely distributed and copied. However, it is requested that in any subsequent use of this work, NLM be given appropriate acknowledgment.*

Some minor modifications have been made, including insertion of additional figures (from the NHGRI Talking Glossary [34]), deletion of some of the text not needed here, and minor editorial changes to maintain consistency with the main text.

The original material included here can be retrieved from the following web sites:

- `http://www.ncbi.nlm.nih.gov/About/primer/genetics.html`
- `http://www.genome.gov/glossary`

We gratefully acknowledge the National Library of Medicine for this material.



(a) Cell types (b) Timeline

Figure 1.1: Eukaryotes and prokaryotes. (a) This figure illustrates a typical human cell (*eukaryote*) and a typical bacterium (*prokaryote*). The drawing on the left highlights the internal structures of eukaryotic cells, including the nucleus (light blue), the nucleolus (intermediate blue), mitochondria (orange), and ribosomes (dark blue). The drawing on the right demonstrates how bacterial DNA is housed in a structure called the nucleoid (very light blue), as well as other structures normally found in a prokaryotic cell, including the cell membrane (black), the cell wall (intermediate blue), the capsule (orange), ribosomes (dark blue), and a flagellum (also black). (b) History of life on earth. Figures courtesy the National Library of Medicine.

1.1 What is a Cell

Cells are the structural and functional units of all living organisms. Some organisms, such as bacteria, are unicellular, consisting of a single cell. Other organisms, such as humans, are multicellular, or have many cells—an estimated 100,000,000,000,000 cells! Each cell is an amazing world unto itself: it can take in nutrients, convert these nutrients into energy, carry out specialized functions, and reproduce as necessary. Even more amazing is that each cell stores its own set of instructions for carrying out each of these activities.

Cell Organization

Before we can discuss the various components of a cell, it is important to know what organism the cell comes from. There are two general categories of cells: *prokaryotes* and *eukaryotes* (see Figure 1.1a).

Prokaryotic Organisms

It appears that life arose on earth about 4 billion years ago (see Figure 1.1b. The simplest of cells, and the first types of cells to evolve, were prokaryotic cells—organisms that lack a nuclear membrane, the membrane that surrounds the nucleus of a cell. Bacteria are the best known and most studied form of prokaryotic organisms, although the recent discovery of a second group of prokaryotes, called *archaea*, has provided evidence of a third cellular domain of life and new insights into the origin of life itself.

Prokaryotes are unicellular organisms that do not develop or differentiate into multicellular forms. Some bacteria grow in filaments, or masses of cells, but each cell in the colony is identical and capable of independent existence. The cells may be adjacent to one another because they did not separate after cell division or because they remained enclosed in a common sheath or slime secreted by the cells. Typically though, there is no continuity or communication between the cells. Prokaryotes are capable of inhabiting almost every place on the earth, from the deep ocean, to the edges of hot springs, to just about every surface of our bodies.

Prokaryotes are distinguished from eukaryotes on the basis of nuclear organization, specifically their lack of a nuclear membrane. Prokaryotes also lack any of the intracellular organelles and structures that are characteristic of eukaryotic cells. Most of the functions of organelles, such as mitochondria, chloroplasts, and the Golgi apparatus, are taken over by the prokaryotic plasma membrane. Prokaryotic cells have three architectural regions: appendages called *flagella* and *pili*—proteins attached to the cell surface; a *cell envelope* consisting of a capsule, a *cell wall*, and a *plasma membrane*; and a *cytoplasmic region* that contains the *cell genome* (DNA) and ribosomes and various sorts of inclusions.

Eukaryotic Organisms

Eukaryotes include fungi, animals, and plants as well as some unicellular organisms. Eukaryotic cells are about 10 times the size of a prokaryote and can be as much as 1000 times greater in volume. The major and extremely significant difference between prokaryotes and eukaryotes is that eukaryotic cells contain membrane-bound compartments in which specific metabolic activities take place. Most important among these is the presence of a nucleus, a membrane-delineated compartment that houses the eukaryotic cell's DNA. It is this nucleus that gives the eukaryote—literally, true nucleus—its name.

Eukaryotic organisms also have other specialized structures, called *organelles*, which are small structures within cells that perform dedicated functions. As the name implies, you can think of organelles as small organs. There are a dozen different types of organelles commonly found in eukaryotic cells. In this primer, we will focus our attention on only a handful of organelles and will examine these organelles with an eye to their role at a molecular level in the cell.

The origin of the eukaryotic cell was a milestone in the evolution of life. Although eukaryotes use the same genetic code and metabolic processes as prokaryotes, their higher level of organizational complexity has permitted the development of truly multicellular organisms. Without eukaryotes, the world would lack mammals, birds, fish, invertebrates, mushrooms, plants, and complex single-celled organisms.

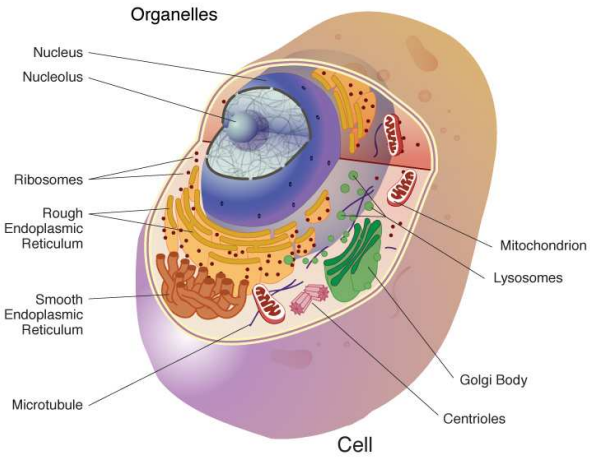


Figure 1.2: An organelle is a subcellular structure that has one or more specific jobs to perform in the cell, much like an organ does in the body. Among the more important cell organelles are the nuclei, which store genetic information; mitochondria, which produce chemical energy; and ribosomes, which assemble proteins.

Cell Structures: The Basics

The Plasma Membrane—A Cell's Protective Coat

The outer lining of a eukaryotic cell is called the *plasma membrane*. This membrane serves to separate and protect a cell from its surrounding environment and is made mostly from a double layer of proteins and lipids, fat-like molecules. Embedded within this membrane are a variety of other molecules that act as channels and pumps, moving different molecules into and out of the cell. A form of plasma membrane is also found in prokaryotes, but in this organism it is usually referred to as the *cell membrane*.

The Cytoskeleton—A Cell's Scaffold

The *cytoskeleton* is an important, complex, and dynamic cell component. It acts to organize and maintain the cell's shape; anchors organelles in place; helps during *endocytosis*, the uptake of external materials by a cell; and moves parts of the cell in processes of growth and motility. There are a great number of proteins associated with the cytoskeleton, each controlling a cell's structure by directing, bundling, and aligning filaments.

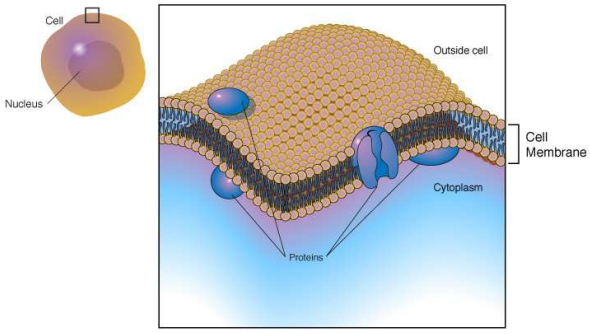


Figure 1.3: The cell membrane, also called the plasma membrane, is found in all cells and separates the interior of the cell from the outside environment. The cell membrane consists of a lipid bilayer that is semipermeable. The cell membrane regulates the transport of materials entering and exiting the cell.

The Cytoplasm—A Cell's Inner Space

Inside the cell there is a large fluid-filled space called the *cytoplasm*, sometimes called the *cytosol*. In prokaryotes, this space is relatively free of compartments. In eukaryotes, the *cytosol* is the "soup" within which all of the cell's organelles reside. It is also the home of the cytoskeleton. The cytosol contains dissolved nutrients, helps break down waste products, and moves material around the cell through a process called *cytoplasmic streaming*. The nucleus often flows with the cytoplasm changing its shape as it moves. The cytoplasm also contains many salts and is an excellent conductor of electricity, creating the perfect environment for the mechanics of the cell. The function of the cytoplasm, and the organelles which reside in it, are critical for a cell's survival.

Genetic Material

Two different kinds of genetic material exist: *deoxyribonucleic acid (DNA)* and *ribonucleic acid (RNA)*. Most organisms are made of DNA, but a few viruses have RNA as their genetic material. The biological information contained in an organism is encoded in its DNA or RNA sequence. Prokaryotic genetic material is organized in a simple circular structure that rests in the cytoplasm. Eukaryotic genetic material is more complex and is divided into discrete units called *genes*. Human genetic material is made up of two distinct components: the *nuclear genome* and the *mitochondrial genome*. The nuclear genome is divided into 24 linear DNA molecules, each contained in a different *chromosome*. The *mitochondrial genome* is a circular DNA molecule separate from the nuclear DNA. Although the mitochondrial genome is very small, it codes for some very important proteins.

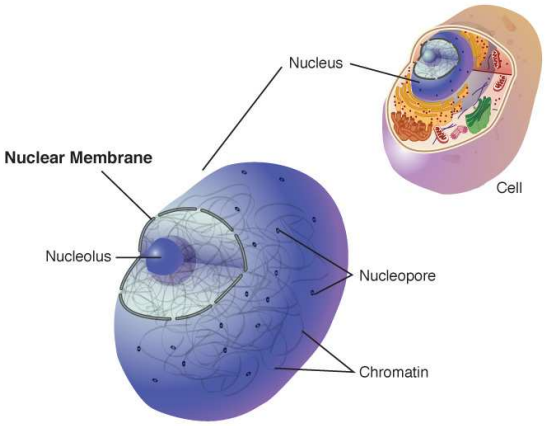


Figure 1.4: A nuclear membrane is a double membrane that encloses the cell nucleus. It serves to separate the chromosomes from the rest of the cell. The nuclear membrane includes an array of small holes or pores that permit the passage of certain materials, such as nucleic acids and proteins, between the nucleus and cytoplasm.

Organelles

The human body contains many different organs, such as the heart, lung, and kidney, with each organ performing a different function. Cells also have a set of "little organs", called *organelles*, that are adapted and/or specialized for carrying out one or more vital functions. Organelles are found only in eukaryotes and are always surrounded by a protective membrane. It is important to know some basic facts about the following organelles.

The Nucleus—A Cell's Center. The *nucleus* is the most conspicuous organelle found in a eukaryotic cell. It houses the cell's chromosomes and is the place where almost all DNA replication and RNA synthesis occur. The nucleus is spheroid in shape and separated from the cytoplasm by a membrane called the *nuclear envelope*. The nuclear envelope isolates and protects a cell's DNA from various molecules that could accidentally damage its structure or interfere with its processing. During processing, DNA is *transcribed*, or synthesized, into a special RNA, called mRNA. This mRNA is then transported out of the nucleus, where it is translated into a specific protein molecule. In prokaryotes, DNA processing takes place in the cytoplasm.

The Ribosome—The Protein Production Machine. Ribosomes are found in both prokaryotes and eukaryotes. The ribosome is a large complex composed of many molecules, including RNAs and proteins, and is responsible for processing the genetic instructions carried by an mRNA. The process of converting an mRNA's genetic code into the exact sequence of amino acids that make up a protein is called *translation*. Protein synthesis is extremely important to all cells, and there-

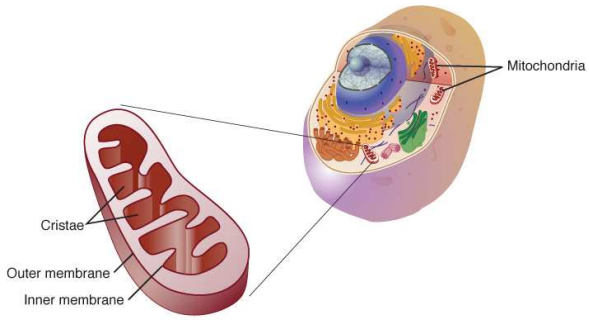


Figure 1.5: Mitochondria are membrane-bound cell organelles (mitochondrion, singular) that generate most of the chemical energy needed to power the cell's biochemical reactions. Chemical energy produced by the mitochondria is stored in a small molecule called adenosine triphosphate (ATP). Mitochondria contain their own small chromosomes. Generally, mitochondria, and therefore mitochondrial DNA, are inherited only from the mother.

fore a large number of ribosomes—sometimes hundreds or even thousands—can be found throughout a cell.

Ribosomes float freely in the cytoplasm or sometimes bind to another organelle called the endoplasmic reticulum. Ribosomes are composed of one large and one small subunit, each having a different function during protein synthesis.

Mitochondria and Chloroplasts—The Power Generators. Mitochondria are self-replicating organelles that occur in various numbers, shapes, and sizes in the cytoplasm of all eukaryotic cells. As mentioned earlier, mitochondria contain their own genome that is separate and distinct from the nuclear genome of a cell. Mitochondria have two functionally distinct membrane systems separated by a space: the *outer membrane*, which surrounds the whole organelle; and the *inner membrane*, which is thrown into folds or shelves that project inward. These inward folds are called *cristae*. The number and shape of cristae in mitochondria differ, depending on the tissue and organism in which they are found, and serve to increase the surface area of the membrane.

Mitochondria play a critical role in generating energy in the eukaryotic cell, and this process involves a number of complex pathways. Let's break down each of these steps so that you can better understand how food and nutrients are turned into energy packets and water. Some of the best energy-supplying foods that we eat contain complex sugars. These complex sugars can be broken down into a less chemically complex sugar molecule called *glucose*. Glucose can then enter the cell through special molecules found in the membrane, called *glucose transporters*. Once inside the cell, glucose is broken down to make *adenosine triphosphate (ATP)*, a form of energy, via two different pathways.

The first pathway, *glycolysis*, requires no oxygen and is referred to as *anaerobic*

(a) Glycolysis (b) Kreb's cycle

Figure 1.6: Cell energy production. Reproduced from Alberts et al. [2]; permission pending.

metabolism. Glycolysis occurs in the cytoplasm outside the mitochondria. During glycolysis, glucose is broken down into a molecule called *pyruvate*. Each reaction is designed to produce some hydrogen ions that can then be used to make energy packets (*ATP*). However, only four ATP molecules can be made from one molecule of glucose in this pathway. In prokaryotes, glycolysis is the only method used for converting energy.

The second pathway, called the *Kreb's cycle*, or the *citric acid cycle*, occurs inside the mitochondria and is capable of generating enough ATP to run all the cell functions. Once again, the cycle begins with a glucose molecule, which during the process of glycolysis is stripped of some of its hydrogen atoms, transforming the glucose into two molecules of *pyruvic acid*. Next, pyruvic acid is altered by the removal of a carbon and two oxygens, which go on to form carbon dioxide. When the *carbon dioxide* is removed, energy is given off, and a molecule called NAD+ is converted into the higher energy form, NADH. Another molecule, *coenzyme A (CoA)*, then attaches to the remaining acetyl unit, forming *acetyl CoA*.

Acetyl CoA enters the Kreb's cycle by joining to a four-carbon molecule called *oxaloacetate*. Once the two molecules are joined, they make a six-carbon molecule called citric acid. Citric acid is then broken down and modified in a stepwise fash-

ion. As this happens, hydrogen ions and carbon molecules are released. The carbon molecules are used to make more carbon dioxide. The hydrogen ions are picked up by NAD and another molecule called *flavin-adenine dinucleotide (FAD)*. Eventually, the process produces the four-carbon oxaloacetate again, ending up where it started off. All in all, the Kreb's cycle is capable of generating from 24 to 28 ATP molecules from one molecule of glucose converted to pyruvate. Therefore, it is easy to see how much more energy we can get from a molecule of glucose if our mitochondria are working properly and if we have oxygen.

Chloroplasts are similar to mitochondria but are found only in plants. Both organelles are surrounded by a double membrane with an intermembrane space; both have their own DNA and are involved in energy metabolism; and both have reticulations, or many foldings, filling their inner spaces. Chloroplasts convert light energy from the sun into ATP through a process called *photosynthesis*.

The Endoplasmic Reticulum and the Golgi Apparatus—Macromolecule Managers. The *endoplasmic reticulum (ER)* is the transport network for molecules targeted for certain modifications and specific destinations, as compared to molecules that will float freely in the cytoplasm. The ER has two forms: the *rough ER* and the *smooth ER*. The rough ER is labeled as such because it has ribosomes adhering to its outer surface, whereas the smooth ER does not. Translation of the mRNA for those proteins that will either stay in the ER or be *exported* (moved out of the cell) occurs at the ribosomes attached to the rough ER. The smooth ER serves as the recipient for those proteins synthesized in the rough ER. Proteins to be exported are passed to the *Golgi apparatus*, sometimes called a Golgi body or Golgi complex, for further processing, packaging, and transport to a variety of other cellular locations.

Lysosomes and Peroxisomes—The Cellular Digestive System. Lysosomes and *peroxisomes* are often referred to as the garbage disposal system of a cell. Both organelles are somewhat spherical, bound by a single membrane, and rich in digestive enzymes, naturally occurring proteins that speed up biochemical processes. For example, lysosomes can contain more than three dozen enzymes for degrading proteins, nucleic acids, and certain sugars called polysaccharides. All of these enzymes work best at a low pH, reducing the risk that these enzymes will digest their own cell should they somehow escape from the lysosome. Here we can see the importance behind compartmentalization of the eukaryotic cell. The cell could not house such destructive enzymes if they were not contained in a membrane-bound system.

One function of a lysosome is to digest foreign bacteria that invade a cell. Other functions include helping to recycle receptor proteins and other membrane components and degrading worn out organelles such as mitochondria. Lysosomes can even help repair damage to the plasma membrane by serving as a membrane patch, sealing the wound.

Peroxisomes function to rid the body of toxic substances, such as hydrogen

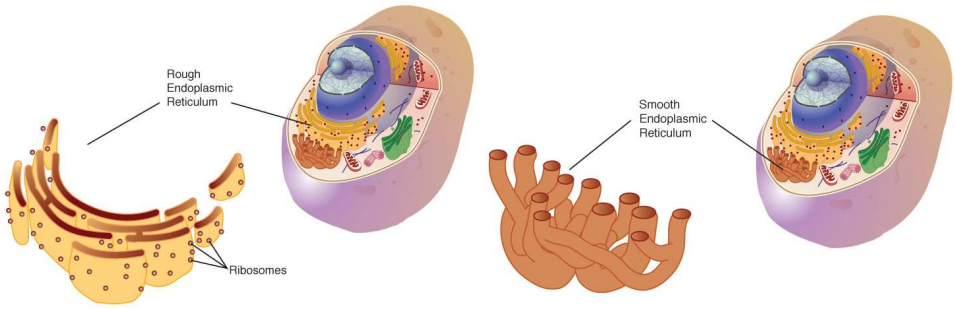


Figure 1.7: Endoplasmic reticulum is a network of membranes inside a cell through which proteins and other molecules move. Proteins are assembled at organelles called ribosomes. (a) When proteins are destined to be part of the cell membrane or exported from the cell, the ribosomes assembling them attach to the endoplasmic reticulum, giving it a rough appearance. (b) Smooth endoplasmic reticulum lacks ribosomes and helps synthesize and concentrate various substances needed by the cell.

peroxide, or other metabolites and contain enzymes concerned with oxygen utilization. High numbers of peroxisomes can be found in the liver, where toxic byproducts are known to accumulate. All of the enzymes found in a peroxisome are imported from the cytosol. Each enzyme transferred to a peroxisime has a special sequence at one end of the protein, called a *PTS* or *peroxisomal targeting signal*, that allows the protein to be taken into that organelle, where they then function to rid the cell of toxic substances.

Peroxisomes often resemble a lysosome. However, peroxisomes are self replicating, whereas lysosomes are formed in the Golgi complex. Peroxisomes also have membrane proteins that are critical for various functions, such as for importing proteins into their interiors and to proliferate and segregate into daughter cells.

Where Do Viruses Fit?

Viruses are not classified as cells and therefore are neither unicellular nor multicellular organisms. Most people do not even classify viruses as “living” because they lack a metabolic system and are dependent on the host cells that they infect to reproduce. Viruses have genomes that consist of either DNA or RNA, and there are examples of viruses that are either double-stranded or single-stranded. Importantly, their genomes code not only for the proteins needed to package its genetic material but for those proteins needed by the virus to reproduce during its infective cycle.

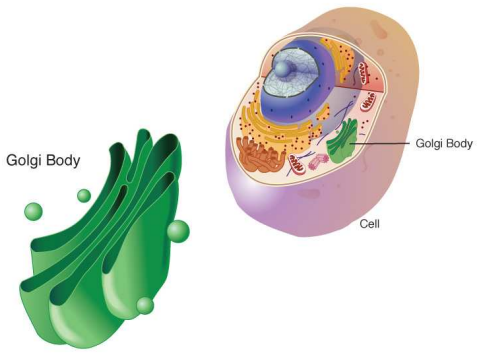


Figure 1.8: A Golgi body, also known as a Golgi apparatus, is a cell organelle that helps process and package proteins and lipid molecules, especially proteins destined to be exported from the cell. Named after its discoverer, Camillo Golgi, the Golgi body appears as a series of stacked membranes.

Making New Cells and Cell Types

For most unicellular organisms, reproduction is a simple matter of *cell duplication*, also known as *replication*. But for multicellular organisms, cell replication and reproduction are two separate processes. Multicellular organisms replace damaged or worn out cells through a replication process called *mitosis*, the division of a eukaryotic cell nucleus to produce two identical *daughter nuclei*. To reproduce, eukaryotes must first create special cells called *gametes*—eggs and sperm—that then fuse to form the beginning of a new organism. Gametes are but one of the many unique cell types that multicellular organisms need to function as a complete organism.

Making New Cells

Most unicellular organisms create their next generation by replicating all of their parts and then splitting into two cells, a type of *asexual reproduction* called *binary fission*. This process spawns not just two new cells, but also two new organisms. Multicellullar organisms replicate new cells in much the same way. For example, we produce new skin cells and liver cells by replicating the DNA found in that cell through mitosis. Yet, producing a whole new organism requires *sexual reproduction*, at least for most multicellular organisms. In the first step, specialized cells called *gametes*—eggs and sperm—are created through a process called meiosis. *Meiosis* serves to reduce the chromosome number for that particular organism by half. In the second step, the sperm and egg join to make a single cell, which restores

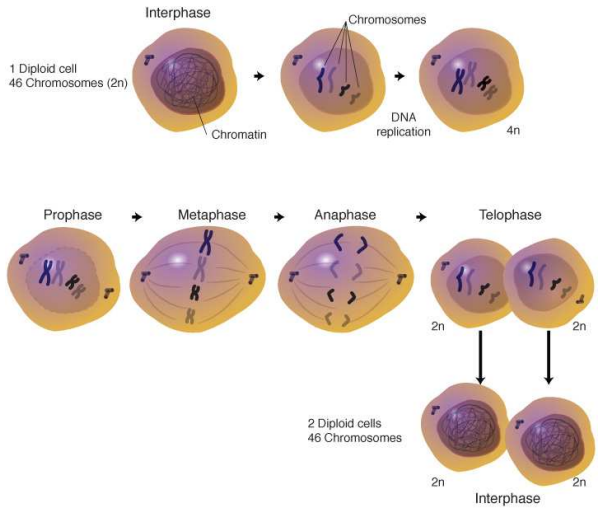


Figure 1.9: Mitosis is a cellular process that replicates chromosomes and produces two identical nuclei in preparation for cell division. Generally, mitosis is immediately followed by the equal division of the cell nuclei and other cell contents into two daughter cells.

the chromosome number. This joined cell then divides and differentiates into different cell types that eventually form an entire functioning organism.

Mitosis. Every time a cell divides, it must ensure that its DNA is shared between the two daughter cells. Mitosis is the process of "divvying up" the genome between the daughter cells. To easier describe this process, let's imagine a cell with only one chromosome. Before a cell enters mitosis, we say the cell is in *interphase*, the state of a eukaryotic cell when not undergoing division. Every time a cell divides, it must first replicate all of its DNA. Because chromosomes are simply DNA wrapped around protein, the cell replicates its chromosomes also. These two chromosomes, positioned side by side, are called *sister chromatids* and are identical copies of one another. Before this cell can divide, it must separate these sister chromatids from one another. To do this, the chromosomes have to condense. This stage of mitosis is called *prophase*. Next, the nuclear envelope breaks down, and a large protein network, called the *spindle*, attaches to each sister chromatid. The chromosomes are now aligned perpendicular to the spindle in a process called *metaphase*. Next, "molecular motors" pull the chromosomes away from the metaphase plate to the spindle poles of the cell. This is called *anaphase*. Once this process is completed, the cells divide, the nuclear envelope reforms, and the chromosomes relax and decondense during *telophase*. The cell can now replicate its DNA again during interphase and go through mitosis once more.

Meiosis. *Meiosis* is a specialized type of cell division that occurs during the formation of gametes. Although meiosis may seem much more complicated than mitosis,

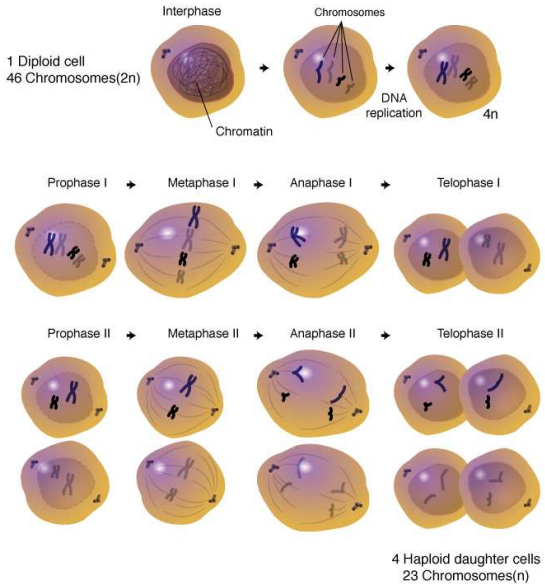


Figure 1.10: Meiosis is the formation of egg and sperm cells. In sexually reproducing organisms, body cells are diploid, meaning they contain two sets of chromosomes (one set from each parent). To maintain this state, the egg and sperm that unite during fertilization must be haploid, meaning they each contain a single set of chromosomes. During meiosis, diploid cells undergo DNA replication, followed by two rounds of cell division, producing four haploid sex cells.

it is really just two cell divisions in sequence. Each of these sequences maintains strong similarities to mitosis.

Meiosis I refers to the first of the two divisions and is often called the *reduction division*. This is because it is here that the chromosome complement is reduced from *diploid* (two copies) to *haploid* (one copy). Interphase in meiosis is identical to interphase in mitosis. At this stage, there is no way to determine what type of division the cell will undergo when it divides. Meiotic division will only occur in cells associated with male or female sex organs. *Prophase I* is virtually identical to prophase in mitosis, involving the appearance of the *chromosomes*, the development of the spindle apparatus, and the breakdown of the nuclear membrane. Metaphase I is where the critical difference occurs between meiosis and mitosis. In mitosis, all of the chromosomes line up on the metaphase plate in no particular order. In Metaphase I, the chromosome pairs are aligned on either side of the metaphase plate. It is during this alignment that the chromatid arms may overlap and temporarily fuse, resulting in what is called *crossovers*. During *Anaphase I*, the spindle fibers contract, pulling the homologous pairs away from each other and toward each pole of the cell. In *Telophase I*, a cleavage furrow typically forms, followed by *cytokinesis*, the changes that occur in the cytoplasm of a cell during

nuclear division; but the nuclear membrane is usually not reformed, and the chromosomes do not disappear. At the end of Telophase I, each daughter cell has a single set of chromosomes, half the total number in the original cell, that is, while the original cell was diploid; the daughter cells are now haploid.

Meiosis II is quite simply a mitotic division of each of the haploid cells produced in Meiosis I. There is no Interphase between Meiosis I and Meiosis II, and the latter begins with *Prophase II*. At this stage, a new set of spindle fibers forms and the chromosomes begin to move toward the equator of the cell. During *Metaphase II*, all of the chromosomes in the two cells align with the metaphase plate. In *Anaphase II*, the centromeres split, and the spindle fibers shorten, drawing the chromosomes toward each pole of the cell. In *Telophase II*, a cleavage furrow develops, followed by cytokinesis and the formation of the nuclear membrane. The chromosomes begin to fade and are replaced by the *granular chromatin*, a characteristic of interphase. When Meiosis II is complete, there will be a total of four daughter cells, each with half the total number of chromosomes as the original cell. In the case of *male structures*, all four cells will eventually develop into *sperm cells*. In the case of the *female life cycles* in higher organisms, three of the cells will typically abort, leaving a single cell to develop into an egg cell, which is much larger than a sperm cell.

Recombination—The Physical Exchange of DNA. All organisms suffer a certain number of small *mutations*, or random changes in a DNA sequence, during the process of DNA replication. These are called *spontaneous mutations* and occur at a rate characteristic for that organism. *Genetic recombination* refers more to a large-scale rearrangement of a DNA molecule. This process involves pairing between complementary strands of two parental duplex, or double-stranded DNAs, and results from a physical exchange of chromosome material.

The position at which a gene is located on a chromosome is called a *locus*. In a given individual, one might find two different versions of this gene at a particular locus. These alternate gene forms are called *alleles*. During Meiosis I, when the chromosomes line up along the metaphase plate, the two strands of a chromosome pair may physically cross over one another. This may cause the strands to break apart at the crossover point and reconnect to the other chromosome, resulting in the exchange of part of the chromosome.

Recombination results in a new arrangement of maternal and paternal alleles on the same chromosome. Although the same genes appear in the same order, the alleles are different. This process explains why offspring from the same parents can look so different. In this way, it is theoretically possible to have any combination of parental alleles in an offspring, and the fact that two alleles appear together in one offspring does not have any influence on the statistical probability that another offspring will have the same combination. This theory of "*independent assortment*" of alleles is fundamental to genetic inheritance. However, having said that, there is an exception that requires further discussion.

The frequency of recombination is actually not the same for all gene combinations. This is because recombination is greatly influenced by the proximity of one gene to another. If two genes are located close together on a chromosome, the likelihood that a recombination event will separate these two genes is less than if they were farther apart. *Linkage* describes the tendency of genes to be inherited together as a result of their location on the same chromosome. *Linkage disequilibrium* describes a situation in which some combinations of genes or genetic markers occur more or less frequently in a population than would be expected from their distances apart. Scientists apply this concept when searching for a gene that may cause a particular disease. They do this by comparing the occurrence of a specific DNA sequence with the appearance of a disease. When they find a high correlation between the two, they know they are getting closer to finding the appropriate gene sequence.

Binary Fission—How Bacteria Reproduce. Bacteria reproduce through a fairly simple process called *binary fission*, or the reproduction of a living cell by division into two equal, or near equal, parts. As just noted, this type of asexual reproduction theoretically results in two identical cells. However, bacterial DNA has a relatively high mutation rate. This rapid rate of genetic change is what makes bacteria capable of developing resistance to antibiotics and helps them exploit invasion into a wide range of environments.

Similar to more complex organisms, bacteria also have mechanisms for exchanging genetic material. Although not equivalent to sexual reproduction, the end result is that a bacterium contains a combination of traits from two different *parental* cells. Three different modes of exchange have thus far been identified in bacteria.

Conjunction involves the direct joining of two bacteria, which allows their circular DNAs to undergo recombination. Bacteria can also undergo *transformation* by absorbing remnants of DNA from dead bacteria and integrating these fragments into their own DNA. Lastly, bacteria can exchange genetic material through a process called *transduction*, in which genes are transported into and out of the cell by bacterial viruses, called *bacteriophages*, or by *plasmids*, an autonomous self-replicating extrachromosomal circular DNA.

Viral Reproduction. Because viruses are acellular and do not use ATP, they must utilize the machinery and metabolism of a host cell to reproduce. For this reason, viruses are called *obligate intracellular parasites*. Before a virus has entered a host cell, it is called a virion—a package of viral genetic material. *Virions*—infectious viral particles—can be passed from host to host either through direct contact or through a vector, or carrier. Inside the organism, the virus can enter a cell in various ways. Bacteriophages—bacterial viruses—attach to the cell wall surface in specific places. Once attached, enzymes make a small hole in the cell wall, and the virus injects its DNA into the cell. Other viruses (such as HIV) enter the host

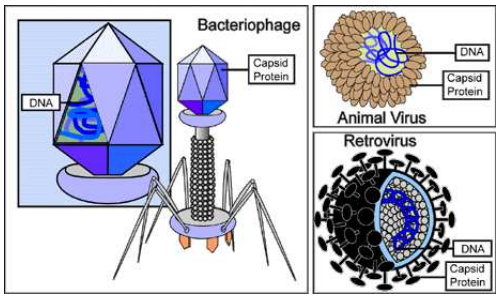


Figure 1.11: Types of viruses. This illustration depicts three types of viruses: a bacterial virus, otherwise called a bacteriophage (left center); an animal virus (top right); and a retrovirus (bottom right). Viruses depend on the host cell that they infect to reproduce. When found outside of a host cell, viruses, in their simplest forms, consist only of genomic nucleic acid, either DNA or RNA (depicted as blue), surrounded by a protein coat, or capsid.

via *endocytosis*, the process whereby cells take in material from the external environment. After entering the cell, the virus's genetic material begins the destructive process of taking over the cell and forcing it to produce new viruses.

There are three different ways genetic information contained in a viral genome can be reproduced. The form of genetic material contained in the *viral capsid*, the protein coat that surrounds the nucleic acid, determines the exact replication process. Some viruses have DNA, which once inside the host cell is replicated by the host along with its own DNA. Then, there are two different replication processes for viruses containing RNA. In the first process, the viral RNA is directly copied using an enzyme called *RNA replicase*. This enzyme then uses that RNA copy as a template to make hundreds of duplicates of the original RNA. A second group of RNA-containing viruses, called the *retroviruses*, uses the enzyme reverse transcriptase to synthesize a complementary strand of DNA so that the virus's genetic information is contained in a molecule of DNA rather than RNA. The viral DNA can then be further replicated using the host cell machinery.

Steps Associated with Viral Reproduction.

1. *Attachment*, sometimes called *absorption*: The virus attaches to receptors on the host cell wall.

2. *Penetration*: The nucleic acid of the virus moves through the plasma membrane and into the cytoplasm of the host cell. The capsid of a phage, a bacterial virus, remains on the outside. In contrast, many viruses that infect animal cells enter the host cell intact.

3. *Replication*: The viral genome contains all the information necessary to produce new viruses. Once inside the host cell, the virus induces the host cell to synthesize the necessary components for its replication.

4. *Assembly*: The newly synthesized viral components are assembled into new viruses.

5. *Release*: Assembled viruses are released from the cell and can now infect other cells, and the process begins again.

When the virus has taken over the cell, it immediately directs the host to begin manufacturing the proteins necessary for virus reproduction. The host produces three kinds of proteins: *early proteins*, enzymes used in nucleic acid replication; *late proteins*, proteins used to construct the virus coat; and *lytic proteins*, enzymes used to break open the cell for viral exit. The final viral product is assembled spontaneously, that is, the parts are made separately by the host and are joined together by chance. This self-assembly is often aided by molecular *chaperones*, or proteins made by the host that help the capsid parts come together.

The new viruses then leave the cell either by exocytosis or by lysis. Envelope-bound animal viruses instruct the host's endoplasmic reticulum to make certain proteins, called *glycoproteins*, which then collect in clumps along the cell membrane. The virus is then discharged from the cell at these exit sites, referred to as exocytosis. On the other hand, bacteriophages must break open, or *lyse*, the cell to exit. To do this, the phages have a gene that codes for an enzyme called *lysozyme*. This enzyme breaks down the cell wall, causing the cell to swell and burst. The new viruses are released into the environment, killing the host cell in the process.

One family of animal viruses, called the retroviruses, contains RNA genomes in their virus particles but synthesize a DNA copy of their genome in infected cells. Retroviruses provide an excellent example of how viruses can play an important role as models for biological research. Studies of these viruses are what first demonstrated the synthesis of DNA from RNA templates, a fundamental mode for transferring genetic material that occurs in both eukaryotes and prokaryotes.

Why Study Viruses?. *Viruses* are important to the study of *molecular and cellular biology* because they provide simple systems that can be used to manipulate and investigate the functions of many cell types. We have just discussed how viral replication depends on the metabolism of the infected cell. Therefore, the study of viruses can provide fundamental information about aspects of cell biology and metabolism. The rapid growth and small genome size of bacteria make them excellent tools for experiments in biology. Bacterial viruses have also further simplified the study of bacterial genetics and have deepened our understanding of the basic mechanisms of molecular genetics. Because of the complexity of an animal cell genome, viruses have been even more important in studies of animal cells than in studies of bacteria. Numerous studies have demonstrated the utility of animal viruses as probes for investigating different activities of eukaryotic cells. Other examples in which animal viruses have provided important models for biological research of their host cells include studies of *DNA replication*, *transcription*, *RNA processing*, and *protein transport*.

Deriving New Cell Types

Look closely at the human body, and it is clear that not all cells are alike. For example, cells that make up our skin are certainly different from cells that make up our inner organs. Yet, all of the different cell types in our body are all *derived*, or arise, from a single, fertilized egg cell through differentiation. *Differentiation* is the process by which an unspecialized cell becomes specialized into one of the many cells that make up the body, such as a heart, liver, or muscle cell. During differentiation, certain genes are turned on, or become *activated*, while other genes are switched off, or *inactivated*. This process is intricately regulated. As a result, a differentiated cell will develop specific structures and perform certain functions.

Mammalian Cell Types. Three basic categories of cells make up the mammalian body: *germ cells*, *somatic cells*, and *stem cells*. Each of the approximately 100 trillion cells in an adult human has its own copy, or copies, of the genome, with the only exception being certain cell types that lack nuclei in their fully differentiated state, such as red blood cells. The majority of these cells are *diploid*, or have two copies of each chromosome. These cells are called *somatic cells*. This category of cells includes most of the cells that make up our body, such as skin and muscle cells. *Germ line cells* are any line of cells that give rise to *gametes*—eggs and sperm—and are continuous through the generations. *Stem cells*, on the other hand, have the ability to divide for indefinite periods and to give rise to specialized cells. They are best described in the context of normal human development.

Human development begins when a sperm fertilizes an egg and creates a single cell that has the potential to form an entire organism. In the first hours after fertilization, this cell divides into identical cells. Approximately 4 days after fertilization and after several cycles of cell division, these cells begin to specialize, forming a hollow sphere of cells, called a *blastocyst*. The blastocyst has an outer layer of cells, and inside this hollow sphere, there is a cluster of cells called the inner *cell mass*. The cells of the inner cell mass will go on to form virtually all of the tissues of the human body. Although the cells of the inner cell mass can form virtually every type of cell found in the human body, they cannot form an organism. Therefore, these cells are referred to as *pluripotent*, that is, they can give rise to many types of cells but not a whole organism. Pluripotent stem cells undergo further specialization into stem cells that are committed to give rise to cells that have a particular function. Examples include blood stem cells that give rise to red blood cells, white blood cells, and platelets, and skin stem cells that give rise to the various types of skin cells. These more specialized stem cells are called *multipotent*—capable of giving rise to several kinds of cells, tissues, or structures.

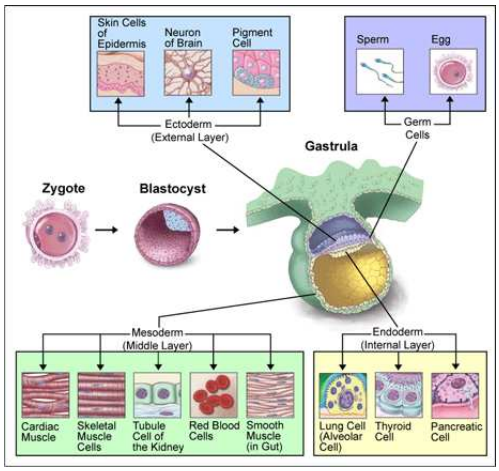


Figure 1.12: Differentiation of human tissues. Human development begins when a sperm fertilizes an egg and creates a single cell that has the potential to form an entire organism, called the zygote (top panel, mauve). In the first hours after fertilization, this cell divides into identical cells. These cells then begin to specialize, forming a hollow sphere of cells, called a blastocyst (second panel, purple). The blastocyst has an outer layer of cells (yellow), and inside this hollow sphere, there is a cluster of cells called the inner cell mass (light blue). The inner cell mass can give rise to the germ cells—eggs and sperm—as well as cells derived from all three germ layers (ectoderm, light blue; mesoderm, light green; and endoderm, light yellow), depicted in the bottom panel, including nerve cells, muscle cells, skin cells, blood cells, bone cells, and cartilage. Reproduced with permission from the Office of Science Policy, the National Institutes of Health.

The Working Cell: DNA, RNA, and Protein Synthesis

DNA Replication

DNA replication, or the process of duplicating a cell's genome, is required every time a cell divides. Replication, like all cellular activities, requires specialized proteins for carrying out the job. In the first step of replication, a special protein, called a *helicase*, unwinds a portion of the parental DNA double helix. Next, a molecule of *DNA polymerase*—a common name for two categories of enzymes that influence the synthesis of DNA— binds to one strand of the DNA. DNA polymerase begins to move along the DNA strand in the 3' to 5' direction, using the single-stranded DNA as a template. This newly synthesized strand is called the *leading strand* and is necessary for forming new nucleotides and reforming a double helix. Because DNA synthesis can only occur in the 5' to 3' direction, a second DNA polymerase molecule is used to bind to the other template strand as the double helix opens. This molecule synthesizes discontinuous segments of polynucleotides, called *Okazaki fragments*. Another enzyme, called *DNA ligase*, is responsible for stitching these fragments together into what is called the *lagging strand*.

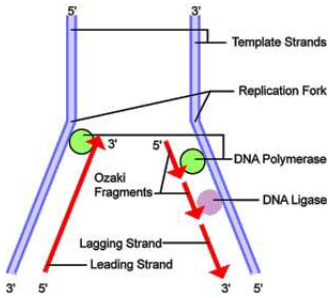


Figure 1.13: An overview of DNA replication. Before a cell can divide, it must first duplicate its DNA. This figure provides an overview of the DNA replication process. In the first step, a portion of the double helix (blue) is unwound by a helicase. Next, a molecule of DNA polymerase (green) binds to one strand of the DNA. It moves along the strand, using it as a template for assembling a leading strand (red) of nucleotides and reforming a double helix. Because DNA synthesis can only occur 5' to 3', a second DNA polymerase molecule (also green) is used to bind to the other template strand as the double helix opens. This molecule must synthesize discontinuous segments of polynucleotides (called Okazaki Fragments). Another enzyme, DNA Ligase (yellow), then stitches these together into the lagging strand.

The average human chromosome contains an enormous number of nucleotide pairs that are copied at about 50 base pairs per second. Yet, the entire replication process takes only about an hour. This is because there are many *replication origin sites* on a eukaryotic chromosome. Therefore, replication can begin at some origins earlier than at others. As replication nears completion, "bubbles" of newly replicated DNA meet and fuse, forming two new molecules.

With multiple replication origin sites, one might ask, how does the cell know which DNA has already been replicated and which still awaits replication? To date, two *replication control mechanisms* have been identified: one positive and one negative. For DNA to be replicated, each replication origin site must be bound by a set of proteins called the *Origin Recognition Complex*. These remain attached to the DNA throughout the replication process. Specific accessory proteins, called *licensing factors*, must also be present for initiation of replication. Destruction of these proteins after initiation of replication prevents further replication cycles from occurring. This is because licensing factors are only produced when the nuclear membrane of a cell breaks down during mitosis.

DNA Transcription—Making mRNA

DNA transcription refers to the synthesis of RNA from a DNA template. This process is very similar to DNA replication. Of course, there are different proteins that direct transcription. The most important enzyme is *RNA polymerase*, an enzyme that influences the synthesis of RNA from a DNA template. For transcription to

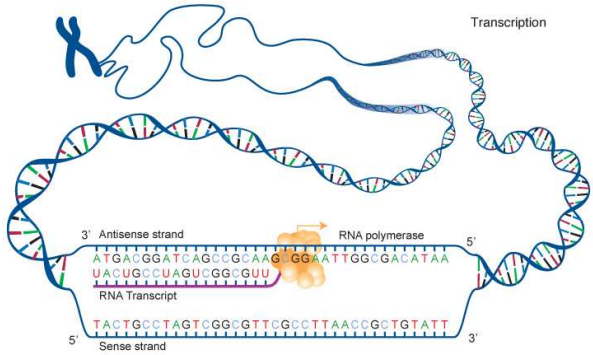


Figure 1.14: Transcription is the process of making an RNA copy of a gene sequence. This copy, called a messenger RNA (mRNA) molecule, leaves the cell nucleus and enters the cytoplasm, where it directs the synthesis of the protein, which it encodes.

be initiated, RNA polymerase must be able to recognize the beginning sequence of a gene so that it knows where to start synthesizing an mRNA. It is directed to this initiation site by the ability of one of its subunits to recognize a specific DNA sequence found at the beginning of a gene, called the *promoter sequence*. The promoter sequence is a unidirectional sequence found on one strand of the DNA that instructs the RNA polymerase in both where to start synthesis and in which direction synthesis should continue. The RNA polymerase then unwinds the double helix at that point and begins synthesis of a RNA strand complementary to one of the strands of DNA. This strand is called the *antisense* or *template strand*, whereas the other strand is referred to as the *sense* or coding strand. Synthesis can then proceed in a unidirectional manner.

Although much is known about transcript processing, the signals and events that instruct RNA polymerase to stop transcribing and drop off the DNA template remain unclear. Experiments over the years have indicated that processed eukaryotic messages contain a *poly(A) addition signal* (AAUAAA) at their 3' end, followed by a string of adenines. This poly(A) addition, also called the *poly(A) site*, contributes not only to the addition of the poly(A) tail but also to transcription termination and the release of RNA polymerase from the DNA template. Yet, transcription does not stop here. Rather, it continues for another 200 to 2000 bases beyond this site before it is aborted. It is either before or during this termination process that the nascent transcript is *cleaved*, or cut, at the poly(A) site, leading to the creation of two RNA molecules. The upstream portion of the newly formed, or *nascent*, RNA then undergoes further modifications, called *post-transcriptional modification*, and becomes mRNA. The downstream RNA becomes unstable and is rapidly degraded.

Although the importance of the poly(A) addition signal has been established, the contribution of sequences further downstream remains uncertain. A recent study

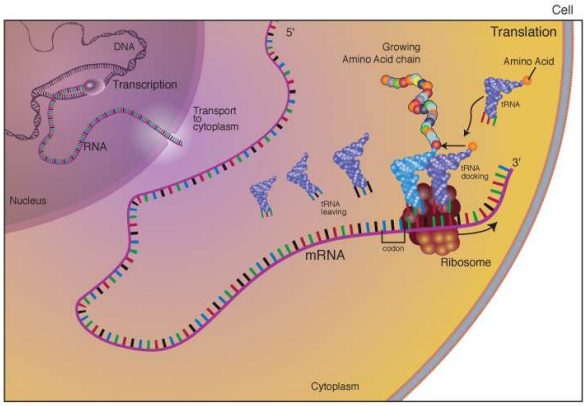


Figure 1.15: Translation is the process of translating the sequence of a messenger RNA (mRNA) molecule to a sequence of amino acids during protein synthesis. The genetic code describes the relationship between the sequence of base pairs in a gene and the corresponding amino acid sequence that it encodes. In the cell cytoplasm, the ribosome reads the sequence of the mRNA in groups of three bases to assemble the protein.

suggests that a defined region, called the *termination region*, is required for proper transcription termination. This study also illustrated that transcription termination takes place in two distinct steps. In the first step, the nascent RNA is cleaved at specific subsections of the termination region, possibly leading to its release from RNA polymerase. In a subsequent step, RNA polymerase disengages from the DNA. Hence, RNA polymerase continues to transcribe the DNA, at least for a short distance.

Protein Translation—How Do Messenger RNAs Direct Protein Synthesis?

The cellular machinery responsible for synthesizing proteins is the *ribosome*. The ribosome consists of structural RNA and about 80 different proteins. In its inactive state, it exists as two subunits: a *large subunit* and a *small subunit*. When the small subunit encounters an mRNA, the process of *translating* an mRNA to a protein begins. In the large subunit, there are two sites for amino acids to bind and thus be close enough to each other to form a bond. The "*A site*" accepts a new *transfer RNA*, or tRNA—the adaptor molecule that acts as a translator between mRNA and protein—bearing an amino acid. The "P site"*P site* binds the tRNA that becomes attached to the growing chain.

As we just discussed, the adaptor molecule that acts as a translator between mRNA and protein is a specific RNA molecule, the tRNA. Each tRNA has a specific *acceptor site* that binds a particular triplet of nucleotides, called a *codon*, and an *anti-codon site* that binds a sequence of three unpaired nucleotides, the anti-codon, which can then bind to the the codon. Each tRNA also has a specific

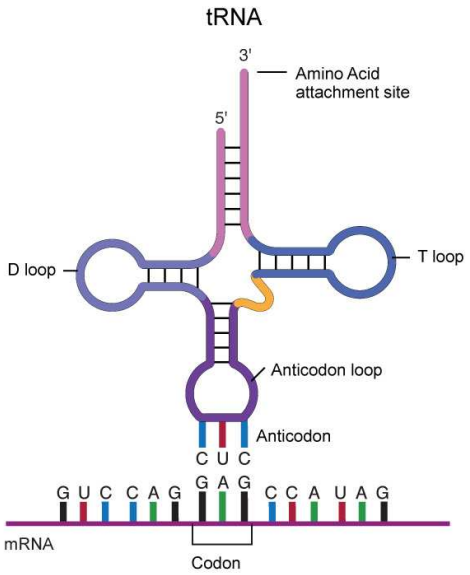


Figure 1.16: Transfer RNA (tRNA) is a small RNA molecule that participates in protein synthesis. Each tRNA molecule has two important areas: a trinucleotide region called the anticodon and a region for attaching a specific amino acid. During translation, each time an amino acid is added to the growing chain, a tRNA molecule forms base pairs with its complementary sequence on the messenger RNA (mRNA) molecule, ensuring that the appropriate amino acid is inserted into the protein.

charger protein, called an *aminoacyl tRNA synthetase*. This protein can only bind to that particular tRNA and attach the correct amino acid to the acceptor site.

The *start signal* for translation is the codon ATG, which codes for methionine. Not every protein necessarily starts with methionine, however. Oftentimes this first amino acid will be removed in later processing of the protein. A tRNA charged with methionine binds to the translation start signal. The large subunit binds to the mRNA and the small subunit, and so begins *elongation*, the formation of the polypeptide chain. After the first charged tRNA appears in the A site, the ribosome shifts so that the tRNA is now in the P site. New charged tRNAs, corresponding the codons of the mRNA, enter the A site, and a bond is formed between the two amino acids. The first tRNA is now released, and the ribosome shifts again so that a tRNA carrying two amino acids is now in the P site. A new charged tRNA then binds to the A site. This process of elongation continues until the ribosome reaches what is called a *stop codon*, a triplet of nucleotides that signals the termination of translation. When the ribosome reaches a stop codon, no aminoacyl tRNA binds to the empty A site. This is the ribosome signal to break apart into its large and small subunits, releasing the new protein and the mRNA. Yet, this isn't always the end of the story. A protein will often undergo further modification, called *post-*

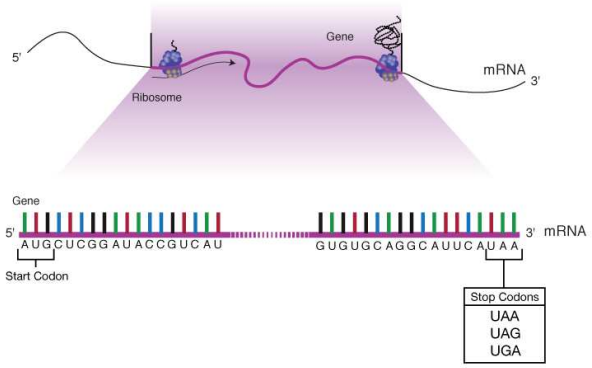


Figure 1.17: A stop codon is a trinucleotide sequence within a messenger RNA (mRNA) molecule that signals a halt to protein synthesis. The genetic code describes the relationship between the sequence of DNA bases (A, C, G, and T) in a gene and the corresponding protein sequence that it encodes. The cell reads the sequence of the gene in groups of three bases. Of the 64 possible combinations of three bases, 61 specify an amino acid, while the remaining three combinations are stop codons.

translational modification. For example, it might be cleaved by a protein-cutting enzyme, called a protease, at a specific place or have a few of its amino acids altered.

DNA Repair Mechanisms

Maintenance of the accuracy of the DNA genetic code is critical for both the long- and short-term survival of cells and species. Sometimes, normal cellular activities, such as duplicating DNA and making new gametes, introduce changes or *mutations* in our DNA. Other changes are caused by exposure of DNA to chemicals, radiation, or other adverse environmental conditions. No matter the source, genetic mutations have the potential for both positive and negative effects on an individual as well as its species. A positive change results in a slightly different version of a gene that might eventually prove beneficial in the face of a new disease or changing environmental conditions. Such beneficial changes are the cornerstone of evolution. Other mutations are considered *deleterious*, or result in damage to a cell or an individual. For example, errors within a particular DNA sequence may end up either preventing a vital protein from being made or encoding a defective protein. It is often these types of errors that lead to various disease states.

The potential for DNA damage is counteracted by a vigorous surveillance and repair system. Within this system, there are a number of enzymes capable of repairing damage to DNA. Some of these enzymes are specific for a particular type of damage, whereas others can handle a range of mutation types. These systems

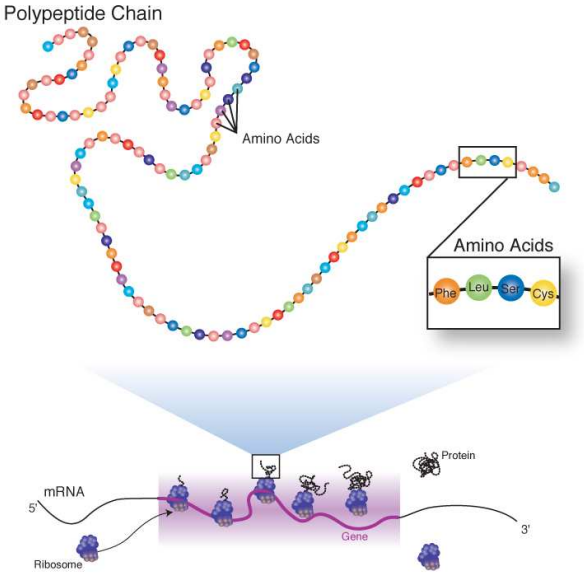


Figure 1.18: A peptide is one or more amino acids linked by chemical bonds. The term also refers to the type of chemical bond that joins the amino acids together. A series of linked amino acids is a polypeptide. The cell's proteins are made from one or more polypeptides.

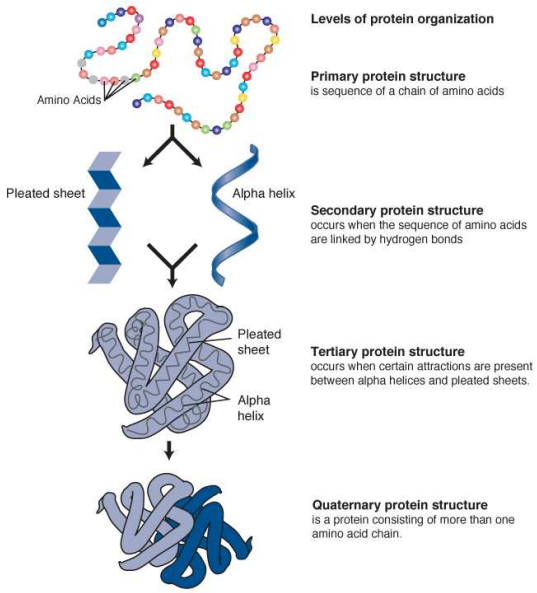


Figure 1.19: Proteins are an important class of molecules found in all living cells. A protein is composed of one or more long chains of amino acids, the sequence of which corresponds to the DNA sequence of the gene that encodes it. Proteins play a variety of roles in the cell, including structural (cytoskeleton), mechanical (muscle), biochemical (enzymes), and cell signaling (hormones). Proteins are also an essential part of diet.

also differ in the degree to which they are able to restore the normal, or *wild-type*, sequence.

Categories of DNA Repair Systems.

- *Photoreactivation* is the process whereby genetic damage caused by ultraviolet radiation is reversed by subsequent illumination with visible or near-ultraviolet light.

- *Nucleotide excision repair* is used to fix DNA lesions, such as single-stranded breaks or damaged bases, and occurs in stages. The first stage involves recognition of the damaged region. In the second stage, two enzymatic reactions serve to remove, or excise, the damaged sequence. The third stage involves synthesis by DNA polymerase of the excised nucleotides using the second intact strand of DNA as a template. Lastly, DNA ligase joins the newly synthesized segment to the existing ends of the originally damaged DNA strand.

- *Recombination repair*, or *post-replication repair*, fixes DNA damage by a strand exchange from the other daughter chromosome. Because it involves homologous recombination, it is largely error free.

- *Base excision repair* allows for the identification and removal of wrong bases, typically attributable to *deamination*—the removal of an amino group (NH2)—of normal bases as well as from chemical modification.

- *Mismatch repair* is a multi-enzyme system that recognizes inappropriately matched bases in DNA and replaces one of the two bases with one that "matches" the other. The major problem here is recognizing which of the mismatched bases is incorrect and therefore should be removed and replaced.

- *Adaptive/inducible repair* describes several protein activities that recognize very specific modified bases. They then transfer this modifying group from the DNA to themselves, and, in doing so, destroy their own function. These proteins are referred to as inducible because they tend to regulate their own synthesis. For example, exposure to modifying agents induces, or turns on, more synthesis and therefore adaptation.

- *SOS repair* or *inducible error-prone repair* is a repair process that occurs in bacteria and is induced, or switched on, in the presence of potentially lethal stresses, such as UV irradiation or the inactivation of genes essential for replication. Some responses to this type of stress include *mutagenesis*—the production of mutations—or cell elongation without cell division. In this type of repair process, replication of the DNA template is extremely inaccurate. Obviously, such a repair system must be a desperate recourse for the cell, allowing replication past a region where the wild-type sequence has been lost.

From Cells to Genomes

Understanding what makes up a cell and how that cell works is fundamental to all of the biological sciences. Appreciating the similarities and differences between cell types is particularly important to the fields of cell and molecular biology. These fundamental similarities and differences provide a unifying theme, allowing the principles learned from studying one cell type to be extrapolated and generalized to other cell types.

Perhaps the most fundamental property of all living things is their ability to reproduce. All cells arise from pre-existing cells, that is, their genetic material must be replicated and passed from parent cell to progeny. Likewise, all multicellular organisms inherit their genetic information specifying structure and function from their parents. The next section of the genetics primer, What is a Genome, details how genetic information is replicated and transmitted from cell to cell and organism to organism.

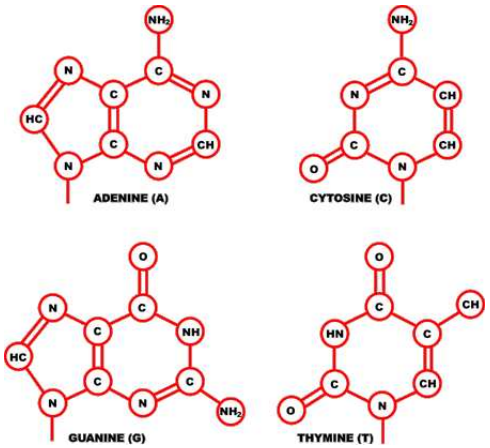


Figure 1.20: The four DNA bases. Each DNA base is made up of the sugar 2'-deoxyribose linked to a phosphate group and one of the four bases depicted above: adenine (top left), cytosine (top right), guanine (bottom left), and thymine (bottom right).

1.2 What is a Genome

Life is specified by *genomes*. Every organism, including humans, has a genome that contains all of the biological information needed to build and maintain a living example of that organism. The biological information contained in a genome is encoded in its *deoxyribonucleic acid (DNA)* and is divided into discrete units called *genes*. Genes code for proteins that attach to the genome at the appropriate positions and switch on a series of reactions called gene expression.

The Physical Structure of the Human Genome

Nuclear DNA

Inside each of our cells lies a *nucleus*, a membrane-bounded region that provides a sanctuary for genetic information. The nucleus contains long strands of DNA that encode this genetic information. A *DNA* chain is made up of four chemical bases: *adenine* (A) and *guanine* (G), which are called *purines*, and *cytosine* (C) and *thymine* (T), referred to as *pyrimidines*. Each base has a slightly different composition, or combination of oxygen, carbon, nitrogen, and hydrogen. In a DNA chain, every base is attached to a sugar molecule (deoxyribose) and a phosphate molecule, resulting in a nucleic acid or *nucleotide*. Individual nucleotides are linked through the phosphate group, and it is the precise order, or sequence, of nucleotides that determines the product made from that gene.

A DNA chain, also called a strand, has a sense of direction, in which one end is chemically different than the other. The so-called 5' end terminates in a 5' phosphate group (-PO4); the 3' end terminates in a 3' hydroxyl group (-OH). This is

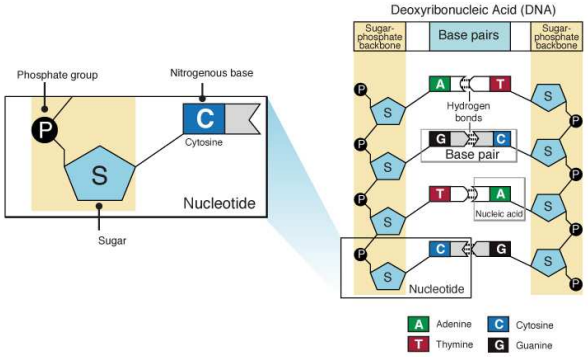


Figure 1.21: A nucleotide is the basic building block of nucleic acids. RNA and DNA are polymers made of long chains of nucleotides. A nucleotide consists of a sugar molecule (either ribose in RNA or deoxyribose in DNA) attached to a phosphate group and a nitrogen-containing base. The bases used in DNA are adenine (A), cytosine (C), guanine (G), and thymine (T). In RNA, the base uracil (U) takes the place of thymine.

important because DNA strands are always synthesized in the 5' to 3' direction.

The DNA that constitutes a gene is a double-stranded molecule consisting of two chains running in opposite directions. The chemical nature of the bases in double-stranded DNA creates a slight twisting force that gives DNA its characteristic gently coiled structure, known as the double helix. The two strands are connected to each other by chemical pairing of each base on one strand to a specific partner on the other strand. Adenine (A) pairs with thymine (T), and guanine (G) pairs with cytosine (C). Thus, *A-T* and *G-C base pairs* are said to be *complementary*. This complementary base pairing is what makes DNA a suitable molecule for carrying our genetic information—one strand of DNA can act as a *template* to direct the synthesis of a complementary strand. In this way, the information in a DNA sequence is readily copied and passed on to the next generation of cells.

Organelle DNA

Not all genetic information is found in nuclear DNA. Both plants and animals have an organelle—a "little organ" within the cell— called the *mitochondrion*. Each mitochondrion has its own set of genes. Plants also have a second organelle, the *chloroplast*, which also has its own DNA. Cells often have multiple mitochondria, particularly cells requiring lots of energy, such as active muscle cells. This is because mitochondria are responsible for converting the energy stored in macromolecules into a form usable by the cell, namely, the *adenosine triphosphate (ATP)* molecule. Thus, they are often referred to as the power generators of the cell.

Unlike *nuclear DNA* (the DNA found within the nucleus of a cell), half of which comes from our mother and half from our father, mitochondrial DNA is only inher-

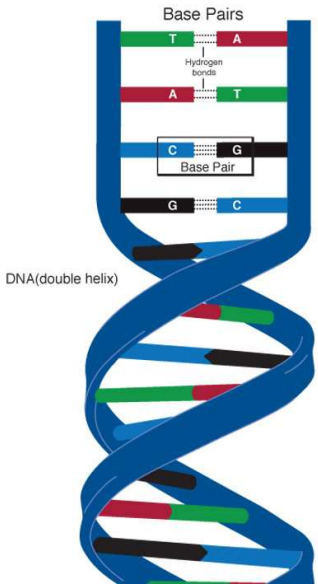


Figure 1.22: A base pair is two chemical bases bonded to one another forming a ”rung of the DNA ladder.” The DNA molecule consists of two strands that wind around each other like a twisted ladder. Each strand has a backbone made of alternating sugar (deoxyribose) and phosphate groups. Attached to each sugar is one of four bases–adenine (A), cytosine (C), guanine (G), or thymine (T). The two strands are held together by hydrogen bonds between the bases, with adenine forming a base pair with thymine, and cytosine forming a base pair with guanine.

ited from our mother. This is because mitochondria are only found in the female gametes or “eggs” of sexually reproducing animals, not in the male gamete, or sperm. Mitochondrial DNA also does not recombine; there is no shuffling of genes from one generation to the other, as there is with nuclear genes.

Large numbers of mitochondria are found in the tail of sperm, providing them with an engine that generates the energy needed for swimming toward the egg. However, when the sperm enters the egg during fertilization, the tail falls off, taking away the father’s mitochondria.

Why Is There a Separate Mitochondrial Genome?

The energy-conversion process that takes place in the mitochondria takes place *aerobically*, in the presence of oxygen. Other energy conversion processes in the cell take place *anaerobically*, or without oxygen. The independent aerobic function of these organelles is thought to have evolved from bacteria that lived inside of other simple organisms in a mutually beneficial, or *symbiotic*, relationship, providing them with aerobic capacity. Through the process of evolution, these tiny organisms became incorporated into the cell, and their genetic systems and cellular functions

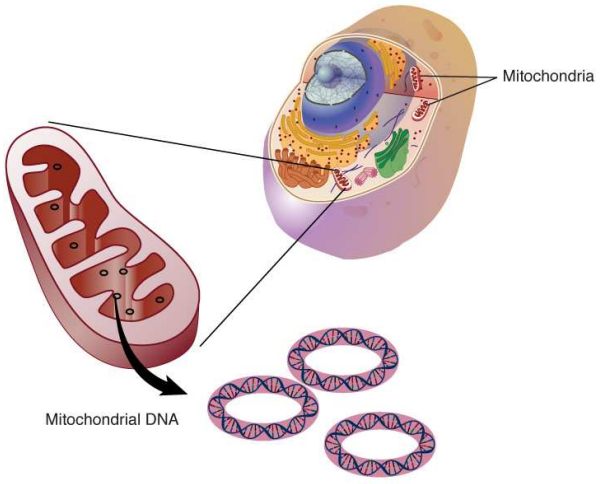


Figure 1.23: Mitochondrial DNA is the small circular chromosome found inside mitochondria. The mitochondria are organelles found in cells that are the sites of energy production. The mitochondria, and thus mitochondrial DNA, are passed from mother to offspring.

became integrated to form a single functioning cellular unit. Because mitochondria have their own DNA, RNA, and ribosomes, this scenario is quite possible. This theory is also supported by the existence of a eukaryotic organism, called the amoeba, which lacks mitochondria. Therefore, amoeba must always have a symbiotic relationship with an aerobic bacterium.

Why Study Mitochondria?

There are many diseases caused by mutations in *mitochondrial DNA (mtDNA)*. Because the mitochondria produce energy in cells, symptoms of mitochondrial diseases often involve degeneration or functional failure of tissue. For example, mtDNA mutations have been identified in some forms of diabetes, deafness, and certain inherited heart diseases. In addition, mutations in mtDNA are able to accumulate throughout an individual's lifetime. This is different from mutations in nuclear DNA, which has sophisticated repair mechanisms to limit the accumulation of mutations. Mitochondrial DNA mutations can also concentrate in the mitochondria of specific tissues. A variety of deadly diseases are attributable to a large number of accumulated mutations in mitochondria. There is even a theory, the *Mitochondrial Theory of Aging*, that suggests that accumulation of mutations in mitochondria contributes to, or drives, the aging process. These defects are associated with Parkinson's and Alzheimer's disease, although it is not known whether

the defects actually cause or are a direct result of the diseases. However, evidence suggests that the mutations contribute to the progression of both diseases.

In addition to the critical cellular energy-related functions, mitochondrial genes are useful to evolutionary biologists because of their maternal inheritance and high rate of mutation. By studying patterns of mutations, scientists are able to reconstruct patterns of migration and evolution within and between species. For example, mtDNA analysis has been used to trace the migration of people from Asia across the Bering Strait to North and South America. It has also been used to identify an ancient maternal lineage from which modern man evolved.

Ribonucleic Acids

Just like DNA, *ribonucleic acid (RNA)* is a chain, or polymer, of nucleotides with the same 5' to 3' direction of its strands. However, the ribose sugar component of RNA is slightly different chemically than that of DNA. RNA has a 2' oxygen atom that is not present in DNA. Other fundamental structural differences exist. For example, uracil takes the place of the thymine nucleotide found in DNA, and RNA is, for the most part, a single-stranded molecule. DNA directs the synthesis of a variety of RNA molecules, each with a unique role in cellular function. For example, all genes that code for proteins are first made into an RNA strand in the nucleus called a *messenger RNA (mRNA)*. The mRNA carries the information encoded in DNA out of the nucleus to the protein assembly machinery, called the *ribosome*, in the cytoplasm. The ribosome complex uses mRNA as a template to synthesize the exact protein coded for by the gene.

In addition to mRNA, DNA codes for other forms of RNA, including ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), and small nuclear RNAs (snRNAs). rRNAs and tRNAs participate in protein assembly whereas snRNAs aid in a process called splicing —the process of editing of mRNA before it can be used as a template for protein synthesis.

Proteins

Although DNA is the carrier of genetic information in a cell, proteins do the bulk of the work. Proteins are long chains containing as many as 20 different kinds of amino acids. Each cell contains thousands of different proteins: *enzymes* that make new molecules and catalyze nearly all chemical processes in cells; *structural components* that give cells their shape and help them move; hormones that transmit signals throughout the body; *antibodies* that recognize foreign molecules; and *transport molecules* that carry oxygen. The genetic code carried by DNA is what specifies the order and number of amino acids and, therefore, the shape and function of the protein.

The "*Central Dogma*"—a fundamental principle of molecular biology—states that genetic information flows from DNA to RNA to protein. Ultimately, however,

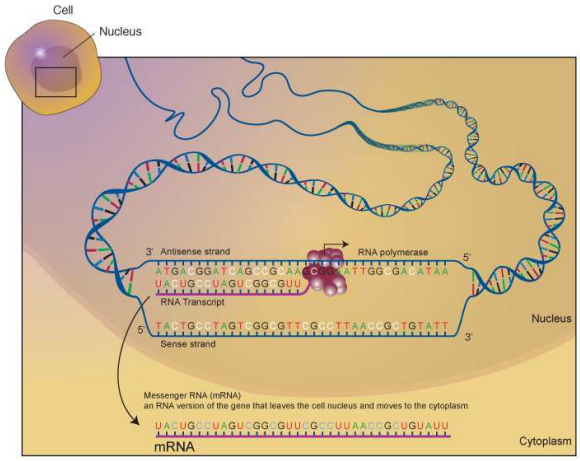
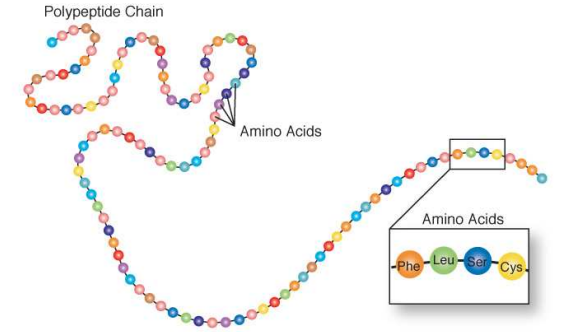


Figure 1.24: Messenger RNA (mRNA) is a single-stranded RNA molecule that is complementary to one of the DNA strands of a gene. The mRNA is an RNA version of the gene that leaves the cell nucleus and moves to the cytoplasm where proteins are made. During protein synthesis, an organelle called a ribosome moves along the mRNA, reads its base sequence, and uses the genetic code to translate each three-base triplet, or codon, into its corresponding amino acid.



Amino Acids

Ala: Alanine	Gln: Glutamine	Leu: Leucine	Ser: Serine
Arg: Arginine	Glu: Glutamic acid	Lys: Lysine	Thr: Threonine
Asn: Asparagine	Gly: Glycine	Met: Methionine	Trp: Tryptophane
Asp:Aspartic acid	His: Histidine	Phe: Phenylalanine	Tyr: Tyrosisne
Cys:Cysteine	Ile: Isoleucine	Pro: Proline	Val: Valine

Figure 1.25: Amino acids are a set of 20 different molecules used to build proteins. Proteins consist of one or more chains of amino acids called polypeptides. The sequence of the amino acid chain causes the polypeptide to fold into a shape that is biologically active. The amino acid sequences of proteins are encoded in the genes.

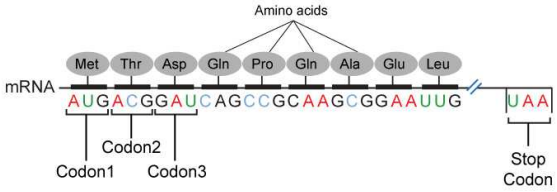


Figure 1.26: A codon is a trinucleotide sequence of DNA or RNA that corresponds to a specific amino acid. The genetic code describes the relationship between the sequence of DNA bases (A, C, G, and T) in a gene and the corresponding protein sequence that it encodes. The cell reads the sequence of the gene in groups of three bases. There are 64 different codons: 61 specify amino acids while the remaining three are used as stop signals.

the genetic code resides in DNA because only DNA is passed from generation to generation. Yet, in the process of making a protein, the encoded information must be faithfully transmitted first to RNA then to protein. Transferring the code from DNA to RNA is a fairly straightforward process called *transcription*. Deciphering the code in the resulting mRNA is a little more complex. It first requires that the mRNA leave the nucleus and associate with a large complex of specialized RNAs and proteins that, collectively, are called the *ribosome*. Here the mRNA is translated into protein by decoding the mRNA sequence in blocks of three RNA bases, called *codons*, where each codon specifies a particular amino acid. In this way, the *ribosomal complex* builds a protein one amino acid at a time, with the order of amino acids determined precisely by the order of the codons in the mRNA.

A given amino acid can have more than one codon. These redundant codons usually differ at the third position. For example, the amino acid serine is encoded by UCU, UCC, UCA, and/or UCG. This redundancy is key to accommodating mutations that occur naturally as DNA is replicated and new cells are produced. By allowing some of the random changes in DNA to have no effect on the ultimate protein sequence, a sort of genetic safety net is created. Some codons do not code for an amino acid at all but instruct the ribosome when to stop adding new amino acids.

The Core Gene Sequence: Introns and Exons

Genes make up about 1 percent of the total DNA in our genome. In the human genome, the coding portions of a gene, called *exons*, are interrupted by intervening sequences, called *introns*. In addition, a eukaryotic gene does not code for a protein in one continuous stretch of DNA. Both exons and introns are "*transcribed*" into mRNA, but before it is transported to the ribosome, the primary mRNA transcript is edited. This editing process removes the introns, joins the exons together, and adds

Table 1.1: RNA triplet codons and their corresponding amino acids.

	U	C	A	G
U	UUU Phenylalanine	UCU Serine	UAU Tyrosine	UGU Cysteine
	UUC Phenylalanine	UCC Serine	UAC Tyrosine	UGC Cysteine
	UUA Leucine	UCA Serine	UAA Stop	UGA Stop
	UUG Leucine	UCG Serine	UAG Stop	UGG Tryptophan
C	CUU Leucine	CCU Proline	CAU Histidine	CGU Arginine
	CUC Leucine	CCC Proline	CAC Histidine	CGC Arginine
	CUA Leucine	CCA Proline	CAA Glutamine	CGA Arginine
	CUG Leucine	CCG Proline	CAG Glutamine	CGG Arginine
A	AUU Isoleucine	ACU Threonine	AAU Asparagine	AGU Serine
	AUC Isoleucine	ACC Threonine	AAC Asparagine	AGC Serine
	AUA Isoleucine	ACA Threonine	AAA Lysine	AGA Arginine
	AUG Methionine	ACG Threonine	AAG Lysine	AGG Arginine
G	GUU Valine	GCU Alanine	GAU Aspartate	GGU Glycine
	GUC Valine	GCC Alanine	GAC Aspartate	GGC Glycine
	GUA Valine	GCA Alanine	GAA Glutamate	GGA Glycine
	GUG Valine	GCG Alanine	GAG Glutamate	GGG Glycine

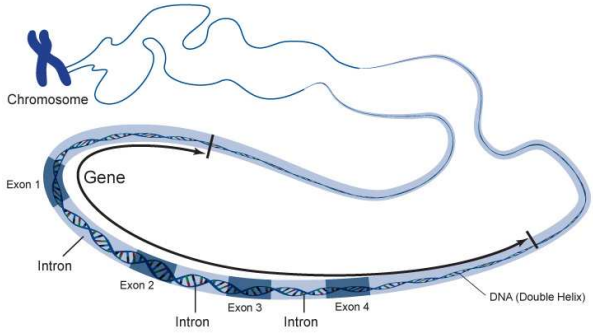


Figure 1.27: An exon is the portion of a gene that codes for amino acids. In the cells of plants and animals, most gene sequences are broken up by one or more DNA sequences called introns. The parts of the gene sequence that are expressed in the protein are called exons, because they are expressed, while the parts of the gene sequence that are not expressed in the protein are called introns, because they come in between–or interfere with–the exons. In the cells of plants and animals, most gene sequences are broken up by one or more introns.

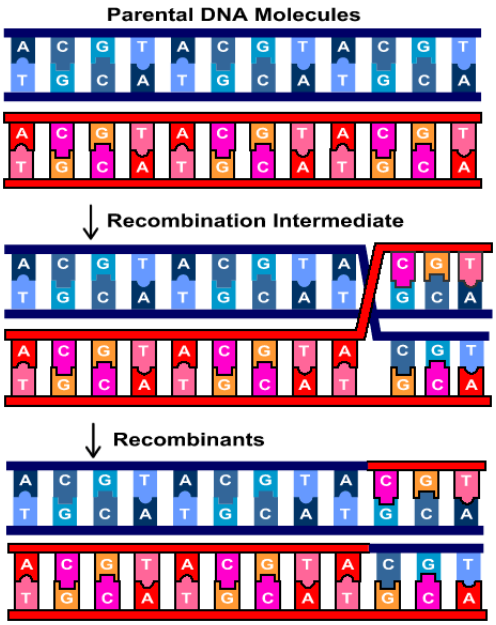


Figure 1.28: Recombination. Recombination involves pairing between complementary strands of two parental duplex DNAs (top and middle panel). This process creates a stretch of hybrid DNA (bottom panel) in which the single strand of one duplex is paired with its complement from the other duplex.

unique features to each end of the transcript to make a "*mature*" mRNA. One might then ask what the purpose of an intron is if it is spliced out after it is transcribed? It is still unclear what all the functions of introns are, but scientists believe that some serve as the site for *recombination*, the process by which progeny derive a combination of genes different from that of either parent, resulting in novel genes with new combinations of exons, the key to evolution.

Gene Prediction Using Computers

When the complete mRNA sequence for a gene is known, computer programs are used to align the mRNA sequence with the appropriate region of the genomic DNA sequence. This provides a reliable indication of the beginning and end of the coding region for that gene. In the absence of a complete mRNA sequence, the boundaries can be estimated by ever-improving, but still inexact, gene prediction software. The problem is the lack of a single sequence pattern that indicates the beginning or end of a eukaryotic gene. Fortunately, the middle of a gene, referred to as the *core gene sequence*–has enough consistent features to allow more reliable predictions.

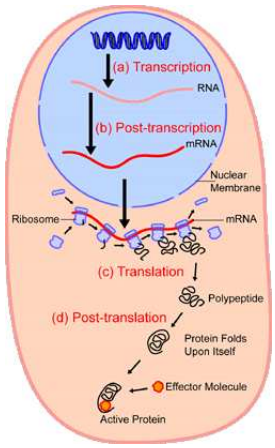


Figure 1.29: An overview of transcription and translation. This drawing provides a graphic overview of the many steps involved in transcription and translation. Within the nucleus of the cell (light blue), genes (DNA, dark blue) are transcribed into RNA. This RNA molecule is then subject to post-transcriptional modification and control, resulting in a mature mRNA molecule (red) that is then transported out of the nucleus and into the cytoplasm (peach), where it undergoes translation into a protein. mRNA molecules are translated by ribosomes (purple) that match the three-base codons of the mRNA molecule to the three-base anti-codons of the appropriate tRNA molecules. These newly synthesized proteins (black) are often further modified, such as by binding to an effector molecule (orange), to become fully active.

From Genes to Proteins: Start to Finish

We just discussed that the journey from DNA to mRNA to protein requires that a cell identify where a gene begins and ends. This must be done both during the transcription and the translation process.

Transcription

Transcription, the synthesis of an RNA copy from a sequence of DNA, is carried out by an enzyme called *RNA polymerase*. This molecule has the job of recognizing the DNA sequence where transcription is initiated, called the *promoter site*. In general, there are two "promoter" sequences upstream from the beginning of every gene. The location and base sequence of each promoter site vary for *prokaryotes* (bacteria) and *eukaryotes* (higher organisms), but they are both recognized by RNA polymerase, which can then grab hold of the sequence and drive the production of an mRNA.

Eukaryotic cells have three different RNA polymerases, each recognizing three classes of genes. *RNA polymerase II* is responsible for synthesis of mRNAs from protein-coding genes. This polymerase requires a sequence resembling TATAA, commonly referred to as the *TATA box*, which is found 25-30 nucleotides upstream

of the beginning of the gene, referred to as the *initiator sequence*.

Transcription terminates when the polymerase stumbles upon a termination, or stop signal. In eukaryotes, this process is not fully understood. Prokaryotes, however, tend to have a short region composed of G's and C's that is able to fold in on itself and form complementary base pairs, creating a stem in the new mRNA. This stem then causes the polymerase to trip and release the *nascent*, or newly formed, mRNA.

Translation

The beginning of *translation*, the process in which the genetic code carried by mRNA directs the synthesis of proteins from amino acids, differs slightly for prokaryotes and eukaryotes, although both processes always initiate at a codon for methionine. For prokaryotes, the ribosome recognizes and attaches at the sequence AGGAGGU on the mRNA, called the *Shine-Delgarno sequence*, that appears just upstream from the methionine (AUG) codon. Curiously, eukaryotes lack this recognition sequence and simply initiate translation at the amino acid methionine, usually coded for by the bases AUG, but sometimes GUG. Translation is terminated for both prokaryotes and eukaryotes when the ribosome reaches one of the three stop codons.

Structural Genes, Junk DNA, and Regulatory Sequences

Over 98 percent of the genome is of unknown function. Although often referred to as "junk" DNA, scientists are beginning to uncover the function of many of these intergenic sequences—the DNA found between genes.

Structural Genes. Sequences that code for proteins are called *structural genes*. Although it is true that proteins are the major components of structural elements in a cell, proteins are also the real workhorses of the cell. They perform such functions as transporting nutrients into the cell; synthesizing new DNA, RNA, and protein molecules; and transmitting chemical signals from outside to inside the cell, as well as throughout the cell—both critical to the process of making proteins.

Regulatory Sequences. A class of sequences called *regulatory sequences* makes up a numerically insignificant fraction of the genome but provides critical functions. For example, certain sequences indicate the beginning and end of genes, sites for initiating replication and recombination, or provide landing sites for proteins that turn genes on and off. Like structural genes, regulatory sequences are inherited; however, they are not commonly referred to as genes.

Other DNA Regions. Forty to forty-five percent of our genome is made up of short sequences that are repeated, sometimes hundreds of times. There are numerous forms of this "*repetitive DNA*", and a few have known functions, such as stabilizing the chromosome structure or inactivating one of the two X chromosomes in

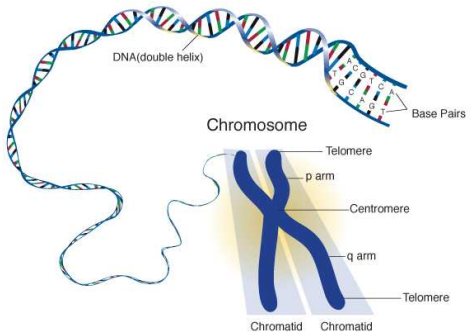


Figure 1.30: A chromosome. A chromosome is composed of a very long molecule of DNA and associated proteins that carry hereditary information. The centromere, shown at the center of this chromosome, is a specialized structure that appears during cell division and ensures the correct distribution of duplicated chromosomes to daughter cells. Telomeres are the structures that seal the end of a chromosome. Telomeres play a critical role in chromosome replication and maintenance by counteracting the tendency of the chromosome to otherwise shorten with each round of replication.

developing females, a process called *X-inactivation*. The most highly repeated sequences found so far in mammals are called "*satellite DNA*" because their unusual composition allows them to be easily separated from other DNA. These sequences are associated with chromosome structure and are found at the *centromeres* (or centers) and *telomeres* (ends) of chromosomes. Although they do not play a role in the coding of proteins, they do play a significant role in chromosome structure, duplication, and cell division. The highly variable nature of these sequences makes them an excellent "*marker*" by which individuals can be identified based on their unique pattern of their satellite DNA.

Another class of non-coding DNA is the "*pseudogene*", so named because it is believed to be a remnant of a real gene that has suffered mutations and is no longer functional. Pseudogenes may have arisen through the duplication of a functional gene, followed by inactivation of one of the copies. Comparing the presence or absence of pseudogenes is one method used by evolutionary geneticists to group species and to determine relatedness. Thus, these sequences are thought to carry a record of our evolutionary history.

How Many Genes Do Humans Have?

In February 2001, two largely independent draft versions of the human genome were published. Both studies estimated that there are 30,000 to 40,000 genes in the human genome, roughly one-third the number of previous estimates. More recently scientists estimated that there are less than 30,000 human genes. However, we still have to make guesses at the actual number of genes, because not all of the human genome sequence is annotated and not all of the known sequence has been assigned

a particular position in the genome.

So, how do scientists estimate the number of genes in a genome? For the most part, they look for tell-tale signs of genes in a DNA sequence. These include: *open reading frames*, stretches of DNA, usually greater than 100 bases, that are not interrupted by a stop codon such as TAA, TAG or TGA; *start codons* such as ATG; specific sequences found at *splice junctions*, a location in the DNA sequence where RNA removes the non-coding areas to form a continuous gene transcript for translation into a protein; and *gene regulatory sequences*. This process is dependent on computer programs that search for these patterns in various sequence databases and then make predictions about the existence of a gene.

From One Gene—One Protein to a More Global Perspective

Only a small percentage of the 3 billion bases in the human genome becomes an expressed gene product. However, of the approximately 1 percent of our genome that is expressed, 40 percent is alternatively spliced to produce multiple proteins from a single gene. *Alternative splicing* refers to the cutting and pasting of the primary mRNA transcript into various combinations of mature mRNA. Therefore the one gene—one protein theory, originally framed as "one gene—one enzyme", does not precisely hold.

With so much DNA in the genome, why restrict transcription to a tiny portion, and why make that tiny portion work overtime to produce many alternate transcripts? This process may have evolved as a way to limit the deleterious effects of mutations. Genetic mutations occur randomly, and the effect of a small number of mutations on a single gene may be minimal. However, an individual having many genes each with small changes could weaken the individual, and thus the species. On the other hand, if a single mutation affects several alternate transcripts at once, it is more likely that the effect will be devastating—the individual may not survive to contribute to the next generation. Thus, alternate transcripts from a single gene could reduce the chances that a mutated gene is transmitted.

Gene Switching: Turning Genes On and Off

The estimated number of genes for humans, less than 30,000, is not so different from the 25,300 known genes of Arabidopsis thaliana, commonly called mustard grass. Yet, we appear, at least at first glance, to be a far more complex organism. A person may wonder how this increased complexity is achieved. One answer lies in the regulatory system that turns genes on and off. This system also precisely controls the amount of a gene product that is produced and can further modify the product after it is made. This exquisite control requires multiple regulatory input points. One very efficient point occurs at transcription, such that an mRNA is produced only when a gene product is needed. Cells also regulate gene expression by *post-transcriptional modification*; by allowing only a subset of the mRNAs

to go on to translation; or by restricting translation of specific mRNAs to only when the product is needed. At other levels, cells regulate gene expression through DNA folding, chemical modification of the nucleotide bases, and intricate "*feedback mechanisms*" in which some of the gene's own protein product directs the cell to cease further protein production.

Controlling Transcription

Promoters and Regulatory Sequences. Transcription is the process whereby RNA is made from DNA. It is initiated when an enzyme, *RNA polymerase*, binds to a site on the DNA called a *promoter sequence*. In most cases, the polymerase is aided by a group of proteins called "*transcription factors*" that perform specialized functions, such as DNA sequence recognition and regulation of the polymerase's enzyme activity. Other regulatory sequences include *activators*, *repressors*, and *enhancers*. These sequences can be *cis-acting* (affecting genes that are adjacent to the sequence) or *trans-acting* (affecting expression of the gene from a distant site), even on another chromosome.

The Globin Genes: An Example of Transcriptional Regulation. An example of transcriptional control occurs in the family of genes responsible for the production of globin. Globin is the protein that complexes with the iron-containing heme molecule to make hemoglobin. *Hemoglobin* transports oxygen to our tissues via red blood cells. In the adult, red blood cells do not contain DNA for making new globin; they are ready-made with all of the hemoglobin they will need.

During the first few weeks of life, embryonic globin is expressed in the yolk sac of the egg. By week five of gestation, globin is expressed in early liver cells. By birth, red blood cells are being produced, and globin is expressed in the bone marrow. Yet, the globin found in the yolk is not produced from the same gene as is the globin found in the liver or bone marrow stem cells. In fact, at each stage of development, different globin genes are turned on and off through a process of transcriptional regulation called "*switching*".

To further complicate matters, globin is made from two different protein chains: an alpha-like chain coded for on chromosome 16; and a beta-like chain coded for on chromosome 11. Each chromosome has the embryonic, fetal, and adult form lined up on the chromosome in a sequential order for developmental expression. The developmentally regulated transcription of globin is controlled by a number of cis-acting DNA sequences, and although there remains a lot to be learned about the interaction of these sequences, one known control sequence is an enhancer called the *Locus Control Region (LCR)*. The LCR sits far upstream on the sequence and controls the alpha genes on chromosome 16. It may also interact with other factors to determine which alpha gene is turned on.

Thalassemias are a group of diseases characterized by the absence or decreased production of normal globin, and thus hemoglobin, leading to decreased oxygen in

the system. There are alpha and beta thalassemias, defined by the defective gene, and there are variations of each of these, depending on whether the embryonic, fetal, or adult forms are affected and/or expressed. Although there is no known cure for the thalassemias, there are medical treatments that have been developed based on our current understanding of both gene regulation and cell differentiation. Treatments include blood transfusions, iron chelators, and bone marrow transplants. With continuing research in the areas of gene regulation and cell differentiation, new and more effective treatments may soon be on the horizon, such as the advent of gene transfer therapies.

The Influence of DNA Structure and Binding Domains. Sequences that are important in regulating transcription do not necessarily code for transcription factors or other proteins. Transcription can also be regulated by subtle variations in DNA structure and by chemical changes in the bases to which transcription factors bind. As stated previously, the chemical properties of the four DNA bases differ slightly, providing each base with unique opportunities to chemically react with other molecules. One chemical modification of DNA, called *methylation*, involves the addition of a *methyl group (-CH3)*. Methylation frequently occurs at cytosine residues that are preceded by guanine bases, oftentimes in the vicinity of promoter sequences. The methylation status of DNA often correlates with its functional activity, where inactive genes tend to be more heavily methylated. This is because the methyl group serves to inhibit transcription by attracting a protein that binds specifically to methylated DNA, thereby interfering with polymerase binding. Methylation also plays an important role in *genomic imprinting*, which occurs when both maternal and paternal alleles are present but only one allele is expressed while the other remains inactive. Another way to think of genomic imprinting is as "*parent of origin differences*" in the expression of inherited traits. Considerable intrigue surrounds the effects of DNA methylation, and many researchers are working to unlock the mystery behind this concept.

Controlling Translation

Translation is the process whereby the genetic code carried by an mRNA directs the synthesis of proteins. *Translational regulation* occurs through the binding of specific molecules, called *repressor proteins*, to a sequence found on an RNA molecule. Repressor proteins prevent a gene from being expressed. As we have just discussed, the default state for a gene is that of being expressed via the recognition of its promoter by RNA polymerase. Close to the promoter region is another cis-acting site called the *operator*, the target for the repressor protein. When the repressor protein binds to the operator, RNA polymerase is prevented from initiating transcription, and gene expression is turned off.

Translational control plays a significant role in the process of embryonic development and cell differentiation. Upon fertilization, an egg cell begins to multiply

to produce a ball of cells that are all the same. At some point, however, these cells begin to *differentiate*, or change into specific cell types. Some will become blood cells or kidney cells, whereas others may become nerve or brain cells. When all of the cells formed are alike, the same genes are turned on. However, once differentiation begins, various genes in different cells must become active to meet the needs of that cell type. In some organisms, the egg houses store immature mRNAs that become translationally active only after fertilization. Fertilization then serves to trigger mechanisms that initiate the efficient translation of mRNA into proteins. Similar mechanisms serve to activate mRNAs at other stages of development and differentiation, such as when specific protein products are needed.

Molecular Genetics: The Study of Heredity, Genes, and DNA

As we have just learned, DNA provides a blueprint that directs all cellular activities and specifies the developmental plan of multicellular organisms. Therefore, an understanding of DNA, gene structure, and function is fundamental for an appreciation of the molecular biology of the cell. Yet, it is important to recognize that progress in any scientific field depends on the availability of experimental tools that allow researchers to make new scientific observations and conduct novel experiments. The last section of the genetic primer concludes with a discussion of some of the laboratory tools and technologies that allow researchers to study cells and their DNA.

1.3 Molecular Genetics: Piecing It Together

Molecular genetics is the study of the agents that pass information from generation to generation. These molecules, our *genes*, are long polymers of *deoxyribonucleic acid*, or DNA. Just four chemical building blocks—guanine (G), adenine (A), thymine (T), and cytosine (C)—are placed in a unique order to code for all of the genes in all living organisms.

Genes determine *hereditary traits*, such as the color of our hair or our eyes. They do this by providing instructions for how every activity in every cell of our body should be carried out. For example, a gene may tell a liver cell to remove excess cholesterol from our bloodstream. How does a gene do this? It will instruct the cell to make a particular protein. It is this protein that then carries out the actual work. In the case of excess blood cholesterol, it is the receptor proteins on the outside of a liver cell that bind to and remove cholesterol from the blood. The cholesterol molecules can then be transported into the cell, where they are further processed by other proteins.

Many diseases are caused by *mutations*, or changes in the DNA sequence of a gene. When the information coded for by a gene changes, the resulting protein may not function properly or may not even be made at all. In either case, the cells containing that genetic change may no longer perform as expected. We now know that mutations in genes code for the *cholesterol receptor protein* associated with a disease called *familial hypercholesterolemia*. The cells of an individual with this disease end up having reduced receptor function and cannot remove a sufficient amount of low density lipoprotein (LDL), or bad cholesterol, from their bloodstream. A person may then develop dangerously high levels of cholesterol, putting them at increased risk for both heart attack and stroke.

How do scientists study and find these genetic mutations? They have available to them a variety of tools and technologies to compare a DNA sequence isolated from a healthy person to the same DNA sequence extracted from an afflicted person. Advanced computer technologies, combined with the explosion of genetic data generated from the various whole genome sequencing projects, enable scientists to use these molecular genetic tools to diagnose disease and to design new drugs and therapies. Below is a review of some common laboratory methods that geneticists— scientists who study the inheritance pattern of specific traits—can use to obtain and work with DNA, followed by a discussion of some applications.

Laboratory Tools and Techniques

The methods used by molecular geneticists to obtain and study DNA have been developed through keen observation and adaptation of the chemical reactions and biological processes that occur naturally in all cells. Many of the enzymes that copy DNA, make RNA from DNA, and synthesize proteins from an RNA template were first characterized in bacteria. These basic research results have become

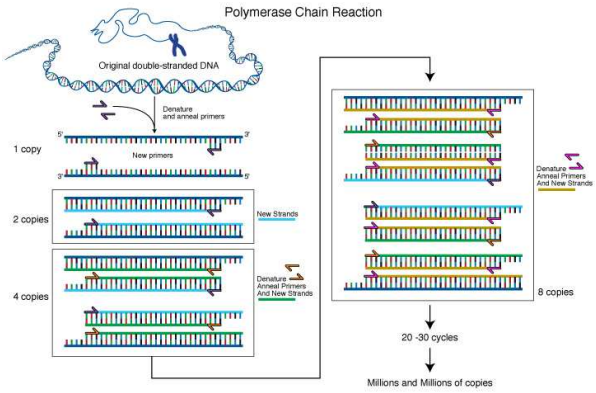


Figure 1.31: Polymerase chain reaction (PCR) is a laboratory technique used to amplify DNA sequences. The method involves using short DNA sequences called primers to select the portion of the genome to be amplified. The temperature of the sample is repeatedly raised and lowered to help a DNA replication enzyme copy the target DNA sequence. The technique can produce a billion copies of the target sequence in just a few hours.

fundamental to our understanding of the function of human cells and have led to immense practical applications for studying a gene and its corresponding protein. For example, large-scale protein production now provides an inexpensive way to generate abundant quantities of certain therapeutic agents, such as insulin for the treatment of diabetes. As science advances, so do the number of tools available that are applicable to the study of molecular genetics.

Obtaining DNA for Laboratory Analysis

Isolating DNA from just a single cell provides a complete set of all a person's genes, that is, two copies of each gene. However, many laboratory techniques require that a researcher have access to hundreds of thousands of copies of a particular gene. One way to obtain this many copies is to isolate DNA from millions of cells grown artificially in the laboratory. Another method, called *cloning*, uses DNA manipulation procedures to produce multiple copies of a single gene or segment of DNA. The *polymerase chain reaction* (PCR) is a third method whereby a specific sequence within a double-stranded DNA is copied, or *amplified*. PCR amplification has become an indispensable tool in a great variety of applications.

Methods for Amplifying DNA

Cloning DNA in Bacteria. The word "*cloning*" can be used in many ways. In this document, it refers to making multiple, exact copies of a particular sequence of

DNA. To make a clone, a target DNA sequence is inserted into what is called a *cloning vector*. A cloning vector is a DNA molecule originating from a virus, plasmid, or the cell of a higher organism into which another DNA fragment of appropriate size can be integrated without interfering with the vector's capacity for self-replication. The target and vector DNA fragments are then *ligated*, or joined together, to create what is called a *recombinant DNA molecule*. Recombinant DNA molecules are usually introduced into Escherichia coli, or E. coli—a common laboratory strain of a bacterium— by *transformation*, the natural DNA uptake mechanism possessed by bacteria. Within the bacterium, the vector directs the multiplication of the recombinant DNA molecule, producing a number of identical copies. The vector replication process is such that only one recombinant DNA molecule can propagate within a single bacterium; therefore, each resulting clone contains multiple copies of just one DNA insert. The DNA can then be isolated using the techniques described earlier.

A *restriction enzyme* is a protein that binds to a DNA molecule at a specific sequence and makes a double-stranded cut at, or near, that sequence. Restriction enzymes have specialized applications in various scientific techniques, such as manipulating DNA molecules during cloning. These enzymes can cut DNA in two different ways. Many make a simple double-stranded cut, giving a sequence what are called *blunt* or *flush ends*. Others cut the two DNA strands at different positions, usually just a few nucleotides apart, such that the resulting DNA fragments have short single-stranded overhangs, called *sticky* or *cohesive ends*. By carefully choosing the appropriate restriction enzymes, a researcher can cut out a target DNA sequence, open up a cloning vector, and join the two DNA fragments to form a recombinant DNA molecule.

More on Cloning Vectors. In general, a bacterial genome consists of a single, circular chromosome. They can also contain much smaller extrachromosomal genetic elements, called *plasmids*, that are distinct from the normal bacterial genome and are nonessential for cell survival under normal conditions. Plasmids are capable of copying themselves independently of the chromosome and can easily move from one bacterium to another. In addition, some plasmids are capable of integrating into a host genome. This makes them an excellent vehicle, or *vector*, for shuttling target DNA into a bacterial host. By cutting both the target and plasmid DNA with the same restriction enzyme, complementary base pairs are formed on each DNA fragment. These fragments may then be joined together, creating a new circular plasmid that contains the target DNA. This *recombinant plasmid* is then coaxed into a bacterial host where it is copied, or *replicated*, as though it were a normal plasmid.

Bacterial plasmids were the first vectors used to transfer genetic information and are still used extensively. However, their use is sometimes limited by the amount of target DNA they can accept, approximately 15,000 bases, or 15 Kb. With DNA sequences beyond this size, the efficiency of the vector decreases because it

now has trouble entering the cell and replicating itself. However, other vectors have been discovered or created that can accept larger target DNA including: *bacteriophages*, bacterial viruses that accept inserts up to 20 Kb; *cosmids*, recombinant plasmids with bacteriophage components that accept inserts up to 45 Kb; *bacterial artificial chromosomes* (BACs) that accept inserts up to 150 Kb; and *yeast artificial chromosomes* (YACs) that accept inserts up to 1000 kb. Many viruses have also been modified for use as cloning vectors.

Polymerase Chain Reaction (PCR). The *polymerase chain reaction (PCR)* is an amazingly simple technique that results in the exponential *amplification* of almost any region of a selected DNA molecule. It works in a way that is similar to DNA replication in nature. The primary materials, or reagents, used in PCR are:

- *DNA nucleotides*, the building blocks for the new DNA
- *Template DNA*, the DNA sequence that you want to amplify
- *Primers*, single-stranded DNAs between 20 and 50 nucleotides long that are complementary to a short region on either side of the template DNA
- *Taq polymerase*, a heat stable enzyme that drives, or catalyzes, the synthesis of new DNA

Taq polymerase was first isolated from a bacterium that lives in the hot springs in Yellowstone National Park. The Taq polymerase enzyme has evolved to withstand the extreme temperatures in which the bacteria live and can therefore remain intact during the high temperatures used in PCR.

The PCR reaction is carried out by mixing together in a small test tube the template DNA, DNA nucleotides, primers, and Taq polymerase. The primers must anneal, or pair to, the template DNA on either side of the region that is to be amplified, or copied. This means that the DNA sequences of these borders must be known so that the appropriate primers can be made. These oligonucleotides serve to initiate the synthesis of the new complementary strand of DNA. Because Taq polymerase, a form of DNA polymerase that catalyzes the synthesis of new DNA, is incredibly heat stable (thermostable), the reaction mixture can be heated to approximately 90 degrees centigrade without destroying the molecules' enzymatic activity. At this temperature, the newly created DNA strands detach from the template DNA.

The reaction mixture is then cooled again, allowing more primers to anneal to the template DNA and also to the newly created DNA. The Taq polymerase can now carry out a second cycle of DNA synthesis. This cycle of heating, cooling, and heating is repeated over and over. Because each cycle doubles the amount of template DNA in the previous cycle, one template DNA molecule rapidly becomes hundreds of thousands of molecules in just a couple of hours.

PCR has many applications in biology. It is used in DNA mapping, DNA sequencing, and molecular phylogenetics. A modified version of PCR can also be used to amplify DNA copies of specific RNA molecules. Because PCR requires very little starting material, or template DNA, it is frequently used in forensic science and clinical diagnosis.

Preparing DNA for Experimental Analysis

Gel Electrophoresis: Separating DNA Molecules of Different Lengths. Gels are usually made from *agarose*—a chain of sugar molecules extracted from seaweed—or some other synthetic molecule. Purified agarose is generally purchased in a powdered form and is dissolved in boiling water. While the solution is still hot, it is poured into a special gel casting apparatus that contains three basic parts: a tray, a support, and a comb. The tray serves as the mold that will provide the shape and size for the gel. The support prevents the liquid agarose from leaking out of the mold during the solidification process. As the liquid agarose starts to cool, it undergoes what is known as *polymerization*. Rather than staying dissolved in the water, the sugar polymers crosslink with each other, causing the solution to *gel* into a semi-solid matrix much like Jello, only more firm. The support also allows the polymerized gel to be removed from the mold without breaking. The job of the comb is to generate small *wells* into which a DNA sample will be loaded.

Once a gel has polymerized, it is lifted from the casting tray, placed into a running tank, and submerged in a special aqueous buffer, called a *running buffer*. The gel apparatus is then connected to a power supply via two plugs, or *electrodes*. Each plug leads to a thin wire at opposite ends of the tank. Because one electrode is positive and the other is negative, a strong electric current will flow through the tank when the power supply is turned on.

Next, DNA samples of interest are dissolved in a tiny volume of liquid containing a small amount of glycerol. Because glycerol has a density greater than water, it serves to weight down the sample and stops it from floating away once the sample has been loaded into a well. Also, because it is helpful to be able to monitor a DNA sample as it migrates across a gel, charged molecules, called *dyes*, are also added to the sample buffer. These dyes are usually of two different colors and two different *molecular weights*, or sizes. One of the dyes is usually smaller than most, if not all, of the sample DNA fragments and will migrate faster than the smallest DNA sample. The other dye is usually large and will migrate with the larger DNA samples. It is assumed that most of the DNA fragments of interest will migrate somewhere in between these two dyes. Therefore, when the small dye reaches the end of the gel, electrophoresis is usually stopped.

Once the gel has been prepared and loaded, the power supply is turned on. The electric current flowing through the gel causes the DNA fragments to migrate toward the bottom, or *positively charged* end, of the gel. This is because DNA has

an overall negative charge because of the combination of molecules in its structure. Smaller fragments of DNA are less impeded by the crosslinks formed within the polymerized gel than are larger molecules. This means that smaller DNA fragments tend to move faster and farther in a given amount of time. The result is a streak, or *gradient*, of larger to smaller DNA pieces. In those instances where multiple copies of DNA all have the same length, a concentration of DNA occurs at that position in the gel, called a band. Bands can result from a restriction enzyme digest of a sample containing thousands of copies of plasmid DNA, or PCR amplification of a DNA sequence. The banded DNA is then detected by soaking the gel briefly in a solution containing a dye called *ethidium bromide* (EtBr). EtBr is an *intercalating agent*, which means that it is capable of wedging itself into the grooves of DNA, where it remains. The more base pairs present within a DNA fragment, the greater the number of grooves available for EtBr to insert itself. EtBr also fluoresces under ultraviolet (UV) light. Therefore, if a gel soaked in a solution containing EtBr is placed under a UV source, a researcher can actually detect DNA by visualizing where the EtBr fluoresces. Because a scientist always loads and runs a "control" sample that contains multiple fragments of DNA with known sizes, the sizes of the sample DNA fragments can be estimated by comparing the control and sample bands.

DNA Blotting. The porous and thin nature of a gel is ideal for separating DNA fragments using electrophoresis, but as we mentioned earlier, these gels are delicate and rarely usable for other techniques. For this reason, DNA that has been separated by electrophoresis is transferred from a gel to an easy-to-handle inert membrane, a process called *blotting*. The term "blotting" describes the overlaying of the membrane on the gel and the application of a pad to ensure even contact, without disturbing the positions of the DNA fragments. In the first step, the DNA trapped in the gel is *denatured*—the double-stranded DNA is broken into single strands by soaking the gel in an alkaline solution. This readies the DNA for hybridization with a *probe*, a piece of DNA that is complementary to the sequence under investigation. A membrane, usually made of a compound called *nitrocellulose*, is then placed on top of the gel and compressed with a heavy weight. The DNA is transferred from the gel to the membrane by simple capillary action. This procedure reproduces the exact pattern of DNA captured in the gel on the membrane. The membrane can then be probed with a DNA marker to verify the presence of a target sequence.

Southern blotting is the name of the procedure for transferring denatured DNA from an agarose gel to a solid support membrane. This procedure takes advantage of a special property of nitrocellulose, its ability to bind very strongly to single-stranded DNA but not double-stranded DNA. On the other hand, *Northern blotting* refers to any blotting procedure in which electrophoresis is performed using RNA.

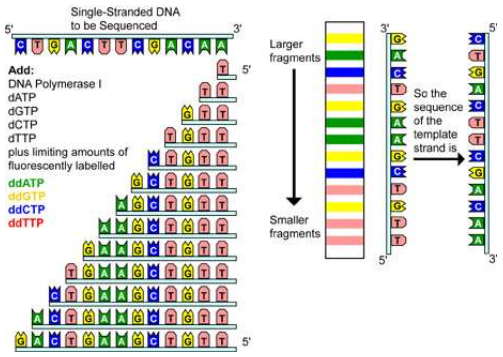


Figure 1.32: Chain termination DNA sequencing. Chain termination sequencing involves the synthesis of new strands of DNA complementary to a single-stranded template (step I). The template DNA is supplied with a mixture of all four deoxynucleotides, four dideoxynucleotides (each labeled with a different colored fluorescent tag), and DNA polymerase (step II). Because all four deoxynucleotides are present, chain elongation proceeds until, by chance, DNA polymerase inserts a dideoxynucleotide. The result is a new set of DNA chains, all of different lengths (step III). The fragments are then separated by size using gel electrophoresis (step IV). As each labeled DNA fragment passes a detector at the bottom of the gel, the color is recorded. The DNA sequence is then reconstructed from the pattern of colors representing each nucleotide sequence (step V).

Methods for Analyzing DNA

Once DNA has been isolated and purified, it can be further analyzed in a variety of ways, such as to identify the presence or absence of specific sequences or to locate nucleotide changes, called mutations, within a specific sequence.

DNA Sequencing. The process of determining the order of the nucleotide bases along a DNA strand is called *sequencing*. In 1977, 24 years after the discovery of the structure of DNA, two separate methods for sequencing DNA were developed: the *chain termination method* and the *chemical degradation method*. Both methods were equally popular to begin with, but, for many reasons, the chain termination method is the method more commonly used today. This method is based on the principle that single-stranded DNA molecules that differ in length by just a single nucleotide can be separated from one another using polyacrylamide gel electrophoresis, described earlier.

The DNA to be sequenced, called the *template DNA*, is first prepared as a single-stranded DNA. Next, a short oligonucleotide is *annealed*, or joined, to the same position on each template strand. The oligonucleotide acts as a primer for the synthesis of a new DNA strand that will be complementary to the template DNA. This technique requires that four nucleotide-specific reactions—one each for G, A, C, and T—be performed on four identical samples of DNA. The four sequencing reactions require the addition of all the components necessary to synthesize and label new DNA, including:

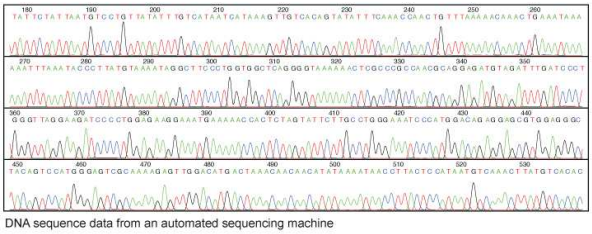


Figure 1.33: DNA sequencing is a laboratory technique used to determine the exact sequence of bases (A, C, G, and T) in a DNA molecule. The DNA base sequence carries the information a cell needs to assemble protein and RNA molecules. DNA sequence information is important to scientists investigating the functions of genes. The technology of DNA sequencing was made faster and less expensive as a part of the Human Genome Project.

- A *DNA template*
- A *primer* tagged with a mildly radioactive molecule or a light-emitting chemical
- *DNA polymerase*, an enzyme that drives the synthesis of DNA
- Four *deoxynucleotides* (G, A, C, and T)
- One *dideoxynucleotide*, either ddG, ddA, ddC, or ddT

After the first deoxynucleotide is added to the growing complementary sequence, DNA polymerase moves along the template and continues to add base after base. The strand synthesis reaction continues until a dideoxynucleotide is added, blocking further elongation. This is because dideoxynucleotides are missing a special group of molecules, called a 3'-hydroxyl group, needed to form a connection with the next nucleotide. Only a small amount of a dideoxynucleotide is added to each reaction, allowing different reactions to proceed for various lengths of time until by chance, DNA polymerase inserts a dideoxynucleotide, terminating the reaction. Therefore, the result is a set of new chains, all of different lengths.

To read the newly generated sequence, the four reactions are run side-by-side on a polyacrylamide sequencing gel. The family of molecules generated in the presence of ddATP is loaded into one lane of the gel, and the other three families, generated with ddCTP, ddGTP, and ddTTP, are loaded into three adjacent lanes. After electrophoresis, the DNA sequence can be read directly from the positions of the bands in the gel.

Variations of this method have been developed for automated sequencing machines. In one method, called *cycle sequencing*, the dideoxynucleotides, not the primers, are tagged with different colored fluorescent dyes; thus, all four reactions occur in the same tube and are separated in the same lane on the gel. As each labeled DNA fragment passes a detector at the bottom of the gel, the color is

recorded, and the sequence is reconstructed from the pattern of colors representing each nucleotide in the sequence.

Impact of Molecular Genetics

Most sequencing and analysis technologies were developed from studies of non-human genomes, notably those of the bacterium Escherichia coli, the yeast Saccharomyces cerevisiae, the fruit fly Drosophila melanogaster, the roundworm Caenorhabditis elegans, and the laboratory mouse Mus musculus. These simpler systems provide excellent models for developing and testing the procedures needed for studying the much more complex human genome.

A large amount of genetic information has already been derived from these organisms, providing valuable data for the analysis of normal human gene regulation, genetic diseases, and evolutionary processes. For example, researchers have already identified single genes associated with a number of diseases, such as cystic fibrosis. As research progresses, investigators will also uncover the mechanisms for diseases caused by several genes or by single genes interacting with environmental factors. Genetic susceptibilities have been implicated in many major disabling and fatal diseases including heart disease, stroke, diabetes, and several kinds of cancer. The identification of these genes and their proteins will pave the way to more effective therapies and preventive measures. Investigators determining the underlying biology of genome organization and gene regulation will also begin to understand how humans develop, why this process sometimes goes awry, and what changes take place as people age.

Part I

Modeling and Analysis

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.

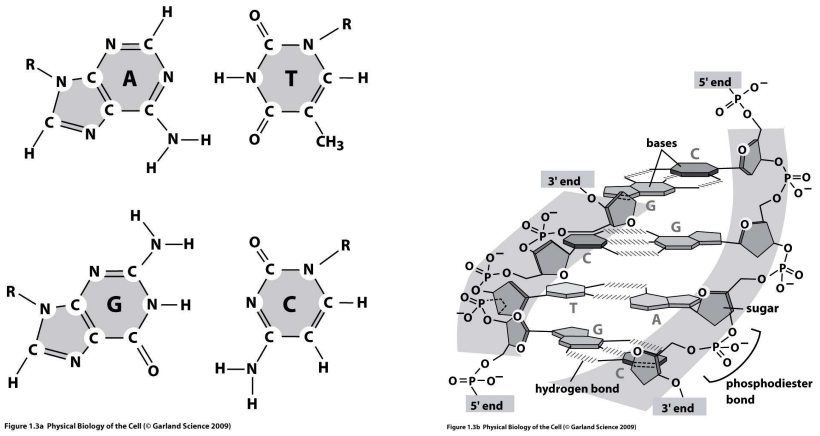
Prerequisites. Readers should have a basic understanding of ordinary differential equations, at the level of Chapter 2 of AM08, and some basic familiarity with cell biology, at the level of the description in Chapter 1.

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.* [35].

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



(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 [35]; used with permission of Garland Science.

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 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.

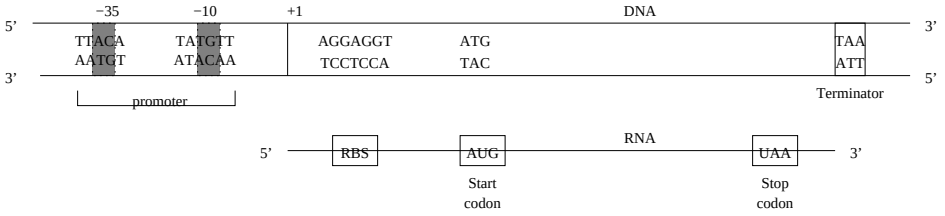


Figure 2.2: Geometric structure of DNA. The layout of the DNA is shown at the top. RNA polymerase binds to the promoter region of the DNA and transcribes the DNA starting at the +1 side and continuing to the termination site.

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 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

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 [11]; permission pending.

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 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. However,

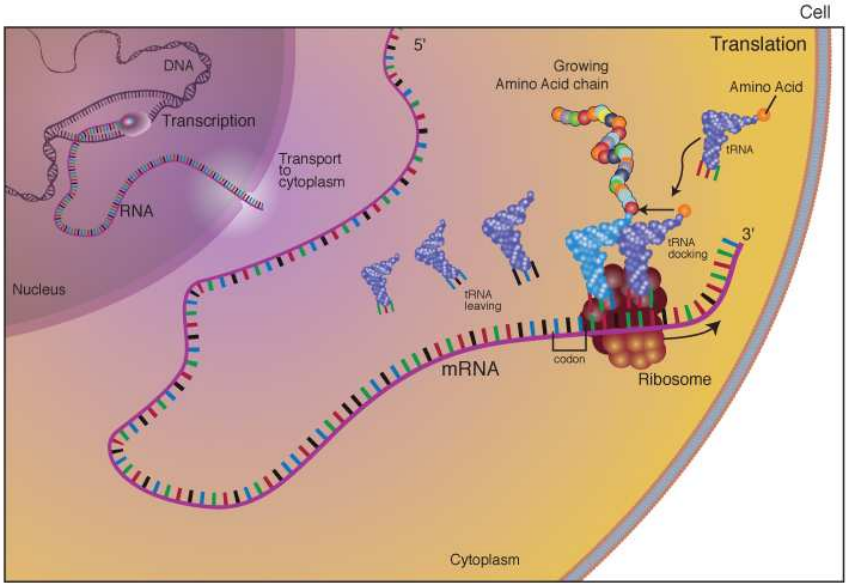


Figure 2.4: Translation is the process of translating the sequence of a messenger RNA (mRNA) molecule to a sequence of amino acids during protein synthesis. The genetic code describes the relationship between the sequence of base pairs in a gene and the corresponding amino acid sequence that it encodes. 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.

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 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

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	

bind to other species in the cell and perform the chemical reactions that underly the behavior of the organism.

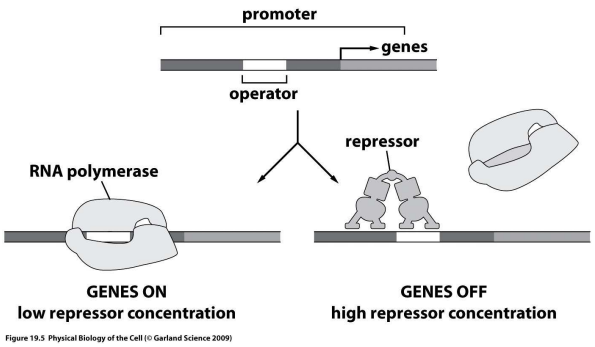
Each of the processes involved in transcription, translation and folding of the protein takes time and affects the dynamics of the cell. Table 2.1 shows the rates of some of the key processes involved in the production of proteins. It is important to note that each of these steps is highly stochastic, 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 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



(a) Repression of gene expression

./coreproc/figures/PKT08_19_06.eps

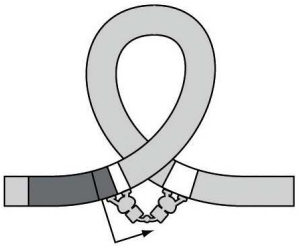
(b) Examples of repressors

Figure 2.5: Repression of gene expression. Figure from Phillips, Kondev and Theriot [35]; used with permission of Garland Science.

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 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



./coreproc/figures/PKT08_08_19.eps

(a) DNA looping (b) *lac* repressor

Figure 2.6: Repression via DNA looping. Figure from Phillips, Kondev and Theriot [35]; used with permission of Garland Science.

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

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

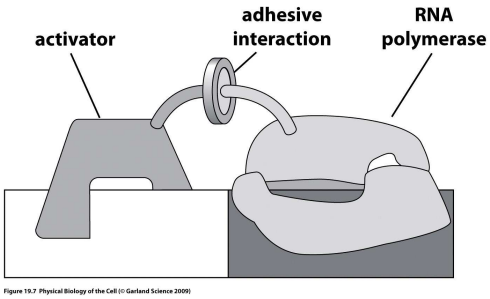
cell undergoes *heat shock* (discussed in more detail in Chapter 5). Another use is to control the timing of the expression of certain genes, as illustrated in Figure 2.9.

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, 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 is present, the RNA polymerase maintains (or regains) its contact with the DNA



(a) Activation mechanism

./coreproc/figures/PKT08_19_08.eps

(b) Examples of activators

Figure 2.8: Activation of gene expression. Figure from Phillips, Kondev and Theriot [35]; used with permission of Garland Science.

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) [**?**], 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,

./coreproc/figures/MBoC09_07_43.eps

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

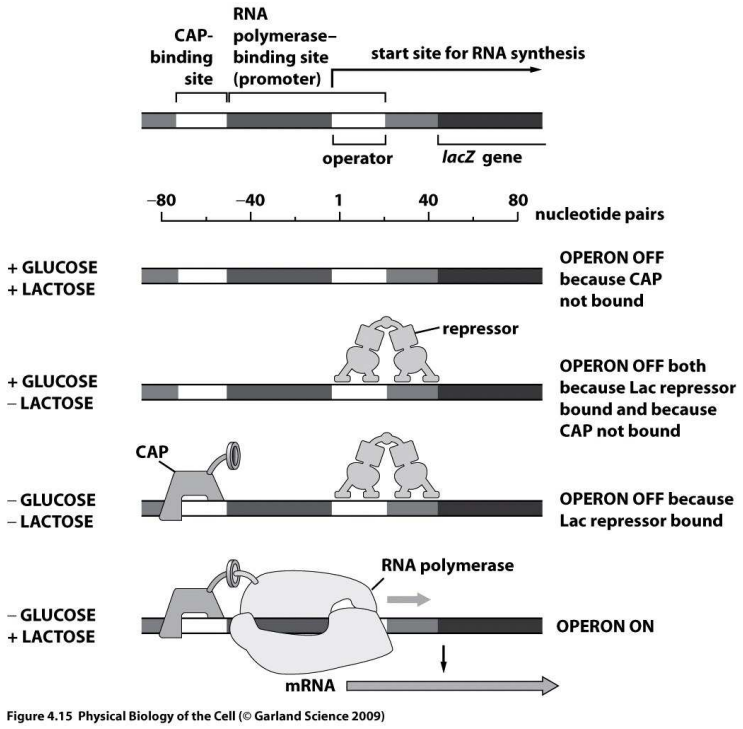


Figure 2.10: Combinatorial logic for the *lac* operator. Figure from Phillips, Kondev and Theriot [35]; used with permission of Garland Science.

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

./coreproc/figures/GNM93-antitermination.eps

Figure 2.11: Antitermination. Reproduced from [20]; permission pending.

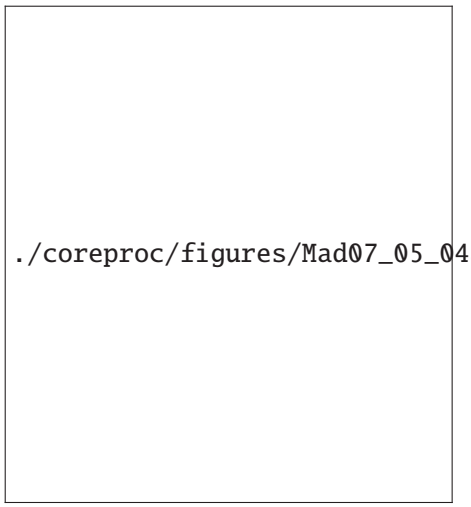


Figure 2.12: Phosphorylation of a protein via a kinase. Reproduced from Madhani [28]; permission pending.

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" 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.

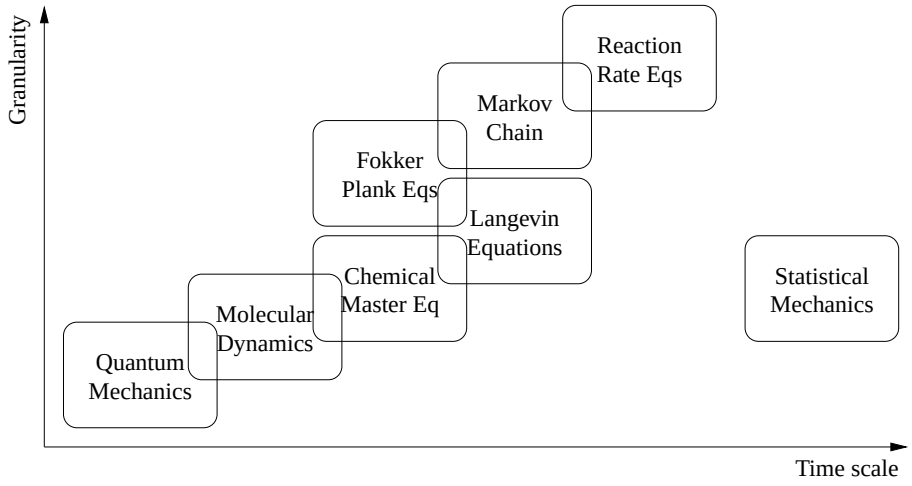


Figure 2.13: Different methods of modeling biomolecular systems.

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.

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.

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

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 [35]; used with permission of Garland Science.

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 (4.14) 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 $c_{\mathrm{A}} = n_{\mathrm{A}}/\Omega$, 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 $x_i = [\mathrm{S}_i]$ 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

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

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

Using 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

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

where A, B, C and D are appropriate species, then the dynamics for the species

concentrations can be written as

$$\begin{aligned}
\frac{d}{dt}A &= k_2C^2 \cdot D - k_1A \cdot B^2, \\
\frac{d}{dt}B &= 2k_2C^2 \cdot D - 2k_1A \cdot B^2, \\
\frac{d}{dt}C &= 2k_1A \cdot B^2 - 2k_2C^2 \cdot D, \\
\frac{d}{dt}D &= k_1A \cdot B^2 - k_2C^2 \cdot D.
\end{aligned} \tag{2.5}$$

Rearranging this equation, we can write the dynamics as

$$\frac{d}{dt}\begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ -2 & 2 \\ 2 & -2 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} k_1A \cdot B^2 \\ k_2C^2 \cdot D \end{pmatrix}. \tag{2.6}$$

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

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

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

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.

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

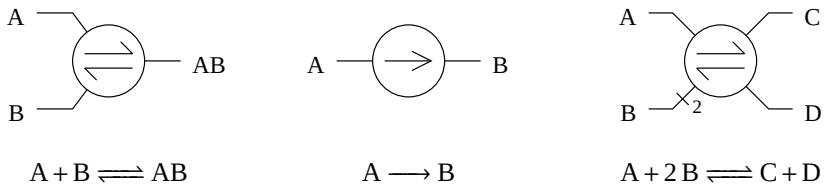


Figure 2.15: Diagrams for chemical reactions.

next chapter to illustrate that the same results can derived using a more formal and rigorous approach.

Simple binding reaction. Consider again the reaction

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

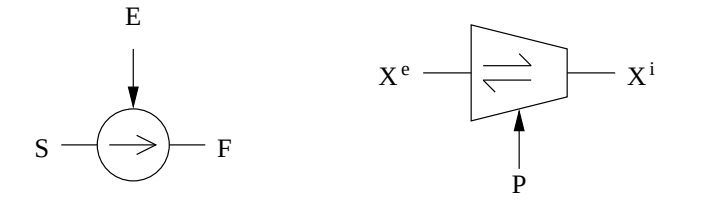
in which we now assume that the total amount of A is conserved and we denote its total concentration by A_{tot}, so that $A + C = A_{tot}$. The corresponding rate equation for C is given by

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

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.



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

Figure 2.16: Diagrams for enzymatic reactions.

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

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

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 $\bar{\mathrm{C}}$ 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

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

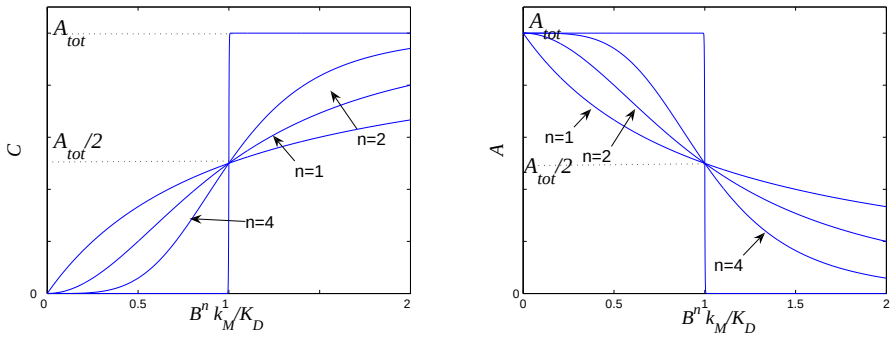


Figure 2.17: Steady state concentrations of the complex C and of A as functions of the concentration of B.

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: RNAP^{c} represents RNA polymerase in the cytoplasm and RNAP^{d} is non-specific binding of RNA polymerase to the DNA. We must similarly keep track of the state of the DNA, to insure that multiple RNA polymerases do not bind to the same section of DNA. Thus we can write DNA^{p} for the promoter region, $\text{DNA}^{g,i}$ for the ith section of a gene g (whose length can depend on the desired resolution) and DNA^{t} for the termination sequence. We write RNAP:DNA to represent RNA polymerase bound to DNA (assumed closed) and $\text{RNAP:DNA}^{\text{o}}$ to indicate the open complex. Finally, we must keep track of the mRNA that is produced by transcription: we write mRNA^{i} to represent an mRNA strand of length i and assume that the length of the gene of interest is N.

Using these various states of the RNA polymerase and locations on the DNA,

we can write a set of reactions modeling the basic elements of transcription as

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

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 $\mathrm{Ribo{:}mRNA^{RBS}}$ indicate the ribosome bound to the ribosome binding site, $\mathrm{Ribo{:}mRNA}^i$ the ribosome bound to the ith codon, $\mathrm{Ribo{:}mRNA^s}$ 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

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

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 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 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 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

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

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, μ is the

growth rate of the cell (which results in dilution of the concentration) and γ_p is the rate of degradation of the mRNA. Since the dilution and degradation terms are of the same form, we will often combine these terms in the mRNA dynamics and use a single coefficient $\bar{\gamma}_p$.

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

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

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

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; $\bar{\delta}_p$ the rate of degradation and dilution of P; and τ_p^f, the time delay associated with folding and other processes required to make the protein functional. The exponential term again accounts for dilution due to cell growth. The degradation and dilution term, parameterized by $\bar{\delta}_p$, captures both rate at which the polypeptide chain is degraded and the rate at which the concentration is diluted due to cell growth.

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

$$\begin{aligned} \frac{dm_p^*(t)}{dt} &= e^{-\mu\tau_p^m} \dot{m}_p(t - \tau_p^m) \\ &= e^{-\mu\tau_p^m} (\alpha_{p,0} - \bar{\gamma}_p m_p(t - \tau_p^m)) = \bar{\alpha}_{p,0} - \bar{\gamma}_p m_p^*(t), \end{aligned} \tag{2.11}$$

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

$$\frac{dP^*(t)}{dt} = \bar{\beta}_{p,0} m_p^*(t - \tau_p^f) - \bar{\delta} P^*(t), \tag{2.12}$$

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

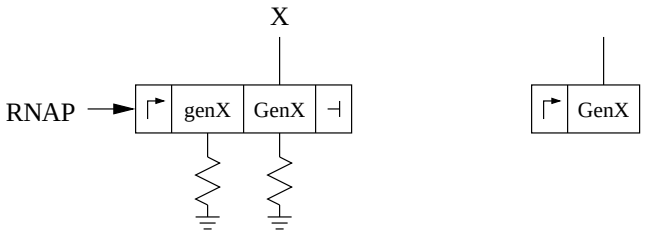


Figure 2.18: Simplified diagrams for protein production. The diagram on the left shows a section of DNA with RNA polymerase as an input, protein concentration as an output and degradation of mRNA and protein. The figure on the right is a simplified view in which only the protein output is indicated.

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

$$\frac{dm_p}{dt} = \alpha_{p,0} - \gamma_p m_p, \qquad \frac{dP}{dt} = \beta_{p,0} m_p - \delta_p P. \tag{2.13}$$

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

Finally, the simplest model for protein production is one in which we only keep track of the basal rate of production of the protein, without including the mRNA dynamics. This essentially amounts to assuming the mRNA dynamics reach steady state quickly and replacing the first differential equation in equation (2.13) 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 terminator at the ends, and then a list of the gene and protein in the middle. The boxes labeled by the gene and protein schematically represent the mRNA and protein concentration, with the line at the left of the DNA represent the input of RNA

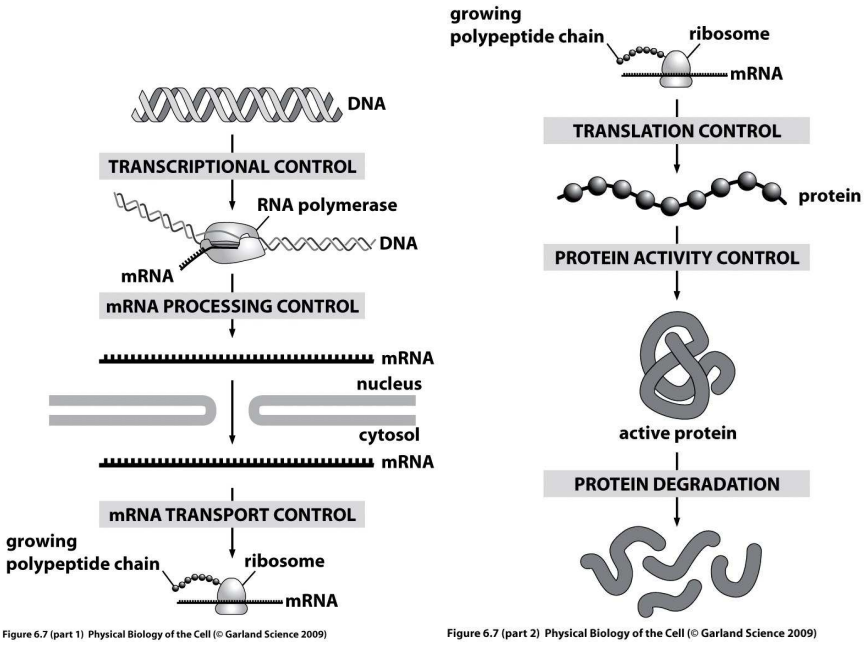


Figure 2.19: Regulation of proteins. Figure from Phillips, Kondev and Theriot [35]; used with permission of Garland Science.

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 (2.11). For a repressor (Rep), we simply have to add a reaction that represents the repressor bound to the promoter:

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

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

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

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

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

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 (4.14):

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

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

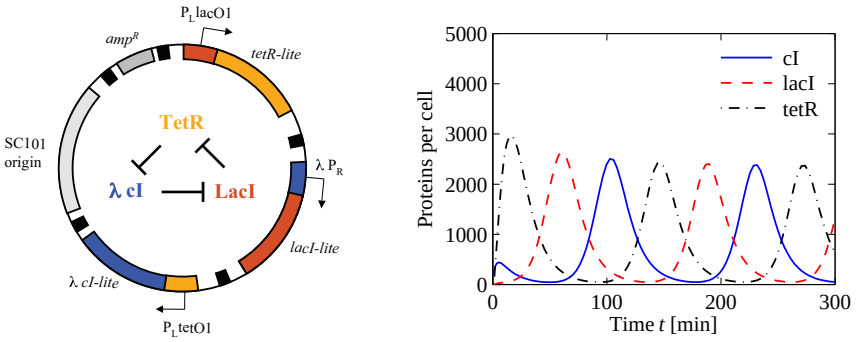
$$\frac{dm_a}{dt} = (1 - P_{\text{bound}}(p_{\text{b}}))\alpha_{a0} - \gamma_a m_a. \tag{2.14}$$

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

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

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* [12, 31]. 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 [14]. 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 (2.14), with A and B replaced by the appropriate combination of TetR, cI and LacI. The state of the

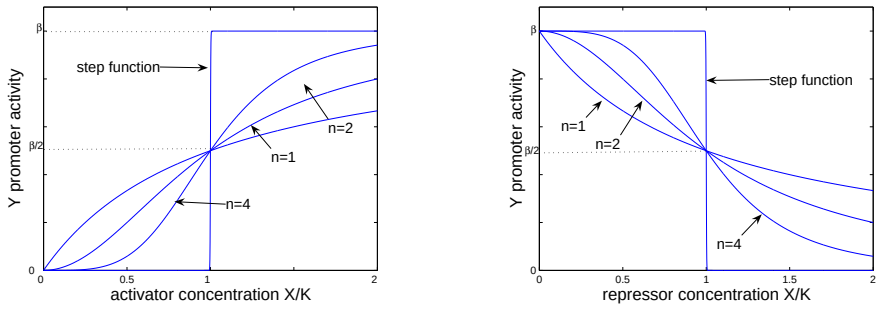


Figure 2.21: Hill function for an activator (left) and for a repressor (right).

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 $n = 2$, $\alpha = 0.5$, $k = 6.25 \times 10^{-4}$, $\alpha_0 = 5 \times 10^{-4}$, $\gamma = 5.8 \times 10^{-3}$, $\beta = 0.12$ and $\delta = 1.2 \times 10^{-3}$ with initial conditions $x(0) = (1, 0, 0, 200, 0, 0)$ (following [14]). ∇

For an activator the Hill function is given by

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

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

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

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,

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

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

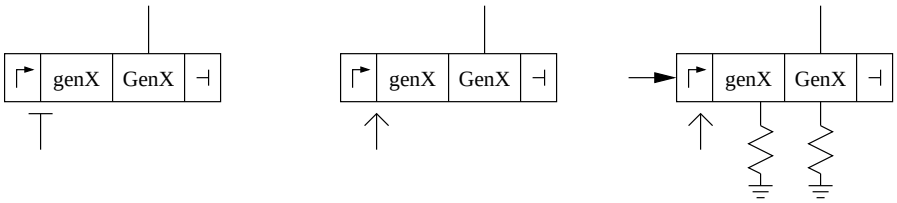


Figure 2.22: Circuit diagrams for transcriptional regulation of a gene. The first two figures represent repression and activation. If desired, additional mechanisms can also be indicated, as shown in the diagram on the right.

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

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

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

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

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

We will often find it convenient to represent the process of regulation in a graphical fashion that hides the specific details of the model that we choose to use. Figure 2.22 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 $\mathrm{p_i}$ 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

$$\frac{dm_y}{dt} = f(p_1, ..., p_n) - \gamma_{\mathrm{y}} m_y, \qquad \frac{dp_y}{dt} = \beta_{\mathrm{y}} m_y - \delta_{\mathrm{y}} p_y, \tag{2.16}$$

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 (2.11):

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

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

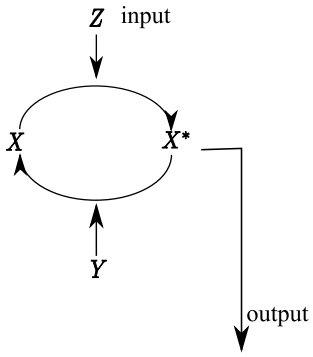


Figure 2.23: Diagram representing a covalent modification cycle.

(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 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

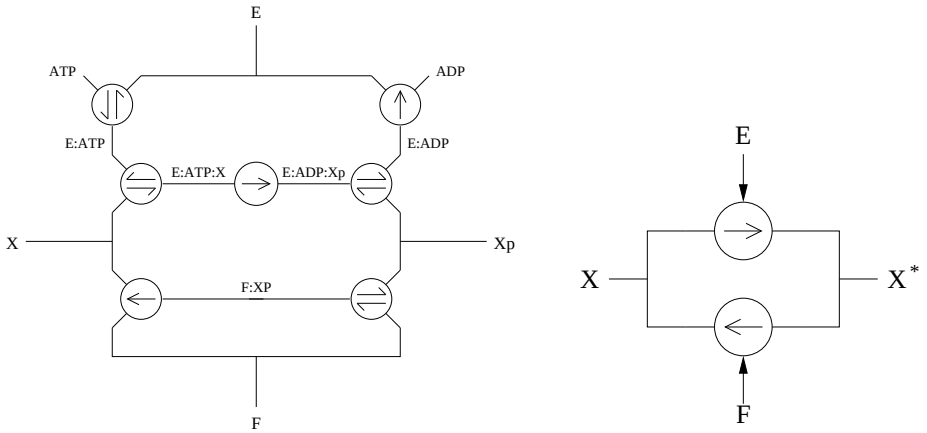
$$k'_f (Y_{tot} - C') \cdot X^* - (k'_r + k'_{cat})C' = 0.$$

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

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



(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.

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

$$\frac{dX^*}{dt} = k_{cat}\frac{Z_{tot}X}{X+K_m} - k'_{cat}\frac{Y_{tot}X^*}{X^*+K'_m}. \tag{2.17}$$

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 (2.17).

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

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

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

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

and give expressions for α and K.

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

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

the steady state values of C and A are

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

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

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

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

2.3 Consider the following modification of the competitive binding reactions:

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

and

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

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

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

2.5 Consider the phosphorylation reactions described in Section 2.5, but suppose that the kinase concentration Z is not constant, but is produced and decays according to the reaction $\mathrm{Z} \underset{k(t)}{\overset{\delta}{\rightleftharpoons}} \emptyset$. How should the system in equation (2.17) be modified? Use a MATLAB simulation to apply a periodic input stimulus $k(t)$ using parameter values: $k_{cat} = k'_{cat} = 10$, $k_f = k'_f = k_r = k'_r = 1$, $\delta = 0.01$. Is the cycle capable of "tracking" the input stimulus? If yes, to what extent? What are the tracking properties depending on?

2.6 Another model for the phosphorylation reactions, referred to as one step reaction model, is given by $\mathrm{Z} + \mathrm{X} \rightleftharpoons \mathrm{X}^* + \mathrm{Z}$ and $\mathrm{Y} + \mathrm{X}^* \rightleftharpoons \mathrm{X} + \mathrm{Y}$, in which the complex formations are neglected. Write down the ODE model and comparing the differential equation of X^* to that of equation (2.17), list the assumptions under which the one step reaction model is a good approximation of the two step reaction model.

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

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

Chapter 3
Dynamic Behavior

In this chapter, we describe some of the tools from dynamical systems and feedback control theory that will be used in the rest of the text to analyze and design biological circuits, building on tools already described in AM08. We focus here on deterministic models and the associated analyses; stochastic methods are given in Chapter 4.

Prerequisites. Readers should have a understanding of the tools for analyzing stability of solutions to ordinary differential equations, at the level of Chapter 4 of AM08. We will also make use of linearized input/output models in state space, based on the techniques described in Chapter 5 of AM08, and sensitivity function methods, described in Chapters 11 and 12 of AM08 and building on the frequency domain techniques described in Chapters 8–10.

3.1 Analysis Near Equilibria

As in the case of many other classes of dynamical systems, a great deal of insight into the behavior of a biological system can be obtained by analyzing the dynamics of the system subject to small perturbations around a known solution. We begin by considering the dynamics of the system near an equilibrium point, which is one of the simplest cases and provides a rich set of methods and tools.

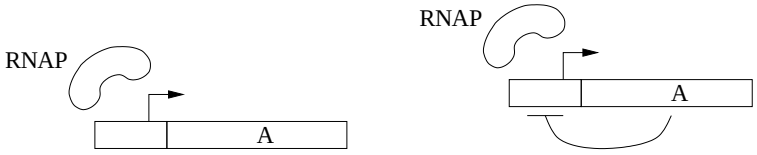
Parametric uncertainty

Consider a general nonlinear system of the from

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

where $x \in \mathbb{R}^n$ is the system state, $\theta \in \mathbb{R}^p$ are the system parameters and $w \in \mathbb{R}^q$ is a set of external inputs. Let $x_e(\theta_0, w_0)$ represent an equilibrium point for fixed parameters θ_0 and external input w_0, so that $f(x_e, \theta_0, w_0) = 0$. The stability of the system around the equilibrium point can be analyzed using the tools described in AM08. Here we focus instead on understanding how the location of the equilibrium point and the dynamics near the equilibrium point vary as a function of changes in the parameters θ and external inputs w.

We start by assuming that $w = 0$ and investigating how x_e depends on θ. The simplest approach is to analytically solve the equation $f(x_e, \theta_0) = 0$ for x_e. However,



(a) Open loop (b) Negative feedback

Figure 3.1: Parameter sensitivity in a genetic circuit. The open loop system (a) consists of a constitutive promoter, while the closed loop circuit (b) is self-regulated with negative feedback (repressor).

this is often difficult to do in closed form and so as an alternative we instead look at the linearized response given by $S_{x_e\theta} = dx_e/d\theta$, the (infinitesimal) change in the equilibrium state due to a change in the parameter. To determine $S_{x_e\theta}$ we begin by differentiating the relationship $f(x_e(\theta),\theta) = 0$ with respect to θ:

$$\frac{df}{d\theta} = \frac{\partial f}{\partial x}\frac{\partial x_e}{\partial \theta} + \frac{\partial f}{\partial \theta} = 0 \quad \Longrightarrow \quad \frac{\partial x_e}{\partial \theta} = -\left(\frac{\partial f}{\partial x}\right)^{-1} \frac{\partial f}{\partial \theta}\bigg|_{(xe,\theta_0)}. \tag{3.1}$$

These quantities can be computed numerically and hence we can evaluate the effect of small (but constant) changes in the parameters θ on the equilibrium state x_e.

A similar analysis can be performed to determine the effects of small (but constant) changes in the external input w. Suppose that x_e depends on both θ and w, with $f(x_e,\theta_0,w_0) = 0$ and θ_0 and w_0 representing the nominal values. Then

$$\frac{\partial x_e}{\partial \theta} = -\left(\frac{\partial f}{\partial x}\right)^{-1} \frac{\partial f}{\partial \theta}\bigg|_{(xe,\theta_0,w_0)}, \qquad \frac{\partial x_e}{\partial w} = -\left(\frac{\partial f}{\partial x}\right)^{-1} \frac{\partial f}{\partial w}\bigg|_{(xe,\theta_0,w_0)}.$$

We see that the vector $\partial f/\partial w$ describes how the specific inputs vary and $(\partial f/\partial x)^{-1}$ indicates how the perturbations are reflected in the equilibrium states. If the system is close to instability then some eigenvalues of $\partial f/\partial x$ will be near zero and hence the inverse could be large, resulting in significant changes in the equilibrium point due to variations in the disturbances (or parameters).

Example 3.1 (Transcriptional regulation)**.** Consider a genetic circuit consisting of a single gene. We wish to study the response of the protein concentration to fluctuations in its parameters in two cases: a *constitutive promoter* (no regulation) and self-repression (negative feedback), illustrated in Figure 3.1. The dynamics of the system are given by

$$\frac{dm}{dt} = F(P) - \gamma m, \qquad \frac{dP}{dt} = \beta m - \delta P,$$

where m is the mRNA concentration and P is the protein concentration.

For the case of no feedback we have $F(p) = \alpha_0$, and the system has an equilibrium point at $m_e = \alpha_0/\gamma$, $P_e = \beta\alpha_0/(\delta\gamma)$. The parameter vector can be taken as $\theta = (\alpha_0, \gamma, \beta, \delta)$. Since we have a simple expression for the equilibrium concentrations, we can compute the sensitivity to the parameters directly:

$$\frac{\partial x_e}{\partial \theta} = \begin{pmatrix} \frac{1}{\gamma} & -\frac{\alpha_0}{\gamma^2} & 0 & 0 \\ \frac{\beta}{\delta\gamma} & -\frac{\beta\alpha_0}{\delta\gamma^2} & \frac{\alpha_0}{\delta\gamma} & -\frac{\beta\alpha_0}{\gamma\delta^2} \end{pmatrix},$$

where the parameters are evaluated at their nominal values, but we leave off the subscript 0 on the individual parameters for simplicity. If we choose the parameters as $\theta_0 = (0.00138, 0.00578, 0.115, 0.00116)$, then the resulting sensitivity matrix evaluates to

$$S^{\text{open}}_{x_e,\theta} \approx \begin{pmatrix} 170 & -41 & 0 & 0 \\ 17000 & -4100 & 210 & -21000 \end{pmatrix}. \tag{3.2}$$

If we look instead at the scaled sensitivity matrix, then the open loop nature of the system yields a particularly simple form:

$$\bar{S}^{\text{open}}_{x_e,\theta} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & -1 & 1 & -1 \end{pmatrix}. \tag{3.3}$$

In other words, a 10% change in any of the parameters will lead to a comparable positive or negative change in the equilibrium values.

For the case of negative regulation, we have

$$F(P) = \frac{\alpha}{K + P^n} + \alpha_0,$$

and the equilibrium points satisfy

$$m_e = \frac{\delta}{\beta} P_e, \qquad \frac{\alpha}{K + P_e^n} + \alpha_0 = \gamma m_e = \frac{\gamma\delta}{\beta} P_e.$$

Rather than attempt to solve for the equilibrium point in closed form, we instead investigate the sensitivity using the computations in equation (3.1). The state, dynamics and parameters are given by

$$x = \begin{pmatrix} m & P \end{pmatrix}, \qquad f(x,\theta) = \begin{pmatrix} F(P) - \gamma m \\ \beta m - \delta P \end{pmatrix}, \qquad \theta = \begin{pmatrix} \alpha_0 & \gamma & \beta & \delta & \alpha & n & K \end{pmatrix}.$$

Note that the parameters are ordered such that the first four parameters match the open loop system. The linearizations are given by

$$\frac{\partial f}{\partial x} = \begin{pmatrix} -\gamma & F'(P_e) \\ \beta & -\delta \end{pmatrix}, \qquad \frac{\partial f}{\partial \theta} = \begin{pmatrix} 1 & -m & 0 & 0 & \frac{1}{K+P^n} & \frac{\alpha P^n \log(P)}{(K+P^n)^2} & \frac{\alpha}{(K+P^n)^2} \\ 0 & 0 & m & -P & 0 & 0 & 0 \end{pmatrix},$$

where again the parameters are taken to be their nominal values. From this we can compute the sensitivity matrix as

$$S_{x,\theta} = \begin{pmatrix} -\frac{\delta \frac{\partial \alpha}{\partial \alpha_0}}{\delta\gamma - \beta F'} & \frac{\delta m}{\delta\gamma - \beta F'} & -\frac{mF'}{\delta\gamma - \beta F'} & \frac{PF'}{\delta\gamma - \beta F'} & -\frac{\delta \frac{\partial \alpha}{\partial \alpha_1}}{\delta\gamma - \beta F'} & -\frac{\delta \frac{\partial \alpha}{\partial n}}{\delta\gamma - \beta F'} & -\frac{\delta \frac{\partial \alpha}{\partial K}}{\delta\gamma - \beta F'} \\ -\frac{\beta \frac{\partial \alpha}{\partial \alpha_0}}{\delta\gamma - \beta F'} & \frac{\beta m}{\delta\gamma - \beta F'} & -\frac{\gamma m}{\delta\gamma - \beta F'} & \frac{\gamma P}{\delta\gamma - \beta F'} & -\frac{\beta \frac{\partial \alpha}{\partial \alpha_1}}{\delta\gamma - \beta F'} & -\frac{\beta \frac{\partial \alpha}{\partial n}}{\delta\gamma - \beta F'} & -\frac{\beta \frac{\partial \alpha}{\partial K}}{\delta\gamma - \beta F'} \end{pmatrix},$$

where $F' = \partial F / \partial P$ and all other derivatives of F are evaluated at the nominal parameter values.

We can now evaluate the sensitivity at the same protein concentration as we use in the open loop case. The equilibrium point is given by

$$x_e = \begin{pmatrix} m_e \\ P_e \end{pmatrix} = \begin{pmatrix} \frac{\alpha_0}{\gamma} \\ \frac{\alpha_0 \beta}{\delta\gamma} \end{pmatrix} = \begin{pmatrix} 0.239 \\ 23.9 \end{pmatrix}$$

and the sensitivity matrix is

$$\bar{S}^{\text{closed}}_{x_e,\theta} \approx \begin{pmatrix} 76.1 & -18.2 & -1.16 & 116. & 0.134 & -0.212 & -0.000117 \\ 7610. & -1820. & 90.8 & -9080. & 13.4 & -21.2 & -0.0117 \end{pmatrix}.$$

The scaled sensitivity matrix becomes

$$\bar{S}^{\text{closed}}_{x_e,\theta} \approx \begin{pmatrix} 0.16 & -0.44 & -0.56 & 0.56 & 0.28 & -1.78 & -3.08 \times 10^{-7} \\ 0.16 & -0.44 & 0.44 & -0.44 & 0.28 & -1.78 & -3.08 \times 10^{-7} \end{pmatrix}. \quad (3.4)$$

Comparing this equation with equation (3.3), we see that there is reduction in the sensitivity with respect to most parameters. In particular, we become less sensitive to those parameters that are not part of the feedback (columns 2–4), but there is higher sensitivity with respect to some of the parameters that are part of the feedback mechanisms (particularly n). ∇

More generally, we may wish to evaluate the sensitivity of a (non-constant) solution to parameter changes. This can be done by computing the function $dx(t)/d\theta$, which describes how the state changes at each instant in time as a function of (small) changes in the parameters θ. We assume $w = 0$ for simplicity of exposition.

Let $x(t; x_0, \theta_0)$ be a solution of the dynamics with initial condition x_0 and parameters θ_0. To compute $dx/d\theta$, we write down a differential equation for how it evolves in time:

$$\begin{aligned} \frac{d}{dt}\left(\frac{dx}{d\theta}\right) = \frac{d}{d\theta}\left(\frac{dx}{dt}\right) &= \frac{d}{d\theta}(f(x,\theta,w)) \\ &= \frac{\partial f}{\partial x}\frac{dx}{d\theta} + \frac{\partial f}{\partial \theta}. \end{aligned}$$

This is a differential equation with $n \times m$ states $S_{ij} = dx_i/d\theta_j$ and with initial condition $S_{ij}(0) = 0$ (since changes to the parameters to not affect the initial conditions).

To solve these equations, we must simultaneously solve for the state x and the sensitivity S (whose dynamics depend on x). Thus, we must solve the set of $n + nm$ coupled differential equations

$$\frac{dx}{dt} = f(x,\theta,w), \qquad \frac{dS_{x\theta}}{dt} = \frac{\partial f}{\partial x}(x,\theta,w)S_{x\theta} + \frac{\partial f}{\partial \theta}(x,\theta,w). \tag{3.5}$$

This differential equation generalizes our previous results by allowing us to evaluate the sensitivity around a (non-constant) trajectory. Note that in the special case that we are at an equilibrium point and the dynamics for $S_{x,\theta}$ are stable, the steady state solution of equation (3.5) is identical to that obtained in equation (3.1). However, equation (3.5) is much more general, allowing us to determine the change in the state of the system at a fixed time T, for example. This equation also does not require that our solution stay near an equilibrium point, it only requires that our perturbations in the parameters are sufficiently small.

Example 3.2 (Repressilator)**.** Consider the example of the repressilator, which was described in Example 2.1. The dynamics of this system can be written as

$$\begin{aligned}
\frac{dm_1}{dt} &= F_{\text{rep}}(P_3) - \gamma m_1 & \frac{dP_1}{dt} &= \beta m_1 - \delta P_1 \\
\frac{dm_2}{dt} &= F_{\text{rep}}(P_1) - \gamma m_2 & \frac{dP_2}{dt} &= \beta m_2 - \delta P_2 \\
\frac{dm_3}{dt} &= F_{\text{rep}}(P_2) - \gamma m_2 & \frac{dP_3}{dt} &= \beta m_3 - \delta P_2,
\end{aligned}$$

where the repressor is modeled using a Hill function

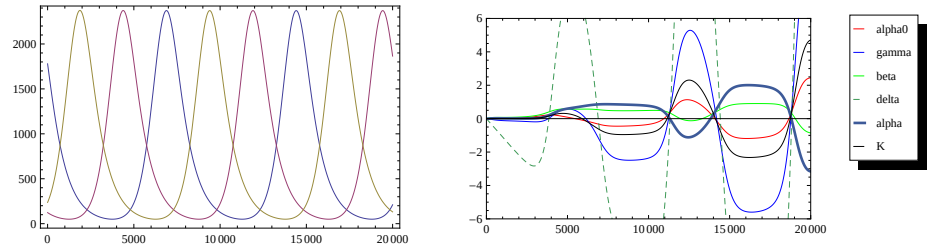
$$F_{\text{rep}}(p) = \frac{\alpha}{K + p^n} + \alpha_0.$$

The dynamics of this system lead to a limit cycle in the protein concentrations, as shown in Figure 3.2a.

We can analyze the sensitivity of the protein concentrations to changes in the parameters using the sensitivity differential equation. Since our solution is periodic, the sensitivity dynamics will satisfy an equation of the form

$$\frac{dS_{x,\theta}}{dt} = A(t)S_{x,\theta} + B(t),$$

where $A(t)$ and $B(t)$ are both periodic in time. Letting $x = (m_1, P_1, m_2, P_2, m_3, P_3)$ and $\theta = (\alpha_0, \gamma, \beta, \delta, \alpha, K)$, we can compute $S_{x,\theta}$ along the limit cycle. If the dynamics for $S_{x,\theta}$ are stable then the resulting solutions will be periodic, showing how the dynamics around the limit cycle depend on the parameter values. The results are shown in Figure 3.2b, where we plot the steady state sensitivity of P_1 as a function of time. We see, for example, that the limit cycle depends strongly on the protein



(a) (b)

Figure 3.2: Repressilator sensitivity plots

degradation and dilution rate γ, indicating that changes in this value can lead to (relatively) large variations in the magnitude of the limit cycle.

∇

Several simulation tools include the ability to do sensitivity analysis of this sort, including `COPASI`.

Frequency domain analysis

Another way to look at the sensitivity of the solutions near equilibria to changes in parameters and inputs is to use frequency domain techniques. Recall that the *frequency response* of a linear system

$$\dot{x} = Ax + Bu$$
$$y = Cx + Du$$

is the response of the system to a sinusoidal input $u = a\sin\omega t$ with input amplitude a and frequency ω. The transfer function for a linear system is given by

$$G_{yu}(s) = C(sI - A)^{-1}B + D$$

and represents the response of a system to an exponential signal of the form $u(t) = e^{st}$ where $s \in \mathbb{C}$. In particular, the response to a sinusoid $u = a\sin\omega t$ is given by $y = Ma\sin(\omega t + \theta)$ where the gain M and phase shift θ can be determined from the transfer function evaluated at $s = i\omega$:

$$G_{yu}(i\omega) = Me^{i\theta}.$$

For finite dimensional linear (or linearized) systems, the transfer function be be written as a ratio of polynomials in s:

$$G(s) = \frac{b(s)}{a(s)}.$$

The values of s at which the numerator vanishes are called the zeros of the transfer function and the values of s at which the denominator vanishes are called the poles.

The transfer function representation of an input/output linear system is essentially equivalent to the state space description, but we reason about the dynamics by looking at the transfer function instead of the state space matrices. For example, it can be shown that the poles of a transfer function correspond to the eigenvalues of the matrix A, and hence the poles determine the stability of the system.

Interconnections between subsystems often have simple representations in terms of transfer functions. Two systems G_1 and G_2 in series (with the output of the first connected to the input of the second) have a combined transfer function $G_{\text{series}}(s) = G_1(s)G_2(s)$ and two systems in parallel (a single input goes to both systems and the outputs are summed) has the transfer function $G_{\text{parallel}}(s) = G_1(s) + G_2(s)$. A common interconnection is two put two systems in feedback form for which the transfer function is given by

$$G_{yr}(s) = \frac{G_1(s)}{G_1(s) + G_2(s)} = \frac{n_1(s)d_2(s)}{n_1(s)d_2(s) + d_1(s)n_2(s)},$$

where $n_i(s)$ and $d_i(s)$ are the numerator and denominator of the individual transfer function. The ease in which the input/output response for interconnected systems can be computed with transfer functions is one of the main motivations for their widespread use in engineering.

Transfer functions are useful representations of linear systems because the properties of the transfer function can be related to the properties of the dynamics. In particular, the shape of the frequency response describes how the system response to inputs and disturbances, as well as allows us to reason about the stability of interconnected systems. The Bode plot of a transfer function gives the magnitude and phase of the frequency response as a function of frequency and the Nyquist plot can be used to reason about stability of a closed loop system from the open loop frequency response. The transfer function for a system can be determined from experiments by measuring the frequency response and fitting a transfer function to the data. Formally, the transfer function corresponds to the ratio of the Laplace transforms of the output to the input.

Returning to our analysis of biomolecular systems, suppose we have a systems whose dynamics can be written as

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

and we wish to understand how the solutions of the system depend on the parameters θ and disturbances w. We focus on the case of an equilibrium solution $x(t; x_0, \theta_0) = x_e$. Let $z = x - x_e$, $\tilde{w} = w - w_0$ and $\tilde{\theta} = \theta - \theta_0$ represent the deviation of the state, input and parameters from their nominal values. We can write the dynamics of the perturbed system using its linearization:

$$\frac{dz}{dt} = \left(\frac{\partial f}{\partial x}\right)_{(x_e,\theta_0,w_0)} \cdot z \quad + \quad \left(\frac{\partial f}{\partial \theta}\right)_{(x_e,\theta_0,w_0)} \cdot \tilde{\theta} \quad + \quad \left(\frac{\partial f}{\partial w}\right)_{(x_e,\theta_0,w_0)} \cdot \tilde{w}.$$

This linear system describes small deviations from $x_e(\theta_0, w_0)$ but allows $\tilde{\theta}$ and $\tilde{w}$ to be time-varying instead of the constant case considered earlier.

To analyze the resulting deviations, it is convenient to look at the system in the frequency domain. Let $y = Cx$ be a set of values of interest. The transfer functions between $\tilde{\theta}$, $\tilde{w}$ and y are given by

$$H_{y\tilde{\theta}}(s) = C(sI - A)^{-1}B_\theta, \qquad H_{y\tilde{w}}(s) = C(sI - A)^{-1}B_w,$$

where

$$A = \left.\frac{\partial f}{\partial x}\right|_{(x_e,\theta_0,w_0)}, \qquad B_\theta = \left.\frac{\partial f}{\partial \theta}\right|_{(x_e,\theta_0,w_0)}, \qquad B_w = \left.\frac{\partial f}{\partial w}\right|_{(x_e,\theta_0,w_0)}.$$

Note that if we let $s = 0$, we get the response to small, constant changes in parameters. For example, the change in the outputs y as a function of constant changes in the parameters is given by

$$H_{y\tilde{\theta}}(0) = CA^{-1}B_\theta = CS_{x,\theta},$$

which matches our previous parametric analysis.

Example 3.3 (Transcriptional regulation)**.** Consider again the case of transcriptional regulation described in Example 3.1. Suppose that the mRNA degradation rate γ can change as a function of time and that we wish to understand the sensitivity with respect to this (time-varying) parameter. Linearizing the dynamics around an equilibrium point

$$A = \begin{pmatrix} -\gamma & F'(p_e) \\ \beta & -\delta \end{pmatrix}, \qquad B_\gamma = \begin{pmatrix} -m_e \\ 0 \end{pmatrix}.$$

For the case of no feedback we have $F(P) = \alpha_0$, and the system has an equilibrium point at $m_e = \alpha_0/\gamma$, $P_e = \beta\alpha_0/(\delta\gamma)$. The transfer function from γ to p is given by

$$G^{\text{ol}}_{P\gamma}(s) = \frac{-\beta m_e}{(s+\gamma)(s+\delta)}.$$

For the case of negative regulation, we have

$$F(P) = \frac{\alpha}{K + P^n} + \alpha_0,$$

and the resulting transfer function is given by

$$G^{\text{cl}}_{P\gamma}(s) = \frac{\beta m_e}{(s+\gamma)(s+\delta)+\beta\sigma}, \qquad \sigma = F'(P_e) = \frac{n\alpha P_e^{n-1}}{(K+P_e^n)^2}.$$

Figure 3.3 shows the frequency response for the two circuits. We see that the feedback circuit attenuates the response of the system to disturbances with low-frequency content but slightly amplifies disturbances at high frequency (compared to the open loop system). ∇

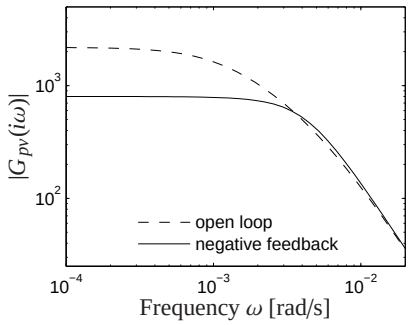


Figure 3.3: Noise attenuation in a genetic circuit.

Robustness analysis

A slightly more general analysis of sensitivity can be accomplished using the control theoretic notions of sensitivity described in AM08, Chapter 12. Rather than just considering static changes to parameter values, we can instead consider the case of *unmodeled dynamics*, in which we allow bounded input/output uncertainties to enter the system dynamics. This can be used to model parameters whose values are unknown and also time-varying, as well as capturing uncertain dynamics that are being ignored or approximated.

To illustrate the basic approach, consider the problem of determining the sensitivity of a set of reactions to a set of additional unmodeled reactions, whose detailed effects are unknown but assumed to be bounded. We set this problem up using the general framework shown in Figure 3.4.

3.2 Analysis of Reaction Rate Equations

The previous section considered analysis techniques for general dynamical systems with small perturbations. In this section, we specialize to the case where the

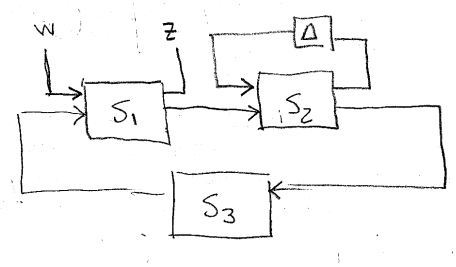


Figure 3.4: Analysis of dynamic uncertainty in a reaction system.

dynamics have the form of a reaction rate equation:

$$\dot{s} = Nv(s, p), \tag{3.6}$$

where s is the vector of species concentrations, p is the vector of reaction parameters, N is the stoichiometry matrix, $v(s, p)$ is the reaction rate (or flux) vector.

Reduced reaction dynamics

When analyzing reaction rate equations, it is often the case that there are conserved quantities in the dynamics. For example, conservation of mass will imply that if all compounds containing a given species are captured by the model, the total mass of that species will be constant. This type of constraint will then give a conserved quantity of the form $c_i = H_i s$ where H_i represents that combinations of species in which the given element appears. Since c_i is constant, it follows that $\dot{c}_i = 0$ and, aggregating the set of all conserved species, we have

$$0 = \dot{c} = H\dot{x} = HNv(s, p) \quad \text{for all } s.$$

If we assume that the vector of fluxes spans $\mathbb{R}^m$ (the range of $v : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^m$), then this implies that the conserved quantities correspond to the left null space of the stoichiometry matrix N.

It is often useful to remove the conserved quantities from the description of the dynamics and write the dynamics for a set of independent species. To do this, we transform the state of the system into two sets of variables:

$$\begin{pmatrix} s_i \\ s_d \end{pmatrix} = \begin{pmatrix} P \\ H \end{pmatrix} s. \tag{3.7}$$

The vector $s_i = Ps$ is the set of independent species and is typically chosen as a subset of the original species of the model (so that the rows P consists of all zeros and a single 1 in the column corresponding to the selected species). The matrix H should span the left null space of N, so that s_d represents the set of dependent concentrations. These dependent species do not necessarily correspond to individual species, but instead are often combinations of species (for example, the total concentration of a given element that appears in a number of molecules that participate in the reaction).

Given the decomposition (3.7), we can rewrite the dynamics of the system in terms of the independent variables s_i. We start by noting that given s_i and s_d, we can reconstruct the full set of species s:

$$s = \begin{pmatrix} P \\ H \end{pmatrix}^{-1} \begin{pmatrix} s_i \\ s_d \end{pmatrix} = Ls_i + c_0, \qquad L = \begin{pmatrix} P \\ H \end{pmatrix}^{-1} \begin{pmatrix} I \\ 0 \end{pmatrix}, \quad c_0 = \begin{pmatrix} P \\ H \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ c \end{pmatrix}$$

where c_0 represents the conserved quantities. We now write the dynamics for s_i as

$$\dot{s}_i = P\dot{s} = PNv(Ls_i + c_0, p) = N_r v_r(s_i, c_0, p), \tag{3.8}$$

where N_r is the *reduced stoichiometry matrix* and v_r is the rate vector with the conserved quantities separated out as constant parameters.

The reduced order dynamics in equation (3.8) represent the evolution of the independent species in the reaction. Given s_i, we can "lift" the dynamics from the independent species to the full set of species by writing $s = Ls_i + c_0$. The vector c_0 represents the values of the conserved quantities, which must be specified in order to compute the values of the full set of species. In addition, since $s = Ls_i + c_0$, we have that

$$\dot{s} = L\dot{s}_i = LN_r v_r(s_i, c_0, p) = LN_r v(s, p),$$

which implies that

$$N = LN_r.$$

Thus, L also "lifts" the reduced stoichiometry matrix from the reduced space to the full space.

Example 3.4 (Enzyme kinetics)**.** Consider an enzymatic reaction

$$\mathrm{S} + \mathrm{E} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_{\mathrm{cat}}} \mathrm{E} + \mathrm{P},$$

whose full dynamics can be written as

$$\frac{d}{dt}\begin{pmatrix} S \\ E \\ ES \\ P \end{pmatrix} = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 1 & 0 \\ 1 & -1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} k_{\mathrm{on}} E \cdot S \\ k_{\mathrm{off}} ES \\ k_{\mathrm{cat}} ES \end{pmatrix}.$$

The conserved quantities are given by

$$H = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & -1 & 0 & 1 \end{pmatrix}.$$

The first of these is the total enzyme concentration $E_T = E + ES$, while the second asserts that the concentration of product P is equal to the free enzyme concentration E minus the substrate concentration S. If we assume that we start with substrate concentration S_0, enzyme concentration E_T and no product or bound enzyme, then the conserved quantities are given by

$$c = \begin{pmatrix} E + ES \\ S - E + P \end{pmatrix} = \begin{pmatrix} E_T \\ S_0 - E_T \end{pmatrix}.$$

There are many possible choices for the set of independent species $s_i = Ps$, but since we are interested in the substrate and the product, we choose P as

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Once P is chosen then we can compute

$$L = \begin{pmatrix} P \\ H \end{pmatrix}^{-1} \begin{pmatrix} I \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ -1 & -1 \\ 0 & 1 \end{pmatrix}, \qquad c_0 = \begin{pmatrix} P \\ H \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ E_T - S_0 \\ S_0 \\ 0 \end{pmatrix},$$

The resulting reduced order dynamics can be computed to be

$$\begin{aligned} \frac{d}{dt}\begin{pmatrix} S \\ P \end{pmatrix} &= \begin{pmatrix} -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} k_{\text{on}}(P + S + E_T - S_0)S \\ k_{\text{off}}(-P - S + S_0) \\ k_{\text{cat}}(-P - S + S_0) \end{pmatrix} \\ &= \begin{pmatrix} -k_{\text{on}}(P + S + E_T - S_0)S - k_{\text{off}}(P + S - S_0) \\ k_{\text{cat}}(S_0 - S - P) \end{pmatrix}. \end{aligned}$$

A simulation of the dynamics is shown in Figure 3.5. We see that the dynamics are very well approximated as being a constant rate of production until we exhaust the substrate (consistent with the Michaelis-Menten approximation).

∇

Metabolic control analysis

Metabolic control analysis (MCA) focuses on the study of the sensitivity of steady state concentrations and fluxes to changes in various system parameters. The basic concepts are equivalent to the sensitivity analysis tools described in Section 3.1, specialized to the case of reaction rate equations. In this section we provide a brief introduction to the key ideas, emphasizing the mapping between the general concepts and MCA terminology (as originally done by Ingalls [24]).

Consider the reduced set of chemical reactions

$$\dot{s}_i = N_r v_r(s_i, p) = N_r v(Ls_i + c_0, p).$$

We wish to compute the sensitivity of the equilibrium concentrations s_e and equilibrium fluxes v_e to the parameters p. We start by linearizing the dynamics around an equilibrium point s_e. Defining $x = s - s_e$, $u = p - p_0$ and $f(x,u) = N_r v(s_e + x, p_0 + u)$, we can write the linearized dynamics as

$$\dot{x} = Ax + Bu, \qquad A = \left(N_r \frac{\partial v}{\partial s} L\right), \quad B = \left(N_r \frac{\partial v}{\partial p}\right), \tag{3.9}$$

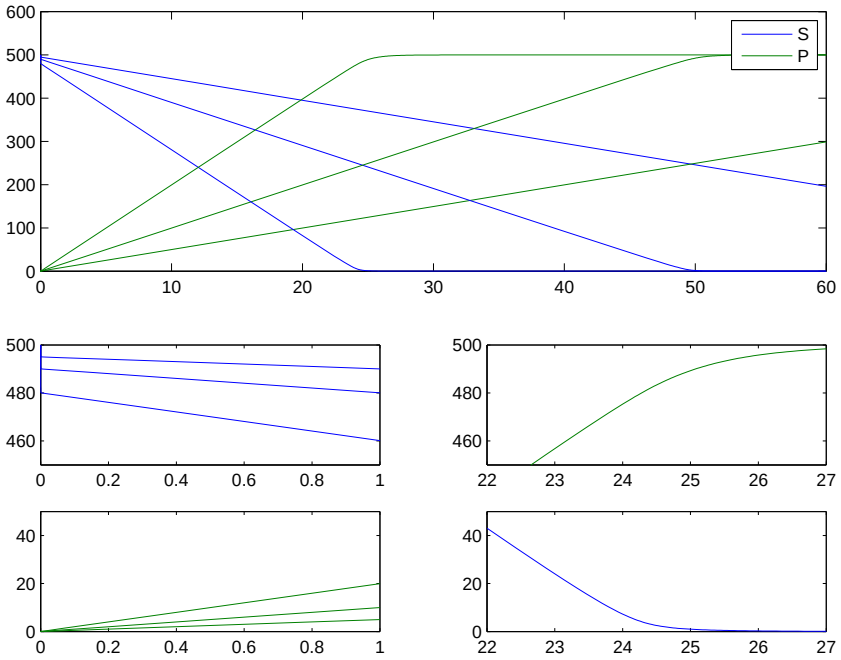


Figure 3.5: Enzyme dynamics. The simulations were carried out $k_{\text{on}} = k_{\text{off}} = 10$, $k_{\text{cat}} = 1$, $S_0 = 500$ and $E_T = 5, 1020$. The top plot shows the concentration of substrate S and product P, with the fastest case corresponding to $E_T = 20$. The figures on the lower left zoom in on the substrate and product concentrations at the initial time and the figures on the lower right at one of the transition times.

which has the form of a linear differential equation with state x and input u.

In metabolic control analysis, the following terms are defined:

$$\bar{\epsilon}_p = \left.\frac{dv}{dp}\right|_{s_e, p_o} \qquad \bar{\epsilon}_p = \text{flux control coefficients}$$

$$\bar{R}^s_p = \qquad \bar{C}^s = \text{concentration control coefficients}$$

$$\bar{R}^s_p = \frac{\partial s_e}{\partial p} = \bar{C}^s \bar{\epsilon}_p$$

$$\bar{R}^v_p = \qquad \bar{C}^v = \text{rate control coefficients}$$

$$\bar{R}^v_p = \frac{\partial v_e}{\partial p} = \bar{C}^v \bar{\epsilon}_p$$

These relationships describe how the equilibrium concentration and equilibrium rates change as a function of the perturbations in the parameters. The two control matrices provide a mapping between the variation in the flux vector evaluated at equilibrium,

$$\left(\frac{\partial v}{\partial p}\right)_{s_e, p_0},$$

and the corresponding differential changes in the equilibrium point, $\partial s_e / \partial p$ and

$\partial v_e/\partial p$. Note that

$$\frac{\partial v_e}{\partial p} \neq \left(\frac{\partial v}{\partial p}\right)_{s_e,p_0}.$$

The left side is the relative change in the equilibrium rates, while the right side is the change in the rate function $v(s,p)$ evaluated at an equilibrium point.

To derive the coefficient matrices $\bar{C}^s$ and $\bar{C}^v$, we simply take the linear equation (3.9) and choose outputs corresponding to s and v:

$$y_s = Ix, \qquad y_v = \frac{\partial v}{\partial s}Lx + \frac{\partial v}{\partial p}u.$$

Using these relationships, we can compute the transfer functions

$$H_s(s) = (sI - A)^{-1}B = \Big[(sI - N_r\frac{\partial v}{\partial s}L)^{-1}N_r\Big]\frac{\partial v}{\partial p},$$
$$H_v(s) = \frac{\partial v}{\partial s}L(sI - A)^{-1}B + \frac{\partial v}{\partial p} = \Big[\frac{\partial v}{\partial s}L(sI - N_r\frac{\partial v}{\partial s}L)^{-1}N_r + I\Big]\frac{\partial v}{\partial p}.$$

Classical metabolic control analysis considers only the equilibrium concentrations, and so these transfer functions would be evaluated at $s = 0$ to obtain the equilibrium equations.

These equations are often normalized by the equilibrium concentrations and parameter values, so that all quantities are expressed as fractional quantities. If we define

$$D^s = \text{diag}\{s_e\}, \qquad D^v = \text{diag}\{v(s_e,p_0)\}, \qquad D^p = \text{diag}\{p_0\},$$

the the normalized coefficient matrices (without the overbar) are given by

$$C^s = (D^s)^{-1}\bar{C}^s D^v, \qquad C^v = (D^v)^{-1}\bar{C}^v D^v,$$
$$R_p^s = (D^s)^{-1}\bar{R}_p^s D^p, \qquad R_p^v = (D^v)^{-1}\bar{R}_p^v D^p.$$

Example 3.5 (Enzyme kinetics)**.** TBA ∇

Flux balance analysis

Flux balance analysis is a technique for studying the relative rate of different reactions in a complex reaction system. We are most interested in the case where there may be multiple pathways in a system, so that the number of reactions m is greater than the number of species n. The dynamics

$$\dot{s} = Nv(s,p)$$

thus have the property that the matrix N has more columns that rows and hence there are multiple reactions that can produce a given set of species. Flux balance is often applied to pathway analysis in metabolic systems to understand the limiting

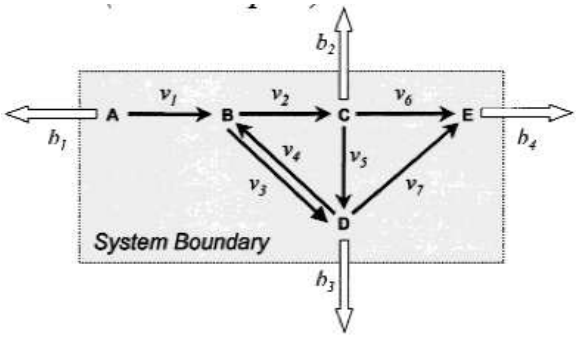


Figure 3.6: Flux balance analysis.

pathways for a given species and the the effects of changes in the network (e.g., through gene deletions) to the production capacity.

To perform a flux balance analysis, we begin by separating the reactions of the pathway into internal fluxes v_i versus exchanges flux v_e, as illustrated in Figure 3.6. The dynamics of the resulting system now be written as

$$\dot{s} = Nv(s,p) = N \begin{pmatrix} v_i \\ v_e \end{pmatrix} = Nv_i(s,p) - b_e,$$

where $b_e = -Nv_e$ represents the effects of external fluxes on the species dynamics. Since the matrix N has more columns that rows, it has a *right* null space and hence there are many different internal fluxes that can produce a given change in species.

In particular, we are interested studying the steady state properties of the system. In this case, we have that $\dot{s} = 0$ and we are left with an algebraic system

$$Nv_i = b_e.$$

Power law formalism

Chemical reaction rate equations are nonlinear differential equations whenever two or more species interact. However, the nonlinearities are very structured: they can be decomposed into a stoichiometry matrix and flux rates, and the flux rates typically consist of either polynomial terms or simple ratios of polynomials (e.g., Michaelis-Menten kinetics or Hill functions). In this section we consider power law representations that exploit these properties and attempt to provide simpler techniques for understand the relationships between species concentrations, parameter values and flux rates. This formalism was developed by Savageau [44] and is also called biochemical systems theory (BST).

The general power law formalism describes a set of reaction dynamics using a set of differential equations of the form

$$\frac{dx_i}{dt} = \sum_r E_r \prod_{j=1}^{n+m} x_j^{\epsilon_j^r} - \sum_s E_s \prod_{j=1}^{n+m} x_j^{\epsilon_j^s}, \quad i = 1, \ldots n. \tag{3.10}$$

Here, x_i is the concentration for species i, with $i = 1, \ldots, n$ representing internal species and $i = n+1, \ldots, m$ representing external species, and the dynamics are broken into two summations. The first sum is over the set of reactions that produce the species x_i and the second is over the reactions that utilize x_i (and so decrease its concentration). The linear coefficients E_r and E_s are the activity levels and correspond to the rate constants (for metabolic networks the rate constants are often proportional to a fixed enzyme level, hence the use of the symbol E). The exponents ϵ_j^r and ϵ_i^s are the *kinetic orders* of the production and utilization reactions.

In this general form, the power law formalism is able to exactly capture mass action kinetics, but it does not provide any additional structure. If we consider a general rate equation of the form $v_i(x_1, \ldots, x_{n+m})$, we can approximate this function in a number of ways. The first is through its linearization,

$$v_i(x_1, \ldots, x_{n+m} \approx v_i(x_{1,e}, \ldots, x_{n+m,e}) + \sum \frac{\partial v}{\partial x_j}(x_j - x_{j,e}) + \text{higher order terms}.$$

We have used exactly this approximation in previous sections.

A different approximation can be obtained by taking a Taylor series expansion for $\log v_i$:

$$\log v_i(x_1, \ldots, x_{n+m} \approx \log v_i(x_{1,e}, \ldots, x_{n+m,e}) + \sum \frac{\partial \log v_i}{\partial \log x_j}(\log x_i - \log x_{i,e}) + \text{higher order terms}.$$

If we define

$$g_{i,j} = \frac{\partial \log v_i}{\partial \log x_j} = \frac{x_j}{v_i} \cdot \frac{\partial v_i}{\partial x_j}$$

and collect terms, we have

$$\log v_i(x) \approx \log \alpha_i + g_{i,1} \log x_1 + \cdots + g_{i,n+m} \log x_{n+m}.$$

Converting this back from log coordinates, we can thus right

$$v_i(x) \approx \alpha_i \prod_{j=1}^{n+m} x_j^{g_{i,j}}.$$

Using this approximation on the sums in equation (3.10), we can approximate the resulting dynamics as

$$\frac{dx_i}{dt} = \alpha_i \prod x_j^{g_i,j} - \beta_i \prod x_j^{h_i,j},$$

where α_i and $g_{i,j}$ are the rate constant and kinetic orders for the production terms and β_i and $h_{i,j}$ are the rate constant and kineeetic orders for reactions that utilize x_i. While this is only an approximation, its form is convenientt for performing equilibrium analyses. In particular, if $\dot{x}_i = 0$ then we can equate the production rate to the utilization rate adn take the log of this expression to obtain

$$\log\alpha_i + \sum g_{i,j}\log x_j = \log\beta_i + \sum h_{i,j}\log x_j.$$

This is now a linear equation for the logs of the concentrations in terms of the various parameters that enter the system.

3.3 Limit Cycle Behavior

Before studying periodic behavior of systems in $\mathbb{R}^n$, we study the behavior of systems in $\mathbb{R}^2$ as several high dimensional systems can be often well approximated by systems in two dimensions by, for example, employing quasi-steady state approximations. For systems in $\mathbb{R}^2$, we will see that there are only two types of solutions: those converging (diverging) from steady states and periodic solutions. That is, chaos can be ruled out in two-dimensional systems.

Consider the system $\dot{x} = f(x)$, in which $f(x)$ is often referred to as vector field, and let $x(t,x_0)$ denote its solution starting at x_0 at time $t = 0$, that is, $\dot{x}(t,x_0) = f(x(t,x_0))$ and $x(0,x_0) = x_0$. We say that $x(t,x_0)$ is a *periodic solution* if there is $T > 0$ such that $x(t,x_0) = x(t+T,x_0)$ for all $t \in \mathbb{R}$. Here, we seek to answer two questions: (a) when does a system $\dot{x} = f(x)$ admit periodic solutions? (b) When are these periodic solutions stable or asymptotically stable?

We first tackle these questions for the case $x \in \mathbb{R}^2$. The first result that we next give provides a simple check to rule out periodic solutions for system in $\mathbb{R}^2$. Specifically, let $(x,y) \in \mathbb{R}^2$ and consider

$$\begin{aligned} \dot{x} &= f(x,y) \\ \dot{y} &= g(x,y), \end{aligned} \tag{3.11}$$

in which the functions g, f are smooth. Then, we have the following result:

Theorem 3.1 (Bendixson's Criterion)**.** *If on a simply connected region $D \subset \mathbb{R}^2$ (i.e., there are no holes in it) the expression*

$$\frac{\partial f}{\partial x} + \frac{\partial g}{\partial y}$$

is not identically zero and does not change sign, then system (3.11) has no closed orbits that lie entirely in D.

Example 3.6. Consider the system

$$\begin{aligned} \dot{x} &= -y^3 + \delta x^3 \\ \dot{y} &= x^3, \end{aligned}$$

with $\delta \geq 0$. We can compute $\frac{\partial f}{\partial x} + \frac{\partial g}{\partial y} = 3\delta x^2$, which is positive in all $\mathbb{R}^2$ if $\delta \neq 0$. If $\delta \neq 0$, we can thus conclude from Bendixson's criterion that there are no periodic solutions. Investigate as an exercise what happens when $\delta = 0$. ∇

In order to provide the main result to state the existence of a stable periodic solution, we need the concept of omega-limit set of a point p, denoted $\omega(p)$. Basically, the omega-limit set $\omega(p)$ denotes the set of all points to which the trajectory of the system starting from p tends as time approaches infinity. This is formally defined in the following definition

Definition 3.1. A point $\bar{x} \in \mathbb{R}^n$ is called an *omega-limit point* of $p \in \mathbb{R}^n$ if there is a sequence of times $\{t_i\}$ with $t_i \to \infty$ for $i \to \infty$ such that $x(t_i, p) \to \bar{x}$ as $i \to \infty$. The *omega limit set* of p, denoted $\omega(p)$, is the set of all omega-limit points of p.

The omega-limit set of a system has several relevant properties, among which the fact that it cannot be empty and that it must be a connected set.

The following theorem, completely characterizes the omega limit set of any point for a system in $\mathbb{R}^2$.

Theorem 3.2 (Poincarè-Bendixson)**.** *Let M be a positively invariant region for the system $\dot{x} = f(x)$ with $x \in \mathbb{R}^2$ (i.e., any trajectory that starts in M stays in M for all $t \geq 0$). Let $p \in M$, then one of the following possibilities holds for $\omega(p)$:*

(i) *$\omega(p)$ is a steady state;*

(ii) *$\omega(p)$ is a closed orbit;*

(iii) *$\omega(p)$ consists of a finite number of steady states and orbits, each starting (for $t = 0$) and ending (for $t \to \infty$) at one of the fixed points.*

This theorem has two important consequences:

1. If the system does not have steady states in M, since $\omega(p)$ is not empty, it must be a periodic solution;

2. If there is only one steady state in M and it is unstable and not a saddle (i.e., the eigenvalues of the linearization at the steady state are both positive), then $\omega(p)$ is a periodic solution.

Example 3.7. Consider the following system in $\mathbb{R}^2$:

$$\begin{aligned} \dot{x} &= x - y - (x^2 + y^2)x \\ \dot{y} &= x + y - (x^2 + y^2)y. \end{aligned}$$

Verify as an exercise that this system admits one equilibrium point only (the origin), which is unstable. Also, show that its trajectories are globally bounded (for example, take a set $x^2 + y^2 = c$ for c large enough and demonstrate that the vector field of the system always points inside the circle $x^2 + y^2 = c$). Therefore, by Poincarè-Bendixson Theorem, we can conclude that the omega-limit set of any point in $\mathbb{R}^2$ different from the origin is a non-zero periodic orbit. ∇

This result holds only for systems in two dimensions. However, there have been recent extensions of this theorem to systems with special structure in $\mathbb{R}^n$. In particular, we have the following result due to Hastings et al. (1977).

Theorem 3.3 (Hastings et al. 1977)**.** *Consider a system $\dot{x} = f(x)$, which is of the form*

$$\begin{aligned} \dot{x}_1 &= f_1(x_n, x_1) \\ \dot{x}_j &= f_j(x_{j-1}, x_j),\ 2 \leq j \leq n \end{aligned}$$

on the set M defined by $x_i \geq 0$ for all i with the following inequalities holding in M:

(i) $\frac{\partial f_i}{\partial x_i} < 0$ *and* $\frac{\partial f_i}{\partial x_{i-1}} > 0$, *for* $2 \leq i \leq n$, *and* $\frac{\partial f_1}{\partial x_n} < 0$*;*

(ii) $f_i(0,0) \geq 0$ *and* $f_1(x_n, 0) > 0$ *for all* $x_n \geq 0$*;*

(iii) The system has a unique steady state $x^* = (x_1^*, ..., x_n^*)$ in M such that $f_1(x_n, x_1) < 0$ if $x_n > x_n^*$ and $x_1 > x_1^*$, while $f_1(x_n, x_1) > 0$ if $x_n < x_n^*$ and $x_1 < x_1^*$;

(iv) $\frac{\partial f_1}{\partial x_1}$ *is bounded above in M.*

Then, if the Jacobian of f at x^* has no repeated eigenvalues and has any eigenvalue with positive real part, then the system has a non-constant periodic solution in M.

This theorem states that for a system with cyclic structure in which the cycle "has negative gain", the instability of the steady state (under some technical assumption) is equivalent to the existence of a periodic solution. This'theorem, however, does not provide information about whether the orbit is attractive or not, that is, of whether it is an omega-limit set of any point in M. This stability result is implied by a more recent theorem due to Mallet-Paret and Smith (1990), for which we provide a simplified statement as follows.

Theorem 3.4 (Mallet-Paret and Smith, 1990)**.** *Consider the system $\dot{x} = f(x)$ with the following cyclic feedback structure*

$$\begin{aligned} \dot{x}_1 &= f_1(x_n, x_1) \\ \dot{x}_j &= f_j(x_{j-1}, x_j),\ 2 \leq j \leq n \end{aligned}$$

on a set M defined by $x_i \geq 0$ for all i with all trajectories starting in M bounded for $t \geq 0$. Then, the omega-limit set $\omega(p)$ of any point $p \in M$ can be one of the following:

(a) A steady state;

(b) A non-constant periodic orbit;

(c) A set of steady states connected by homoclinic or heteroclinic orbits.

A heteroclinic orbit is an orbit that starts (for $t = 0$) at a steady state and ends (for $t \to \infty$) into a different steady state. A homoclinic orbit is an orbit that starts and ends at the same steady state. It is thus clear that a steady state whose linearization admits all positive or all negative eigenvalues cannot have a homoclinic orbit. As a consequence of the theorem, then we have that for a system with cyclic feedback structure that admits one steady state only and at which the linearization has all eigenvalues with positive real part, the omega limit set must be a periodic orbit.

Let for some $\delta_i \in \{1, -1\}$ be $\delta_i \frac{\partial f_i(x, x_{i-1})}{\partial x_{i-1}} > 0$ for all $0 \leq i \leq n$ and define $\Delta := \delta_1 \cdot \ldots \cdot \delta_n$. One can show that the sign of Δ is related to whether the system has one or multiple steady states.

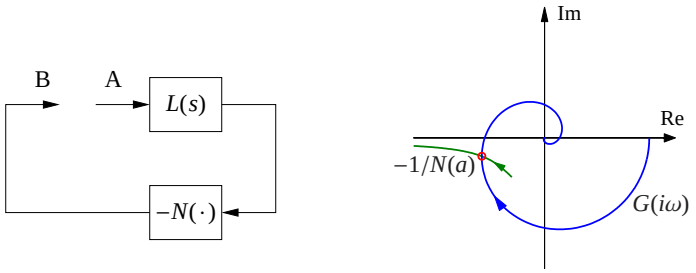
Therefore, a system with a cyclic feedback structure and a unique equilibrium point at which the linearization has all eigenvalues with positive real part admits a stable periodic orbit.

3.4 Analysis Using Describing Functions

Unlike the case of linear systems, where it is possible to full characterize the solutions of a model and there are a wide variety of analysis techniques available, the behavior of nonlinear systems is harder to analyze, especially away from equilibrium points (where the linearization gives a good approximation). One of the more useful techniques for studying the behavior of nonlinear systems is the method of harmonic balance, of which a special case is the method of describing functions. This section explores the use of harmonic balance and describing functions for analyzing nonlinear systems, including the detection and analysis of limit cycles and the propagation of noise through nonlinear systems.

Describing functions (AM08)

For special nonlinear systems like the one shown in Figure 3.7a, which consists of a feedback connection between a linear system and a static nonlinearity, it is possible to obtain a generalization of Nyquist's stability criterion based on the idea of *describing functions*. Following the approach of the Nyquist stability condition, we will investigate the conditions for maintaining an oscillation in the system. If the linear subsystem has low-pass character, its output is approximately sinusoidal even if its input is highly irregular. The condition for oscillation can then be found by exploring the propagation of a sinusoid that corresponds to the first harmonic.



(a) Block diagram (b) Nyquist plot

Figure 3.7: Describing function analysis. A feedback connection between a static nonlinearity and a linear system is shown in (a). The linear system is characterized by its transfer function $L(s)$, which depends on frequency, and the nonlinearity by its describing function $N(a)$, which depends on the amplitude a of its input. The Nyquist plot of $L(i\omega)$ and the plot of the $-1/N(a)$ are shown in (b). The intersection of the curves represents a possible limit cycle.

To carry out this analysis, we have to analyze how a sinusoidal signal propagates through a static nonlinear system. In particular we investigate how the first harmonic of the output of the nonlinearity is related to its (sinusoidal) input. Letting F represent the nonlinear function, we expand $F(e^{i\omega t})$ in terms of its harmonics:

$$F(ae^{i\omega t}) = \sum_{n=0}^{\infty} M_n(a)e^{i(n\omega t+\phi_n(a))},$$

where $M_n(a)$ and $\phi_n(a)$ represent the gain and phase of the nth harmonic, which depend on the input amplitude since the function F is nonlinear. We define the describing function to be the complex gain of the first harmonic:

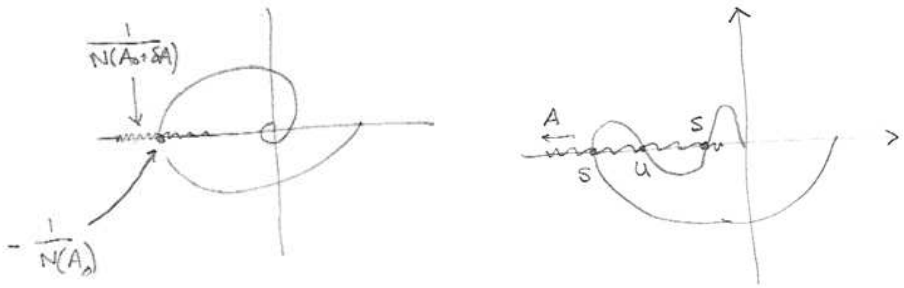
$$N(a) = M_1(a)e^{i\phi_n(a)}. \tag{3.12}$$

The function can also be computed by assuming that the input is a sinusoid and using the first term in the Fourier series of the resulting output.

Arguing as we did when deriving Nyquist's stability criterion, we find that an oscillation can be maintained if

$$L(i\omega)N(a) = -1. \tag{3.13}$$

This equation means that if we inject a sinusoid at A in Figure 3.7, the same signal will appear at B and an oscillation can be maintained by connecting the points. Equation (3.13) gives two conditions for finding the frequency ω of the oscillation and its amplitude a: the phase must be 180°, and the magnitude must be unity. A convenient way to solve the equation is to plot $L(i\omega)$ and $-1/N(a)$ on the same



(a) (b)

Figure 3.8: Heuristic stability of limit cycles using describing functions. (a) To check if a perturbation from amplitude a_0 to amplitude $a_0 + \delta a$ is stabilizing, we check to see if the Nyquist criterion is satisfied for the original frequency response and the perturbed critical point $P_1 = 1/N(a_0 + \delta a)$. (b) An example of a nonlinear system with multiple limit cycles. Stable limit cycles are labeled 's' and unstable limit cycles are labeled 'u'.

diagram as shown in Figure 3.7b. The diagram is similar to the Nyquist plot where the critical point -1 is replaced by the curve $-1/N(a)$ and a ranges from 0 to ∞.

It is possible to define describing functions for types of inputs other than sinusoids. Describing function analysis is a simple method, but it is approximate because it assumes that higher harmonics can be neglected. Excellent treatments of describing function techniques can be found in the texts by Atherton [5] and Graham and McRuer [19].

Example 3.8 (Repressilator)**.** ∇

Stability of limit cycles using describing functions

In order to check the stability of a limit cycle, we must reason about how solutions that have initial conditions near the limit cycle evolve in time and whether they move closer to the limit cycle (asymptotic stability) or diverge from the limit cycle (instability).

We begin by arguing heuristically, using the Nyquist plot in Figure 3.7b. Suppose that we were to consider a perturbed limit cycle with amplitude $a_0 + \delta a$, where a_0 is the amplitude of the limit cycle predicted by the describing function method. If we did so, then the point of intersection of the describing function and the frequency response would move from $P_0 = -1/N(a_0)$ to $P_1 = -1/N(a_0 + \delta a)$, as shown in Figure 3.8a. Now evaluate the Nyquist criterion for the frequency response with critical point P_1. If the criterion indicates that the perturbed system is stable (i.e., no net encirclements of P_1 for a stable process), then intuitively the amplitude of the perturbed solution would decrease and we would return to our original amplitude limit cycle. Conversely, if the Nyquist criterion with critical point P_1 indicates

instability, then the oscillation would grow and hence we can infer that the limit cycle is unstable. Figure 3.8b shows a situation with multiple limit cycles with some stable and some unstable.

While this heuristic method is intuitively appealing, it does not always give the correct answer. Indeed, even the prediction of the existence of a limit cycle using describing functions can be incorrect unless the system satisfies some additional conditions. We present here one such set of conditions, due to Mees [**?**].

Suppose that (ω_0, a_0) satisfies the describing function balance equation $P(i\omega_0) = -1/N(a_0)$ and that the the frequency response curve and the describing function locus are transverse (not tangent) at their intersection. Define

$$\begin{aligned}
\rho(\omega)^2 &= \sum_{k=3,5,9,\ldots} |P(ik\omega_0)|^2, && \text{“gain of harmonics”} \\
p(a)^2 &= \|n(a\sin t)\|_2^2 - |aN(a)|^2, && \text{“first harmonic error”} \\
q(a,\epsilon) &= \|m(a\sin t,\epsilon)\|_2, && \text{“slope bound”} \\
m(x,\epsilon) &= \max\{|N(x+\epsilon) - N(x)|, |N(x-\epsilon) - N(x)|\}.
\end{aligned}$$

Now find an ϵ such that for all (ω, a) near (ω, a_0),

$$\rho(\omega)(p(a) + q(a,\epsilon)) \leq \epsilon$$

and let $\Omega \in \mathbb{R}_+^2$ be the set of (ω, a) such that

$$|N(a) + 1/G(i\omega)| \leq q(a,\epsilon)/a.$$

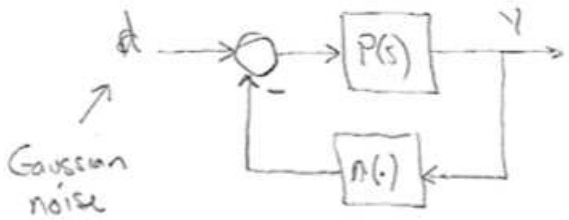
Theorem 3.5. *Suppose Ω is bounded and there exists a unique $(\omega, a_0) \in \Omega$ satisfying the balance equation. Then there exists a periodic solution of the form $y(t) = a\sin(\omega t) + y^*(t)$ with remnant $\|y^*\|_\infty \leq \epsilon$.*

Sketch of proof. Reduced to the contraction mapping theorem, which generates ρ, p and q. □

The basic idea behind this theorem is that if the harmonics around the loop die off sufficiently fast, then we can insure that there is truly a periodic solution and bound the error of the higher harmonics. There is also a graphical version of the stability theorem that checks for “complete intersections” between the describing function locus and the Nyquist curve [**?**].

Mathematically, the stability of a limit cycle can be analyzed by taking the linearization of the system around the (non-equilibrium) solution. To see how this is done, consider a nonlinear system of the form

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



(a)

Figure 3.9: Random input describing function analysis.

that has a solution $x_d(t)$ that is periodic with period T. To compute the linearization of the dynamics around the equilibrium point, we compute the dynamics of the error $e = x - x_d$:

$$\dot{e} = f(x) - f(x_d) = F(e, x_d(t)) \approx A(t)e$$

where $A(t)$ is the time-varying linearization given by

$$A(t) = \frac{\partial F}{\partial e}(e, x_d)\Big|_{e=0, x_d(t)}.$$

The dynamics matrix $A(t)$ is periodic and so the dynamics of the linearization are a given by a periodic, linear ordinary differential equation.

The dynamics of periodic linear systems can be studied using *Floquet* theory, which we briefly review here. Let $\Phi(t, 0)$ be the (T-periodic) fundamental matrix for $\dot{e} = A(t)e$, so that the solution is given by $x(t) = \Phi(t, 0)x(0)$. It can be show that $\Phi(t, 0)$ has the form $\phi(t, 0) = P(t)e^{Ft}$ where $P(t) = P(t + T) \in \mathbb{R}^{n\times n}$ is a periodic matrix and $F \in \mathbb{R}^{n\times n}$ is a constant matrix. We can now check stability by examining the eigenvalues of the matrix e^{FT}, which corresponds to the "first return" map for the system.

Random input describing functions

In addition to allowing prediction and analysis of limit cycles, describing functions can also be used to analyze the propagation of noise through nonlinear feedback systems. This approach is known as the *random input describing function* method.

As in the single input describing function method, we begin with a system in the form of a a linear system with a nonlinear feedback, as shown in Figure 3.9a. To analyze this system, we construct an input that contains both a sinusoid and a random input $r(t)$:

$$y = b + a\sin(\omega t + \phi) + r(t),$$

where b is the bias term, a is the amplitude of the sinusoidal term, ϕ is a uniform random variable and $r(t)$ is a stationary Gaussian random process with variance

σ^2 and correlation $\rho(\tau)$.[1] We approximate the response of the system through the nonlinearity by

$$N(y(t)) \approx N_b b + N_a a \sin(\omega t + \phi) + N_r r(t),$$

where N_b is called the *bias gain*, N_a is the sinusoidal gain and N_r is the stochastic gain. These functions are given by

$$\begin{aligned}
N_b(b,a,\sigma) &= \frac{1}{b}E\{f(y)\} = \frac{1}{(2\pi)^{3/2}\sigma b}\int_0^{2\pi}\int_{-\infty}^{\infty} f(b + a\sin\theta + r(t))e^{-\frac{r^2}{2\sigma^2}}\,dr\,d\theta \\
N_a(b,a,\sigma) &= \frac{2}{a}E\{f(y)\sin\theta\} = \frac{2}{(2\pi)^{3/2}\sigma a}\int_0^{2\pi}\int_{-\infty}^{\infty} f(b + a\sin\theta + r(t))\sin\theta e^{-\frac{r^2}{2\sigma^2}}\,dr\,d\theta \\
N_r(b,a,\sigma) &= \frac{1}{\sigma^2}E\{f(y)r\} = \frac{1}{(2\pi)^{3/2}\sigma^3}\int_0^{2\pi}\int_{-\infty}^{\infty} f(b + a\sin\theta + r(t))re^{-\frac{r^2}{2\sigma^2}}\,dr\,d\theta
\end{aligned} \tag{3.14}$$

The random input describing function method has a number of special cases. If we take $\sigma = 0$, then it can be shown that we recover the standard describing function method. If we instead take $a = 0$, we can study how noise propagates through the system. Recall that in the linear case, where the feedback term is given by a constant gain N, the spectral density of the output y is given by

$$S_y(\omega) = H_{yd}(-i\omega)S_d(\omega)H_{yd}(i\omega), \qquad \sigma_y = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_y(\omega)\,d\omega.$$

In the nonlinear case, we replace the feedback gain N with $N_r(\sigma_y)$ so that

$$\tilde{H}_{yd}(s) = \frac{P(s)}{1 + P(s)N_r(\sigma_y)}, \qquad \sigma_y = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{H}_{yd}(-i\omega)S_d(\omega)\tilde{H}_{yd}(i\omega). \tag{3.15}$$

Note that this equation gives an algebraic relationship for σ_y that can be solved and then used to compute $N_r(\sigma)$ and $S_y(\omega)$.

Consider next the case of both a limit cycle and random noise,

$$y(t) = a\sin(\omega t + \phi) + r(t).$$

We now look for solutions of the coupled equations

$$\begin{aligned}
&\tilde{H}_{yd}(s) = \frac{P(s)}{1 + P(s)N_r(\sigma_y)}, \qquad \sigma_y = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{H}_{yd}(-i\omega)S_d(\omega)\tilde{H}_{yd}(i\omega), \\
&N_a(a,\sigma_y)P(i\omega_0) = -1.
\end{aligned} \tag{3.16}$$

If we can find a, σ_y and ω_0 that satisfy all of the equations, then we get a description of $y(t)$.

It is interesting to note that it can sometimes happen that $S_d(\omega)$ can cause an unstable (noiseless) system to be stable. Similarly, we can get a system with $N_r(0,\sigma_y)$ that destabilizes and otherwise stable system.

[1] These are described in more detail in Chapter 4.

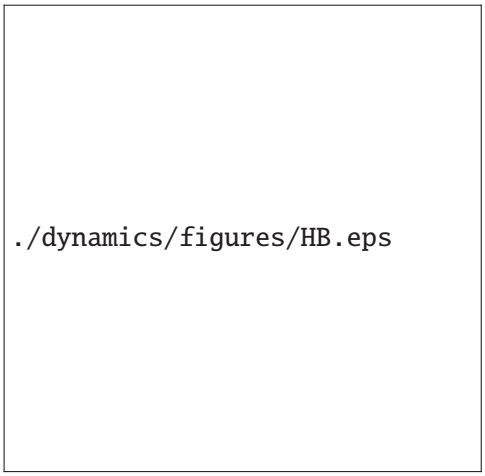


Figure 3.10: Hopf Bifurcation.

3.5 Bifurcations

Hopf bifurcation is a technique that is often used to understand whether a system admits a periodic orbit when some parameter is varied. Usually, such an orbit is a small amplitude periodic orbit that is present in the close vicinity of an unstable steady state.

Consider the system dependent on a parameter α:

$$\dot{x} = g(x,\alpha), x \in \mathbb{R}^n,\ \alpha \in \mathbb{R},$$

and assume that at the steady state $\bar{x}$ corresponding to $\alpha = \bar{\alpha}$ (i.e., $g(\bar{x},\bar{\alpha}) = 0$), the linearization $\frac{\partial g}{\partial x}(\bar{x},\bar{\alpha})$ has a pair of (non zero) imaginary eigenvalues with the remaining $\mathbb{R}^{n-2}$ eigenvalues having negative real parts. Define the new parameter $\mu := \alpha - \bar{\alpha}$ and re-define the system as

$$\dot{x} = f(x,\mu) := g(x,\mu+\bar{\alpha}),$$

so that the linearization $\frac{\partial f}{\partial x}(\bar{x},0)$ has a pair of (non zero) imaginary eigenvalues with the remaining $\mathbb{R}^{n-2}$ eigenvalues having negative real parts. Denote by $\lambda(\mu) = \beta(\mu) + i\omega(\mu)$ the eigenvalue such that $\beta(0) = 0$. Then, if $\frac{\partial \beta}{\partial \mu}(\mu = 0) \neq 0$ the system admits a small amplitude almost sinusoidal periodic orbit for μ small enough and the system is said to go through a Hopf bifurcation at $\mu = 0$. If the small amplitude periodic orbit is stable, the Hopf bifurcation is said *supercritical*, while if it is unstable it is said *subcritical*. Figure 3.10 shows diagrams corresponding to these bifurcations.

In order to determine whether a Hopf bifurcation is supercrictical or subcritical, it is necessary to calculate a "curvature" coefficient, for which there are formulas (Marsden and McCrocken, 1976) and available bifurcation sofwtare, such as

AUTO. In practice, it is often enough to calculate the value $\bar{\alpha}$ of the parameter at which Hopf bifurcation occurs and simulate the system for values of the parameter α close to $\bar{\alpha}$. If a small amplitude limit cycle appears, then the bifurcation must be supercritical.

The Hopf bifurcation result is based on the center manifold theory for nonlinear dynamical systems. For a rigorous treatment of Hopf bifurcation is thus necessary to study center manifold theory first, which is outside the scope of this text. For details, the reader is referred to Wiggins book on dynamical systems and chaos.

3.6 Model Reduction Techniques

The techniques that we have developed in this chapter can be applied to a wide variety of dynamical systems. However, many of the methods require significant computation and hence we would like to reduce the complexity of the models as much as possible before applying them. In this section we review methods for doing such a reduction in the complexity of the models. Most of the techniques are based on the common idea that if we are interested in the slower time scale dynamics of a system, the fast time scale dynamics can be approximated by their equilibrium solutions. This idea was introduced in Chapter 2 in the context of reduced order mechanisms; we present a more mathematical analysis of such systems here.

Singular Perturbation

Let $(x,y) \in D := D_x \times D_y \subset \mathbb{R}^n \times \mathbb{R}^m$ and consider the vector field

$$\dot{x} = f(x,y), \;\; \epsilon\dot{y} = g(x,y), \; (x(0),y(0)) = (x_0,y_0)$$

in which $0 < \epsilon \ll 1$ is a small parameter. Since $\epsilon \ll 1$, the absolute value of the time derivative of y can be much larger than the time derivative of x, resulting in y dynamics that are much faster than the x dynamics. That is, this system has a slow time scale evolution (in x) and a fast time-scale evolution (in y). If we are interested only in the slower time scale, then the above system can be approximated (under suitable conditions) by the *reduced system*

$$\dot{\bar{x}} = f(\bar{x},\bar{y}), \;\; 0 = g(\bar{x},\bar{y}), \; \bar{x}(0) = x_0.$$

Letting $y = \gamma(x)$ (called the *slow manifold*) be the locally unique solution of $g(x,y) = 0$, we can approximate the dynamics in x as

$$\dot{\bar{x}} = f(\bar{x},\gamma(\bar{x})), \; x(0) = x_0.$$

We seek to determine under what conditions the solution $x(t)$ is "close" to the solution $\bar{x}(t)$ of the reduced system. This problem can be addressed by analyzing

the fast dynamics. Letting $\tau = t/\epsilon$ be the fast time scale, we have that

$$\frac{dx}{d\tau} = \epsilon f(x,y), \qquad \frac{dy}{d\tau} = g(x,y), \qquad (x(0),y(0)) = (x_0,y_0),$$

so that when $\epsilon \ll 1$, $x(\tau)$ does not appreciably change. Therefore, the above system in the τ time scale can be approximated by

$$\frac{dy}{d\tau} = g(x_0,y), \qquad y(0) = y_0,$$

in which x is "frozen" at the initial condition. This system is usually referred to as the *boundary layer* system. If for all x_0, we have that $y(\tau)$ converges to $\gamma(x_0)$, then for $t > 0$ we will have that the solution $x(t)$ is well approximated by the solution $\bar{x}(t)$ to the reduced system. This qualitative explanation is more precisely captured by the following theorem (originally due to Tikonov).

Theorem 3.6. *Assume that*

$$\frac{\partial}{\partial y} g(x,y)\bigg|_{y=\gamma(x)} < 0$$

uniformly for $x \in D_x$. Let the solution of the reduced system be uniquely defined for $t \in [0,t_f]$. Then, for all $t_b \in (0,t_f]$ there is a constant $\epsilon^* > 0$ and set $\Omega \subseteq D$ such that

$$x(t) - \bar{x}(t) = O(\epsilon) \text{ uniformly for } t \in [0,t_f],$$
$$y(t) - \gamma(\bar{x}(t)) = O(\epsilon) \text{ uniformly for } t \in [t_b,t_f],$$

provided $\epsilon < \epsilon^*$ and $(x_0,y_0) \in \Omega$.

Example 3.9 (Linear system)**.** Consider the following linear system

$$\begin{aligned} \dot{x}_1 &= -x_1 \\ \dot{x}_2 &= -\frac{1}{\epsilon}x_2 + \frac{1}{\epsilon}x_1, \ \epsilon > 0, \end{aligned} \tag{3.17}$$

in which ϵ is very small. This system has two eigenvalues equal to -1 and $-1/\epsilon$ with corresponding eigenvectors $(1-\epsilon,1)$ and $(0,1)$, respectively. The slow manifold, obtained by multiplying both sides of the second equation in system (3.17) by ϵ and setting $\epsilon = 0$, is given by $x_2 = x_1$ and the boundary layer system is exponentially stable. The reduced system is just given by

$$\dot{\bar{x}}_1 = -\bar{x}_1, \text{ and } \bar{x}_2(t) = \bar{x}_1(t).$$

The trajectories of the system along with the slow manifold are represented in Figure 3.11. The initial conditions that are not on the slow manifold quickly converge to the slow manifold and then they converge to the origin. ∇

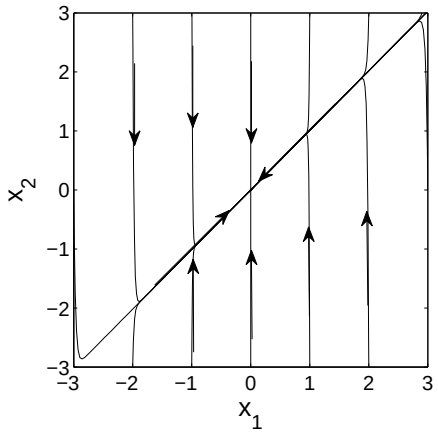


Figure 3.11: Simulation results for the system in equations (3.17). Trajectories in the x_1, x_2 plane.

Example 3.10 (Enzymatic reaction)**.** Let's go back to the enzymatic reaction

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

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. The corresponding ODE system is given by

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

By assuming that $k_r, k_f \gg k_{cat}$, we obtained that approximatively $\frac{dC}{dt} = 0$ and thus that $C = \frac{E_{tot} S}{S + K_m}$, with $K_m = \frac{k_r + k_{cat}}{k_f}$ and $\frac{dP}{dt} = \frac{V_{max} S}{S + K_m}$ with $V_{max} = k_{cat} E_{tot}$. From this, it also follows that

$$\frac{dE}{dt} \approx 0 \text{ and } \frac{dS}{dt} \approx -\frac{dP}{dt}. \tag{3.18}$$

How good is this approximation? By applying the singular perturbation method, we will obtain a clear answer to this question. Specifically, define $a := k_f / k_r$ and

take the system to standard singular perturbation form by defining the small parameter as $\epsilon := \frac{k_{cat}}{k_r}$, so that $k_f = \frac{k_{cat}}{\epsilon} a$, $k_r = \frac{k_{cat}}{\epsilon}$, and the system becomes

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

One cannot directly apply singular perturbation theory on this system because one can verify from the linearization of the first three equations that the boundary layer dynamics is not locally exponentially stable as there are two zero eigenvalues. This is because the three variables E, S, C are not independent. Specifically, $E = E_{tot} - C$ and $S + C + P = S(0) = S_{tot}$, assuming that initially we have S in amount $S(0)$ and no amount of P and C in the system. Given these conservation laws, the system can be re-written as

$$\begin{aligned}
\epsilon \frac{dC}{dt} &= a k_{cat} (E_{tot} - C) \cdot (S_{tot} - C - P) - k_{cat} C - \epsilon k_{cat} C \\
\frac{dP}{dt} &= k_{cat} C.
\end{aligned}$$

Under the assumption made in the analysis of the enzymatic reaction that $S_{tot} \gg E_{tot}$, we have that $C \ll S_{tot}$ so that the equations finally become

$$\begin{aligned}
\epsilon \frac{dC}{dt} &= a k_{cat} (E_{tot} - C) \cdot (S_{tot} - P) - k_{cat} C - \epsilon k_{cat} C \\
\frac{dP}{dt} &= k_{cat} C.
\end{aligned}$$

One can verify (show as an exercise) that in this system, the boundary layer dynamics is locally exponentially stable, so that setting $\epsilon = 0$ one obtains $\bar{C} = \frac{E_{tot}(S_{tot} - \bar{P})}{(S_{tot} - P) + K_m} =: g(\bar{P})$ and thus that the slow dynamics of the system are given by

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

From the conservation law $\bar{S} + \bar{C} + \bar{P} = S(0) = S_{tot}$, we obtain that $\frac{d\bar{S}}{dt} = -\frac{d\bar{P}}{dt} - \frac{d\bar{C}}{dt}$, in which now $\frac{d\bar{C}}{dt} = \frac{\partial g}{\partial P}(\bar{P}) \cdot \frac{dP}{dt}$. Therefore

$$\frac{d\bar{S}}{dt} = -\frac{d\bar{P}}{dt}\left(1 + \frac{\partial g}{\partial P}(\bar{P})\right), \; \bar{S}(0) = S_{tot} - g(\bar{P}(0)) - \bar{P}(0) \tag{3.19}$$

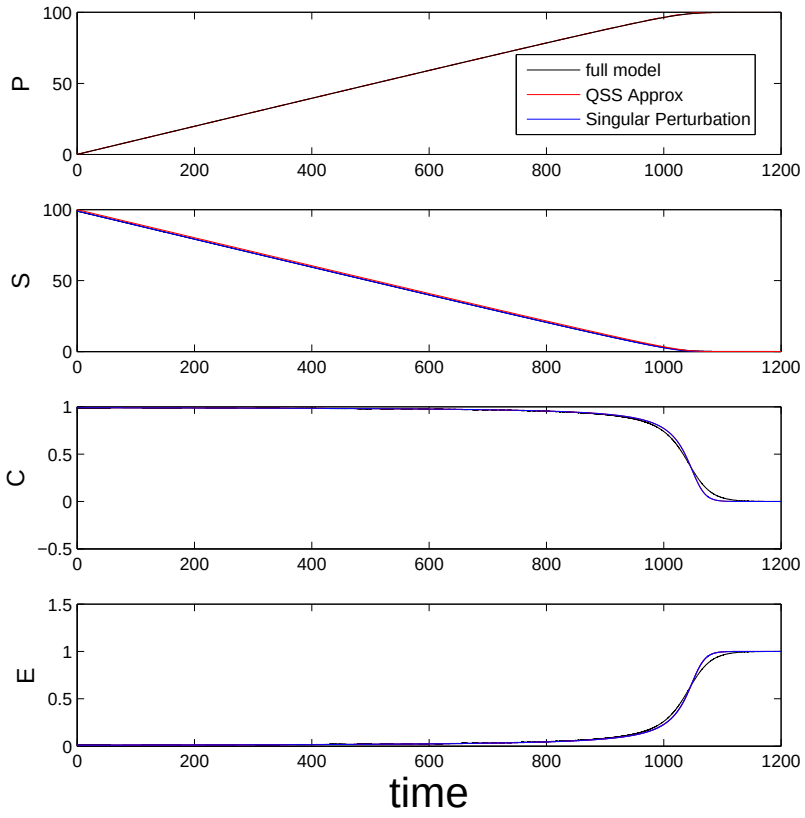


Figure 3.12: Simulation results for the enzymatic reaction comparing the approximations from singular perturbation and from the quasi-steady state approximation. Here, we have $S_{tot} = 100$, $E_{tot} = 1$, $k_r = k_f = 10$, and $k_{cat} = 0.1$.

and

$$\frac{d\bar{E}}{dt} = -\frac{d\bar{C}}{dt} = -\frac{\partial g}{\partial P}(\bar{P})\frac{d\bar{P}}{dt},\ E(0) = E_{tot} - g(\bar{P}(0)), \tag{3.20}$$

which are different from expressions (3.18). Specifically, these expressions are close to those in (3.18) only when $\frac{\partial g}{\partial P}(\bar{P})$ is small enough. In the plots of Figure 3.12, we show the time trajectories of the original system, of the Michaelis-Menten quasi-steady state approximation, and of the singular perturbation approximation. The trajectories of $E(t)$ and of $S(t)$ for the quasi-steady state approximation have been obtained from the conservation laws once $P(t)$ and $C(t)$ are determined. The trajectories of these variables for the singular perturbation approximation have been obtained directly integrating equations (3.19) and (3.20). Notice that the quasi-steady state approximations $\frac{dC}{dt} \approx 0$ and $\frac{dE}{dt} \approx 0$ are well representing the

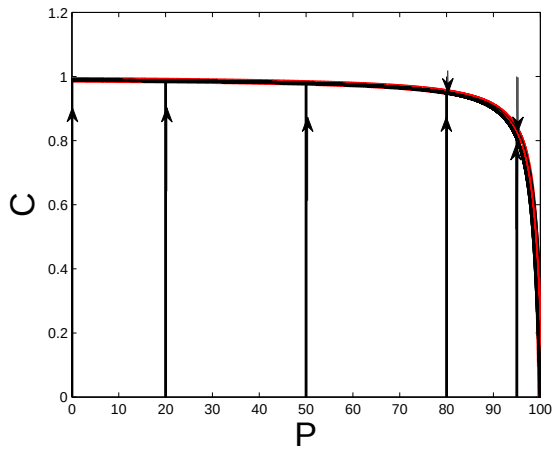


Figure 3.13: The slow manifold of the system $C = g(P)$ is shown in red. In black, we show the trajectories of the the full system. These trajectories collapse into an ϵ-neighbor of the slow manifold. Here, we have $S_{tot} = 100$, $E_{tot} = 1$, $k_r = k_f = 10$, and $k_{cat} = 0.1$.

dynamics of the C and E variables only while $S(t)$ is large enough. By contrast, equations (3.19-3.20) well represent the system even when the substrate goes to zero. In Figure 3.13, we show the curve $C = g(P)$ (in red) and the trajectories of the full system in black. All of the trajectories of the system immediately collapse into an ϵ-neighbor of the curve $C = g(P)$. ∇

Balanced truncation

Principle component analysis (PCA)

Chapter 4
Stochastic Behavior

In this chapter we explore stochastic behavior in biomolecular systems, building on our preliminary discussion of stochastic modeling in Section 2.2. We begin by reviewing the various methods for modeling stochastic processes, including the chemical master equation (CME), the chemical Langevin equation (CLE) and the Fokker-Planck equation (FPE). Given a stochastic description, we can then analyze the behavior of the system using a variety of stochastic simulation and analysis tools. In many cases, we must simplify the dynamics of the system in order to obtain a tractable model, and we describe several methods for doing so, including finite state projection, linearization and Markov chain representations. We also investigate how to use data to identify some the structure and parameters of stochastic models.

Prerequisites. This chapter makes use of a variety of topics in stochastic processes that are not covered in AM08. Readers should have a good working knowledge of basic probability and some exposure to simple stochastic processes (e.g., Brownian motion).

4.1 Stochastic Systems

We begin by briefly introducing the general notions of stochastic systems in continuous time and with continuous states. Some of the material in this section is drawn from the AM08 supplement on Optimization-Based Control Systems [32].

Review of random variables

Random variables and processes are defined in terms of an underlying *probability space* that captures the nature of the stochastic system we wish to study. A probability space has three elements:

- a *sample space* Ω that represents the set of all possible outcomes;
- a set of *events* $\mathcal{F}$ the captures combinations of elementary outcomes that are of interest; and
- a *probability measure* $\mathcal{P}$ that describes the likelihood of a given event occurring.

Ω can be any set, either with a finite, countable or infinite number of elements. The event space $\mathcal{F}$ consists of subsets of Ω. There are some mathematical limits on the properties of the sets in $\mathcal{F}$, but these are not critical for our purposes here. The probability measure $\mathcal{P}$ is a mapping from $\mathcal{P} : \mathcal{F} \to [0,1]$ that assigns a probability to each event. It must satisfy the property that given any two disjoint sets $A, B \subset \mathcal{F}$, $P(A \cup B) = P(A) + P(B)$. The term *probability distribution* is also to describe a probability measure.

With these definitions, we can model many different stochastic phenomena. Given a probability space, we can choose samples $\omega \in \Omega$ and identify each sample with a collection of events chosen from $\mathcal{F}$. These events should correspond to phenomena of interest and the probability measure $\mathcal{P}$ should capture the likelihood of that even occurring in the system that we are modeling. This definition of a probability space is very general and allows us to consider a number of situations as special cases.

A *random variable* X is a function $X : \Omega \to S$ that gives a value in S, called the state space, for any sample $\omega \in \Omega$. Given a subset $A \subset S$, we can write the probability that $X \in A$ as

$$P(X \in A) = P(\omega \in \Omega : X(\omega) \in A).$$

We will often find it convenient to omit ω when working random variables and hence we write $X \in S$ rather than the more correct $X(\omega) \in S$.

A *discrete random variable* X is a variable that can take on any value from a discrete set S with some probability for each element of the set. We model a discrete random variable by its *probability mass function* $p_X(s)$, which gives the probability that the random variable X takes on the specific value $s \in S$:

$$p_X(s) = \text{probability that } X \text{ takes on the value } s \in S.$$

The sum of the probabilities over the entire set of states must be unity, and so we have that

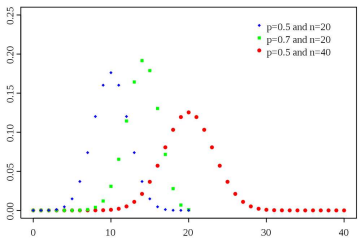
$$\sum_{s \in S} p_X(s) = 1.$$

If A is a subset of S, then we can write $P(X \in A)$ for the probability that X will take on some value in the set A. It follows from our definition that

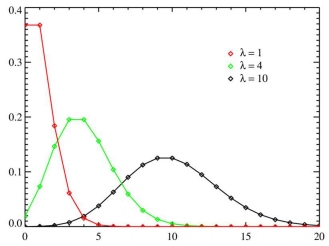
$$P(X \in A) = \sum_{s \in A} p(s).$$

Definition 4.1 (Bernoulli distribution)**.** The Bernoulli distribution is used to model a random variable that takes the value 1 with probability p and 0 with probability $1 - p$:

$$P(X = 1) = p, \qquad P(X = 0) = 1 - p.$$



(a) Binomial distribution



(b) Poisson distribution

Figure 4.1: Probability mass functions for common discrete distributions.

Alternatively, it can be written in terms of its probability mass function

$$p(s) = \begin{cases} p & s = 1 \\ 1 - p & s = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Bernoulli distributions are used to model independent experiments with binary outcomes, such as flipping a coin.

Definition 4.2 (Binomial distribution)**.** The *binomial distribution* models the probability of successful trials in n experiments, given that a single experiment has probability of success p. If we let K_n be a random variable that indicates the number of success in n trials, then the binomial distribution is given by

$$p_{K_n}(k) = P(K_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

for $k = 1, \ldots, n$. The probability mass function is shown in Figure 4.1a.

Definition 4.3 (Poisson distribution)**.** The *Poisson distribution* is used to describe the probability that a given number of events will occur in a fixed interval of time t. The Poisson distribution is defined as

$$p_{N_t}(k) = P(N_t = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}, \tag{4.1}$$

where N_t is the number of events that occur in a period t and λ is a real number parameterizing the distribution. This distribution can be considered as a model for a counting process, where we assume that the average rate of occurrences in a period t is given by λt and λ represents the rate of the counting process. Figure 4.1b shows the form of the distribution for different values of k and λt.

A *continuous (real-valued) random variable X* is a variable that can take on any value in the set of real numbers $\mathbb{R}$. We can model the random variable X according to its *probability distribution P*:

$$P(x_l \leq X \leq x_u) = \text{probability that } x \text{ takes on a value in the range } x_l, x_u.$$

More generally, we write $P(A)$ as the probability that an event A will occur (e.g., $A = \{x_l \leq X \leq x_u\}$). It follows from the definition that if X is a random variable in the range $[L, U]$ then $P(L \leq X \leq U) = 1$. Similarly, if $Y \in [L, U]$ then $P(L \leq X \leq Y) = 1 - P(Y \leq X \leq U)$.

We characterize a random variable in terms of the *probability density function* (pdf) $p(x)$. The density function is defined so that its integral over an interval gives the probability that the random variable takes its value in that interval:

$$P(x_l \leq X \leq x_u) = \int_{x_l}^{x_u} p(x)dx. \tag{4.2}$$

It is also possible to compute $p(x)$ given the distribution P as long as the distribution is suitably smooth:

$$p(x) = \left.\frac{\partial P(x_l \leq x \leq x_u)}{\partial x_u}\right|_{\substack{x_l \text{ fixed}, \\ x_u = x,}} \qquad x > x_l.$$

We will sometimes write $p_X(x)$ when we wish to make explicit that the pdf is associated with the random variable X. Note that we use capital letters to refer to a random variable and lower case letters to refer to a specific value.

Definition 4.4 (Uniform distribution)**.** The *uniform distribution* on an interval $[L, U]$ assigns equal probability to any number in the interval. Its pdf is given by

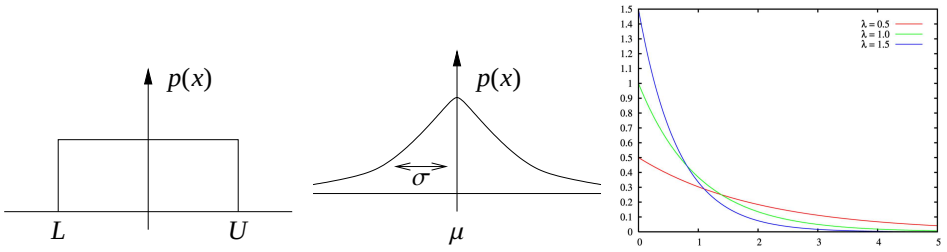
$$p(x) = \frac{1}{U - L}. \tag{4.3}$$

The uniform distribution is illustrated in Figure 4.2a.

Definition 4.5 (Gaussian distribution)**.** The *Gaussian distribution* (also called a *normal distribution*) has a pdf of the form

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}. \tag{4.4}$$

The parameter μ is called the *mean* of the distribution and σ is called the *standard deviation* of the distribution. Figure 4.2b shows a graphical representation a Gaussian pdf.



(a) Uniform distribution (b) Gaussian distribution (c) Exponentialdistribution

Figure 4.2: Probability density function (pdf) for uniform, Gaussian and exponential distributions.

Definition 4.6 (Exponential distribution)**.** The exponential distribution is defined for positive numbers and has a pdf of the form

$$p(x) = \lambda e^{-\lambda x}, \qquad x > 0$$

where λ is a parameter defining the distribution. A plot of the pdf for an exponential distribution is shown in Figure 4.2c. The exponential distribution can be shown to describe the amount of time between two events in a Poisson process.

Properties of random variables

We now define a number of properties of collections of random variables. We focus on the continuous random variable case, but unless noted otherwise these concepts can all be defined similarly for discrete random variables (using the probability mass function in place of the probability density function).

If two random variables are related, we can talk about their *joint probability distribution*: $P_{X,Y}(A, B)$ is the probability that both event A occurs for X and B occurs for Y. This is sometimes written as $P(A \cap B)$, where we abuse notation by implicitly assuming that A is associated with X and B with Y. For continuous random variables, the joint probability distribution can be characterized in terms of a *joint probability density function*

$$P(x_l \le X \le x_u, y_l \le Y \le y_u) = \int_{y_l}^{y_u} \int_{x_l}^{x_u} p(x,y)\, dx dy. \tag{4.5}$$

The joint pdf thus describes the relationship between X and Y, and for sufficiently smooth distributions we have

$$p(x,y) = \left. \frac{\partial^2 P(x_l \le X \le x_u, y_l \le Y \le y_u)}{\partial x_u \partial y_u} \right|_{\substack{x_l, y_l \text{ fixed}, \\ x_u = x, y_u = y,}} \qquad \begin{array}{l} x > x_l, \\ y > y_l. \end{array}$$

We say that X and Y are *independent* if $p(x,y) = p(x)p(y)$, which implies that $P_{X,Y}(A,B) = P_X(A)P_Y(B)$ for events A associated with X and B associated with Y. Equivalently, $P(A \cap B) = P(A)P(B)$ if A and B are independent.

The *conditional probability* for an event A given that an event B has occurred, written as $P(A \mid B)$, is given by

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}. \tag{4.6}$$

If the events A and B are independent, then $P(A \mid B) = P(A)$. Note that the individual, joint and conditional probability distributions are all different, so we should really write $P_{X,Y}(A \cap B)$, $P_{X|Y}(A \mid B)$ and $P_Y(B)$.

If X is dependent on Y then Y is also dependent on X. *Bayes' theorem* relates the conditional and individual probabilities:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}, \qquad P(B) \neq 0. \tag{4.7}$$

Bayes' theorem gives the conditional probability of event A on event B given the inverse relationship (B given A). It can be used in situations in which we wish to evaluate a hypothesis H given data D when we have some model for how likely the data is given the hypothesis, along with the unconditioned probabilities for both the hypothesis and the data.

The analog of the probability density function for conditional probability is the *conditional probability density function* $p(x \mid y)$

$$p(x \mid y) = \begin{cases} \dfrac{p(x,y)}{p(y)} & 0 < p(y) < \infty \\ 0 & \text{otherwise.} \end{cases} \tag{4.8}$$

It follows that

$$p(x,y) = p(x \mid y)p(y) \tag{4.9}$$

and

$$\begin{aligned} P(x_l \leq X \leq x_u \mid y) &:= P(x_l \leq X \leq x_u \mid Y = y) \\ &= \int_{x_l}^{x_u} p(x \mid y)dx = \frac{\int_{x_l}^{x_u} p(x,y)dx}{p(y)}. \end{aligned} \tag{4.10}$$

If X and Y are independent than $p(x \mid y) = p(x)$ and $p(y \mid x) = p(y)$. Note that $p(x,y)$ and $p(x \mid y)$ are different density functions, though they are related through equation (4.9). If X and Y are related with joint probability density function $p(x,y)$ and conditional probability density function $p(x \mid y)$ then

$$p(x) = \int_{-\infty}^{\infty} p(x,y)dy = \int_{-\infty}^{\infty} p(x \mid y)p(y)dy.$$

Example 4.1 (Conditional probability for sum)**.** Consider three random variables X, Y and Z related by the expression

$$Z = X + Y.$$

In other words, the value of the random variable Z is given by choosing values from two random variables X and Y and adding them. We assume that X and Y are independent Gaussian random variables with mean μ_1 and μ_2 and standard deviation $\sigma = 1$ (the same for both variables).

Clearly the random variable Z is not independent of X (or Y) since if we know the values of X then it provides information about the likely value of Z. To see this, we compute the joint probability between Z and X. Let

$$A = \{x_l \le x \le x_u\}, \qquad B = \{z_l \le z \le z_u\}.$$

The joint probability of both events A and B occurring is given by

$$\begin{aligned} P_{X,Z}(A \cap B) &= P(x_l \le x \le x_u, z_l \le x + y \le z_u) \\ &= P(x_l \le x \le x_u, z_l - x \le y \le z_u - x). \end{aligned}$$

We can compute this probability by using the probability density functions for X and Y:

$$\begin{aligned} P(A \cap B) &= \int_{x_l}^{x_u} \Big(\int_{z_l - x}^{z_u - x} p_Y(y) dy \Big) p_X(x) dx \\ &= \int_{x_l}^{x_u} \int_{z_l}^{z_u} p_Y(z - x) p_X(x) dz dx =: \int_{z_l}^{z_u} \int_{x_l}^{x_u} p_{Z,X}(z, x) dx dz. \end{aligned}$$

Using Gaussians for X and Y we have

$$\begin{aligned} p_{Z,X}(z, x) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z - x - \mu_Y)^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x - \mu_X)^2} \\ &= \frac{1}{2\pi} e^{-\frac{1}{2}((z - x - \mu_Y)^2 + (x - \mu_X)^2)}. \end{aligned}$$

A similar expression holds for $p_{Z,Y}$. ∇

Given a random variable X, we can define various standard measures of the distribution. The *expectation* or *mean* of a random variable is defined as

$$\mathbb{E}\{X\} = \langle X \rangle = \int_{-\infty}^{\infty} x\, p(x)\, dx,$$

and the *mean square* of a random variable is

$$\mathbb{E}\{X^2\} = \langle X^2 \rangle = \int_{-\infty}^{\infty} x^2\, p(x)\, dx.$$

If we let μ represent the expectation (or mean) of X then we define the *variance* of X as

$$\mathbb{E}\{(X-\mu)^2\} = \langle (X - \langle X \rangle)^2 \rangle = \int_{-\infty}^{\infty} (x-\mu)^2 \, p(x)\, dx.$$

We will often write the variance as σ^2. As the notation indicates, if we have a Gaussian random variable with mean μ and (stationary) standard deviation σ, then the expectation and variance as computed above return μ and σ^2.

Example 4.2 (Exponential distribution)**.** The exponential distribution has mean and variance given by

$$\mu = \frac{1}{\lambda}, \qquad \sigma^2 = \frac{1}{\lambda^2}.$$

The exponential distribution can be shown to describe the amount of time between two events in a Poisson process. ∇

Several useful properties follow from the definitions.

Proposition 4.1 (Properties of random variables)**.**

1. *If X is a random variable with mean μ and variance σ^2, then αX is random variable with mean αX and variance $\alpha^2\sigma^2$.*

2. *If X and Y are two random variables, then $\mathbb{E}\{\alpha X + \beta Y\} = \alpha\mathbb{E}\{X\} + \beta\mathbb{E}\{Y\}$.*

3. *If X and Y are Gaussian random variables with means μ_X, μ_Y and variances σ_X^2, σ_Y^2,*

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}, \qquad p(y) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2},$$

then $X+Y$ is a Gaussian random variable with mean $\mu_Z = \mu_X + \mu_Y$ and variance $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$,

$$p(x+y) = \frac{1}{\sqrt{2\pi\sigma_Z^2}} e^{-\frac{1}{2}\left(\frac{x+y-\mu_Z}{\sigma_Z}\right)^2}.$$

Proof. The first property follows from the definition of mean and variance:

$$\mathbb{E}\{\alpha X\} = \int_{-\infty}^{\infty} \alpha x \, p(x)\, dx = \alpha \int_{-\infty}^{\infty} \alpha x \, p(x)\, dx = \alpha\mathbb{E}\{X\}$$
$$\mathbb{E}\{(\alpha X)^2\} = \int_{-\infty}^{\infty} (\alpha x)^2 \, p(x)\, dx = \alpha^2 \int_{-\infty}^{\infty} x^2 \, p(x)\, dx = \alpha^2\mathbb{E}\{X^2\}.$$

The second property follows similarly, remembering that we must take the expectation using the joint distribution (since we are evaluating a function of two random variables):

$$\begin{aligned}\mathbb{E}\{\alpha X+\beta Y\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\alpha x+\beta y)\,p_{X,Y}(x,y)\,dxdy \\ &= \alpha\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x\,p_{X,Y}(x,y)\,dxdy+\beta\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}y\,p_{X,Y}(x,y)\,dxdy \\ &= \alpha\int_{-\infty}^{\infty}x\,p_X(x)\,dx+\beta\int_{-\infty}^{\infty}y\,p_Y(y)\,dy=\alpha\mathbb{E}\{X\}+\beta\mathbb{E}\{Y\}.\end{aligned}$$

The third item is left as an exercise. □

Introduction to random processes

A *random process* is a collection of time-indexed random variables. Formally, we consider a random process X to be a joint mapping of sample and a time to a state: $X:\Omega\times\mathcal{T}\to S$, where $\mathcal{T}$ is an appropriate time set. We view this mapping as a generalized random variable: a sample corresponds to choosing an entire function of time. Of course, we can always fix the time and interpret $X(\omega,t)$ as a regular random variable, with $X(\omega,t')$ representing a different random variable if $t\neq t'$. Our description of random processes will consist of describing how the random variable at a time t relates to the value of the random variable at an earlier time s. To build up some intuition about random processes, we will begin with the discrete time case, where the calculations are a bit more straightforward, and then proceed to the continuous time case.

A *discrete-time random process* is a stochastic system characterized by the *evolution* of a sequence of random variables $X[k]$, where k is an integer. As an example, consider a discrete-time linear system with dynamics

$$X[k+1]=AX[k]+BU[k]+FW[k],\qquad Y[k]\quad=CX[k]+V[k]. \tag{4.11}$$

As in AM08, $X\in\mathbb{R}^n$ represents the state of the system, $U\in\mathbb{R}^p$ is the vector of inputs and $Y\in\mathbb{R}^q$ is the vector of outputs. The (possibly vector-valued) signal W represents disturbances to the process dynamics and V represents noise in the measurements. To try to fix the basic ideas, we will take $u=0$, $n=1$ (single state) and $F=1$ for now.

We wish to describe the evolution of the dynamics when the disturbances and noise are not given as deterministic signals, but rather are chosen from some probability distribution. Thus we will let $W[k]$ be a collection of random variables where the values at each instant k are chosen from a probability distribution with pdf $p_{W,k}$. As the notation indicates, the distributions might depend on the time instant k, although the most common case is to have a *stationary* distribution in which the distributions are independent of k (defined more formally below).

In addition to stationarity, we will often also assume that distribution of values of W at time k is independent of the values of W at time l if $k \neq l$. In other words, $W[k]$ and $W[l]$ are two separate random variables that are independent of each other. We say that the corresponding random process is *uncorrelated* (also defined more formally below). As a consequence of our independence assumption, we have that

$$\mathbb{E}\{W[k]W[l]\} = \mathbb{E}\{W^2[k]\}\delta(k-l) = \begin{cases} \mathbb{E}\{W^2[k]\} & k = l \\ 0 & k \neq l. \end{cases}$$

In the case that $W[k]$ is a Gaussian with mean zero and (stationary) standard deviation σ, then $\mathbb{E}\{W[k]W[l]\} = \sigma^2\,\delta(k-l)$.

We next wish to describe the evolution of the state x in equation (4.11) in the case when W is a random variable. In order to do this, we describe the state x as a sequence of random variables $X[k]$, $k = 1, \cdots, N$. Looking back at equation (4.11), we see that even if $W[k]$ is an uncorrelated sequence of random variables, then the states $X[k]$ are not uncorrelated since

$$X[k+1] = AX[k] + FW[k],$$

and hence the probability distribution for X at time $k+1$ depends on the value of X at time k (as well as the value of W at time k), similar to the situation in Example 4.1.

Since each $X[k]$ is a random variable, we can define the mean and variance as $\mu[k]$ and $\sigma^2[k]$ using the previous definitions at each time k:

$$\mu[k] := \mathbb{E}\{X[k]\} = \int_{-\infty}^{\infty} x\,p(x,k)\,dx,$$
$$\sigma^2[k] := \mathbb{E}\{(X[k]-\mu[k])^2\} = \int_{-\infty}^{\infty} (x-\mu[k])^2\,p(x,k)\,dx.$$

To capture the relationship between the current state and the future state, we define the *correlation function* for a random process as

$$\rho(k_1,k_2) := \mathbb{E}\{X[k_1]X[k_2]\} = \int_{-\infty}^{\infty} x_1 x_2\,p(x_1,x_2;k_1,k_2)\,dx_1 dx_2$$

The function $p(x_i,x_j;k_1,k_2)$ is the *joint probability density function*, which depends on the times k_1 and k_2. A process is *stationary* if $p(x,k+d) = p(x,d)$ for all k, $p(x_i,x_j;k_1+d,k_2+d) = p(x_i,x_j;k_1,k_2)$, etc. In this case we can write $p(x_i,x_j;d)$ for the joint probability distribution. We will almost always restrict to this case. Similarly, we will write $p(k_1,k_2)$ as $p(d) = p(k,k+d)$.

We can compute the correlation function by explicitly computing the joint pdf (see Example 4.1) or by directly computing the expectation. Suppose that we take

a random process of the form (4.11) with $x[0] = 0$ and W having zero mean and standard deviation σ. The correlation function is given by

$$\begin{aligned}\mathbb{E}\{X[k_1]X[k_2]\} &= E\Big\{(\sum_{i=0}^{k_1-1} A^{k_1-i}BW[i])(\sum_{j=0}^{k_2-1} A^{k_2-j}BW[j])\Big\} \\ &= E\Big\{\sum_{i=0}^{k_1-1}\sum_{j=0}^{k_2-1} A^{k_1-i}BW[i]W[j]BA^{k_2-j}\Big\}.\end{aligned}$$

We can now use the linearity of the expectation operator to pull this inside the summations:

$$\begin{aligned}\mathbb{E}\{X[k_1]X[k_2]\} &= \sum_{i=0}^{k_1-1}\sum_{j=0}^{k_2-1} A^{k_1-i}B\mathbb{E}\{W[i]W[j]\}BA^{k_2-j} \\ &= \sum_{i=0}^{k_1-1}\sum_{j=0}^{k_2-1} A^{k_1-i}B\sigma^2\delta(i-j)BA^{k_2-j} \\ &= \sum_{i=0}^{k_1-1} A^{k_1-i}B\sigma^2BA^{k_2-i}.\end{aligned}$$

Note that the correlation function depends on k_1 and k_2.

We can see the dependence of the correlation function on the time more clearly by letting $d = k_2 - k_1$ and writing

$$\begin{aligned}\rho(k,k+d) = \mathbb{E}\{X[k]X[k+d]\} &= \sum_{i=0}^{k_1-1} A^{k-i}B\sigma^2BA^{d+k-i} \\ &= \sum_{j=1}^{k} A^jB\sigma^2BA^{j+d} = \Big(\sum_{j=1}^{k} A^jB\sigma^2BA^j\Big)A^d.\end{aligned}$$

In particular, if the discrete time system is stable then $|A| < 1$ and the correlation function decays as we take points that are further departed in time (d large). Furthermore, if we let $k \to \infty$ (i.e., look at the steady state solution) then the correlation function only depends on d (assuming the sum converges) and hence the steady state random process is stationary.

In our derivation so far, we have assumed that $X[k+1]$ only depends on the value of the state at time k (this was implicit in our use of equation (4.11) and the assumption that $W[k]$ is independent of X). This particular assumption is known as the *Markov property* for a random process: a Markovian process is one in which the distribution of possible values of the state at time k depends only on the values of the state at the prior time and not earlier. Written more formally, we say that a discrete random process is Markovian if

$$p_{X,k}(x \mid X[k-1], X[k-2], \ldots, X[0]) = p_{X,k}(x \mid X[k-1]).$$

Markov processes are roughly equivalent to state space dynamical systems, where the future evolution of the system can be completely characterized in terms of the current value of the state (and not it history of values prior to that).

Continuous time random processes

We now consider the case where our time index is no longer discrete, but instead varies continuously. A fully rigorous derivation requires careful use of measure theory and is beyond the scope of this text, so we focus here on the concepts that will be useful for modeling and analysis of important physical properties.

A *continuous-time random process* is a stochastic system characterized by the evolution of a random variable $X(t)$, $t \in [0, T]$. We are interested in understanding how the (random) state of the system is related at separate times. The process is defined in terms of the "correlation" of $X(t_1)$ with $X(t_2)$. We assume, as above, that the process is described by continuous random variables, but the discrete state case (with time still modeled as a real variable) can be handled in a similar fashion.

We call $X(t) \in \mathbb{R}^n$ the *state* of the random process at time t. For the case $n > 1$, we have a vector of random processes:

$$X(t) = \begin{pmatrix} X_1(t) \\ \vdots \\ X_n(t) \end{pmatrix}$$

We can characterize the state in terms of a (vector-valued) time-varying pdf,

$$P(x_l \le X_i(t) \le x_u) = \int_{x_l}^{x_u} p_{X_i}(x; t) dx.$$

Note that the state of a random process is not enough to determine the next state (otherwise it would be a deterministic process). We typically omit indexing of the individual states unless the meaning is not clear from context.

We can characterize the dynamics of a random process by its statistical characteristics, written in terms of joint probability density functions:

$$P(x_{1l} \le X_i(t_1) \le x_{1u}, x_{2l} \le X_j(t_2) \le x_{2u}) = \int_{x_{2l}}^{x_{2u}} \int_{x_{1l}}^{x_{1u}} p_{X_i,Y_i}(x_1, x_2; t_1, t_2)\, dx_1 dx_2$$

The function $p(x_i, x_j; t_1, t_2)$ is called a *joint probability density function* and depends both on the individual states that are being compared and the time instants over which they are compared. Note that if $i = j$, then p_{X_i,X_i} describes how X_i at time t_1 is related to X_i at time t_2.

In general, the distributions used to describe a random process depend on the specific time or times that we evaluate the random variables. However, in some

cases the relationship only depends on the difference in time and not the absolute times (similar to the notion of time invariance in deterministic systems, as described in AM08). A process is *stationary* if $p(x, t+\tau) = p(x,t)$ for all τ, $p(x_i, x_j; t_1 + \tau, t_2 + \tau) = p(x_i, x_j; t_1, t_2)$, etc. In this case we can write $p(x_i, x_j; \tau)$ for the joint probability distribution. Stationary distributions roughly correspond to the steady state properties of a random process and we will often restrict our attention to this case.

In looking at biomolecular systems, we are going to be interested in random processes in which the changes in the state occur when a random event occurs (such as a molecular reaction or binding event). In this case, it is natural to describe the state of the system in terms of a set of times $t_0 < t_1 < t_2 < \cdots < t_n$ and $X(t_i)$ is the random variable that corresponds to the possible states of the system at time t_i. Note that time time instants do not have to be uniformly spaced and most often (for biomolecular systems) they will not be. All of the definitions above carry through, and the process can now be described by a probability distribution of the form

$$P\Big(X(t_i) \in [x_i, x_i + dx_i], i = 1, \ldots, n\Big) = \int \cdots \int p(x_n, x_{n-1}, \ldots, x_0; t_n, t_{n-1}, \ldots, t_0) dx_n\, dx_{n-1}\, dx_1,$$

where dx_i are taken as infinitesimal quantities.

An important class of stochastic systems is those for which the next state of the system depends only on the current state of the system and not the history of the process. Suppose that

$$P\Big(X(t_n) \in [x_n, x_n + dx_n] \mid X(t_i) \in [x_i, x_i + dx_i], i = 1, \ldots, n-1\Big) = P\Big(X(t_n) \in [x_n, x_n + dx_n] \mid X(t_{n-1}) \in [x_{n-1}, x_{n-1} + dx_{n-1}]\Big). \quad (4.12)$$

That is, the probability of being in a given state at time t_n depends *only* on the state that we were in at the previous time instant t_{n-1} and not the entire history of states prior to t_{n-1}. A stochastic process that satisfies this property is called a *Markov process*.

In practice we do not usually specify random processes via the joint probability distribution $p(x_i, x_j; t_1, t_2)$ but instead describe them in terms of a *propogater function*. Let $X(t)$ be a Markov process and define the Markov propogater as

$$\Xi(dt; x, t) = X(t + dt) - X(t), \text{ given } X(t) = x.$$

The propogater function describes how the random variable at time t is related to the random variable at time $t + dt$. Since both $X(t + dt)$ and $X(t)$ are random variables, $\Xi(dt; x, t)$ is also a random variable and hence it can be described by its density function, which we denote as $\Pi(\xi, x; dt, t)$:

$$P(x \le X(t + dt) \le x + \xi) = \int_x^{x+\xi} \Pi(dx, x; dt, t)\, dx.$$

The previous definitions for mean, variance and correlation can be extended to the continuous time, vector-valued case by indexing the individual states:

$$E\{X(t)\} = \begin{pmatrix} E\{X_1(t)\} \\ \vdots \\ E\{X_n(t)\} \end{pmatrix} =: \mu(t)$$

$$E\{(X(t)-\mu(t))(X(t)-\mu(t))^T\} = \begin{pmatrix} E\{X_1(t)X_1(t)\} & \dots & E\{X_1(t)X_n(t)\} \\ & \ddots & \vdots \\ & & E\{X_n(t)X_n(t)\} \end{pmatrix} =: \Sigma(t)$$

$$E\{X(t)X^T(s)\} = \begin{pmatrix} E\{X_1(t)X_1(s)\} & \dots & E\{X_1(t)X_n(s)\} \\ & \ddots & \vdots \\ & & E\{X_n(t)X_n(s)\} \end{pmatrix} =: R(t,s)$$

Note that the random variables and their statistical properties are all indexed by the time t (and s). The matrix $R(t,s)$ is called the *correlation matrix* for $X(t) \in \mathbb{R}^n$. If $t = s$ then $R(t,t)$ describes how the elements of x are correlated at time t (with each other) and in the case that the processes have zero mean, $R(t,t) = \Sigma(t)$. The elements on the diagonal of $\Sigma(t)$ are the variances of the corresponding scalar variables. A random process is uncorrelated if $R(t,s) = 0$ for all $t \neq s$. This implies that $X(t)$ and $X(s)$ are independent random events and is equivalent to $p_{X,Y}(x,y) = p_X(x)p_Y(y)$.

If a random process is stationary, then it can be shown that $R(t+\tau, s+\tau) = R(t,s)$ and it follows that the correlation matrix depends only on $t-s$. In this case we will often write $R(t,s) = R(s-t)$ or simple $R(\tau)$ where τ is the correlation time. The correlation matrix in this case is simply $R(0)$.

In the case where X is also scalar random process, the correlation matrix is also a scalar and we will write $\rho(\tau)$, which we refer to as the (scalar) correlation function. Furthermore, for stationary scalar random processes, the correlation function depends only on the absolute value of the correlation function, so $\rho(\tau) = \rho(-\tau) = \rho(|\tau|)$. This property also holds for the diagonal entries of the correlation matrix since $R_{ii}(s,t) = R_{ii}(t,s)$ from the definition.

Definition 4.7 (Ornstein-Uhlenbeck process)**.** Consider a scalar random process defined by a Gaussian pdf with $\mu = 0$,

$$p(x,t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{x^2}{\sigma^2}},$$

and a correlation function given by

$$\rho(t_1,t_2) = \frac{Q}{2\omega_0} e^{-\omega_0|t_2-t_1|}.$$

The correlation function is illustrated in Figure 4.3. This process is known as an *Ornstein-Uhlenbeck process* and it is a stationary process.

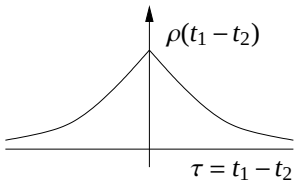


Figure 4.3: Correlation function for a first-order Markov process.

Note on terminology. The terminology and notation for covariance and correlation varies between disciplines. The term covariance is often used to refer to both the relationship between different variables X and Y and the relationship between a single variable at different times, $X(t)$ and $X(s)$. The term "cross-covariance" is used to refer to the covariance between two random vectors X and Y, to distinguish this from the covariance of the elements of X with each other. The term "cross-correlation" is sometimes also used. Finally, the term "correlation coefficient" refers to the normalized correlation $\bar{\rho}(t,s) = \mathbb{E}\{X(t)X(s)\}/\mathbb{E}\{X(t)X(t)\}$..

`MATLAB` has a number of functions to implement covariance and correlation, which mostly match the terminology here:

- `cov(X)` - this returns the variance of the vector `X` that represents samples of a given random variable or the covariance of the columns of a matrix X where the rows represent observations.
- `cov(X, Y)` - equivalent to `cov([X(:), Y(:)])`. Computes the covariance between the columns of X and Y, where the rows are observations.
- `xcorr(X, Y)` - the "cross-correlation" between two random sequences. If these sequences came from a random process, this is correlation function $\rho(t)$.
- `xcov(X, Y)` - this returns the "cross-covariance", which `MATLAB` defines as the "mean-removed cross-correlation".

The `MATLAB` help pages give the exact formulas used for each, so the main point here is to be careful to make sure you know what you really want.

We will also make use of a special type of random process referred to as "white noise". A *white noise process* $X(t)$ satisfies $E\{X(t)\} = 0$ and $R(t,s) = W\delta(s-t)$, where $\delta(\tau)$ is the impulse function and W is called the *noise intensity*. White noise is an idealized process, similar to the impulse function or Heaviside (step) function in deterministic systems. In particular, we note that $\rho(0) = E\{X^2(t)\} = \infty$, so the covariance is infinite and we never see this signal in practice. However, like the step function, it is very useful for characterizing the responds of a linear system, as described in the following proposition. It can be shown that the integral of a white noise process is a Wiener process, and so often white noise is described as the derivative of a Wiener process.

Discrete-state random processes

There are a number of specialized discrete random processes that are relevant for biochemical systems. In this section we give a brief introduction to these processes.

A *birth-death* process is one in which the states of the process represent integer-value counts of different species populations and the transitions between states are restricted to either incrementing (birth) or decrementing (death) a given species. This type of model is often used to represent chemical reactions such as the production and degradation of proteins.

Example 4.3 (Protein production)**.** ∇

A more general type of discrete random process is a *Markov chain*. In a Markov chain, evolution of the discrete states occurs by execution of allowable transitions between two states. Each transition has a specified probability, which is used to determine whether a system will transition from its current state into a different state (corresponding to an allowable transition). An important property, called the *Markov property*, is that the transition probability only depends on the value of the current state, not the previous values of the state.

We define a Markov chain by giving the set of transition probabilities

$$q_{ij}(t,\tau) = P(X(t+\tau) = s_j | X(t) = s_i),$$

where $s_i, s_j \in S$, t is the current time and τ is the time interval over which we are interested. If $q_{ij}(t,\tau) \neq 0$ for some $\tau \neq 0$ then we say that the transition is allowable at time t. If q_{ij} is independent of t then we say that the process is *stationary* and we omit the argument t. In the special case that we are only interested in a fixed τ (i.e., we are using a discrete-time model) then we omit this argument as well.

It is generally difficult to describe the probability of being in a particular state in a Markov process at a given time. Instead, we often resort to describing the steady state distributions, assuming that they exist. For a stationary Markov chain, we can look at the equilibrium distributions, which are those distributions π that satisfy

$$\pi_i = q_{ij}(\tau)\pi_j, \quad \text{for all } i, j.$$

Example 4.4 (Protein expression)**.** ∇

4.2 Stochastic Modeling of Biochemical Systems

Chemical reactions in the cell can be modeled as a collection of stochastic events corresponding to chemical reactions between species, including binding and unbinding of molecules (such as RNA polymerase and DNA), conversion of one set of species into another, and enzymatically controlled covalent modifications such as phosphorylation. In this section we will briefly survey some of the different

representations that can be used for stochastic models of biochemical systems, following the material in the textbooks by Phillips *et al.* [35], Gillespie [17] and Van Kampen [26].

Statistical physics

At the core of many of the reactions and multi-molecular interactions that take place inside of cells is the chemical physics associated with binding between two molecules. One way to capture some of the properties of these interactions is through the use of statistical mechanics and thermodynamics.

As described briefly already in Chapter 2, 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.

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. Consider a setting in which our system is contained within a reservoir. The total (conserved) energy is given by E_{tot} and we let E_r represent the energy in the reservoir. Let $E_s^{(1)}$ and $E_s^{(2)}$ represent two different energy levels for the system of interest and let $W_r(E_r)$ be the number of possible microstates of the reservoir with energy E_r. The laws of statistical mechanics state that the ratio of probabilities of being at the energy levels $E_s^{(1)}$ and $E_s^{(2)}$ is given by the ratio of number of possible states of the reservoir:

$$\frac{P(E_s^{(1)})}{P(E_s^{(2)})} = \frac{W_r(E_{\text{tot}} - E_s^{(1)})}{W_r(E_{\text{tot}} - E_s^{(2)})}. \tag{4.13}$$

Defining the entropy of the system as $S = k_B \ln W$, we can rewrite equation (4.13) as

$$\frac{W_r(E_{\text{tot}} - E_s^{(1)})}{W_r(E_{\text{tot}} - E_s^{(2)})} = \frac{e^{S_r(E_{\text{tot}} - E_s^{(1)})/k_B}}{e^{S_r(E_{\text{tot}} - E_s^{(2)})/k_B}}.$$

We now approximate $S_r(E_{\text{tot}} - E_s)$ in a Taylor series expansion around E_{tot}, under the assumption that $E_r \gg E_s$:

$$S_r(E_{\text{tot}} - E_s) \approx S_r(E_{\text{tot}}) - \frac{\partial S_r}{\partial E} E_s.$$

From the properties of thermodynamics, if we hold the volume and number of molecules constant, then we can define the temperature as

$$\left.\frac{\partial S}{\partial E}\right|_{V,N} = \frac{1}{T}$$

and we obtain

$$\frac{P(E_s^{(1)})}{P(E_s^{(2)})} = \frac{e^{-E_s^{(1)}/k_BT}}{e^{-E_s^{(2)}/k_BT}}.$$

This implies that

$$P(E_s^{(q)}) \propto e^{-E_s^{(q)}/(k_BT)}$$

and hence the probability of being in a microstate q is given by

$$P(q) = \frac{1}{Z} e^{-E_q/(k_BT)}, \tag{4.14}$$

where we have written E_q for the energy of the microstate and Z is a normalizing factor, known as the *partition function*, defined by

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

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.

In order to determine the energy levels associated with different microstates, we will often make use of the *free energy* of the system. Consider an elementary reaction $\mathrm{A} + \mathrm{B} \rightleftharpoons \mathrm{AB}$. Let E be the energy of the system, taken to be operating at pressure P in a volume V. The *enthalpy* of the system is defined as $H = E + PV$ and the *Gibbs free energy* is defined as $G = H - TS$ where T is the temperature of the system and S is its entropy (defined above). The change in bond energy due to the reaction is given by

$$\Delta H = \Delta G + T\Delta S,$$

where the Δ represents the change in the respective quantity. $-\Delta H$ represents the amount of heat that is absorbed from the reservoir, which then affects the entropy of the reservoir.

The resulting formula for the probability of being in a microstate q is given by

$$P(q) = \frac{1}{Z} e^{-\Delta G/k_BT}.$$

Example 4.5 (Ligand-receptor binding)**.** To illustrate how these ideas can be applied in a cellular setting, consider the problem of determining the probability that a ligand binds to a receptor protein, as illustrated in Figure 4.4. We model the system by breaking up the cell into Ω different locations, each of the size of a ligand molecule, and keeping track of the locations of the L ligand molecules. The microstates of the system consist of all possible locations of the ligand molecules, including those in which one of the ligand molecules is bound to the receptor molecule.

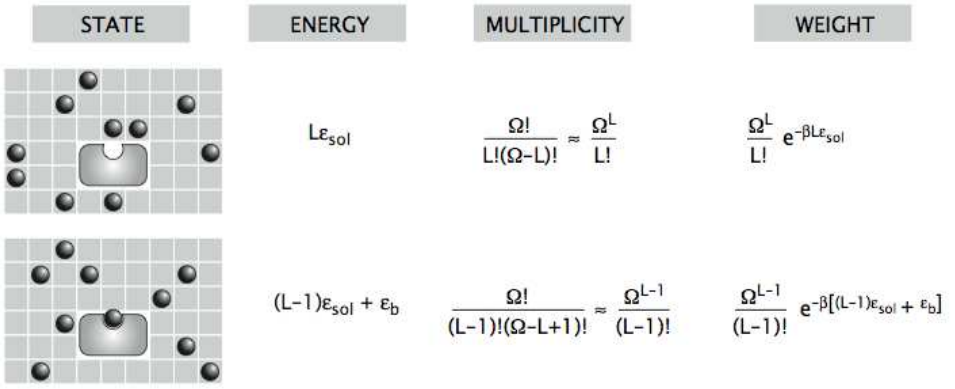


Figure 4.4: Statistical physics description of ligand-receptor binding. The cell is modeled as a compartment with Ω sites, one of which contains a receptor protein. Ligand molecules can occupy any of the sites (first column) and we can compute the Gibbs free energy associated with each configuration (second column). The first row represents all possible microstates in which the receptor protein is not bound, while the second represents all configurations in which one of the ligands binds to the receptor. By accounting for the multiplicity of each microstate (third column), we can compute the weight of the given collection of microstates (fourth column). Figure from Phillips, Kondev and Theriot [35].

To compute the probability that the ligand is bound to the receptor, we must compute the energy associated with each possible microstate and then compute the weighted sum of the microstates corresponding to the ligand being bound, normalized by the partition function. We let E_{sol} represent the free energy associated with a ligand in free solution and E_{bound} represent the free energy associated with the ligand being bound to the receptor. Thus, the energy associated with microstates in which the ligand is not bound to the receptor is given by

$$\Delta G_{\text{sol}} = LE_{\text{sol}}$$

and the energy associated with microstates in which one ligand is bound to the receptor is given by

$$\Delta G_{\text{bound}} = (L-1)E_{\text{sol}} + E_{\text{bound}}.$$

Next, we compute the number of possible ways in which each of these two situations can occur. For the unbound ligand, we have L molecules that can be in any one of Ω locations, and hence the total number of combinations is given by

$$N_{\text{sol}} = \binom{\Omega}{L} = \frac{\Omega!}{L!(\Omega-L)!} \approx \frac{\Omega^L}{L!},$$

where the final approximation is valid in the case when $L \ll \Omega$. Similarly, the number of microstates in which the ligand is bound to the receptor is

$$N_{\text{sol}} = \binom{\Omega}{L-1} = \frac{\Omega!}{(L-1)!(\Omega-L+1)!} \approx \frac{\Omega^{L-1}}{(L-1)!}.$$

Using these two counts, the partition function for the system is given by

$$Z \approx \frac{\Omega^L}{L!} e^{-\frac{LE_{\text{sol}}}{k_B T}} + \frac{\Omega^{L-1}}{(L-1)!} e^{-\frac{(L-1)E_{\text{sol}}+E_{\text{bound}}}{k_B T}}.$$

Finally, we can compute the steady state probability that the ligand is bound by computing the ratio of the weights for the desired states divided by the partition function

$$P_{\text{bound}} = \frac{1}{Z} \cdot \frac{\Omega^{L-1}}{(L-1)!} e^{-\frac{(L-1)E_{\text{sol}}+E_{\text{bound}}}{k_B T}}.$$

∇

While the previous example was carried out for the special case of a ligand molecule binding to a receptor protein, in fact this same type of computation can be used to compute the probability that a transcription factor is attached to a piece of DNA or that two freely moving molecules bind to each other. Each of these cases simply comes down to enumerating all possible microstates, computing the energy associated with each, and then computing the ratio of the sum of the weights for the desired states to the complete partition function.

Chemical Master Equation (CME)

The statistical physics model we have just considered gives a description of the *steady state* properties of the system. In many cases, it is clear that the system reaches this steady state quickly and hence we can reason about the behavior of the system just by modeling the free energy of the system. In other situations, however, we care about the transient behavior of a system or the dynamics of a system that does not have an equilibrium configuration. In these instances, we must extend our formulation to keep track of how quickly the system transitions from one microstate to another, known as the *chemical kinetics* of the system.

To model these dynamics, we return to our enumeration of all possible microstates of the system. Let $P(q,t)$ represent the probability that the system is in microstate q at a given time t. Here q can be any of the very large number of possible microstates for the system. We wish to write an explicit expression for how $P(q,t)$ varies as a function of time, from which we can study the stochastic dynamics of the system.

We begin by assuming we have a set of M reactions R_j, $j = 1,\ldots,M$, with ξ_j representing the change in state associated with reaction R_j. The *propensity function* defines the probability that a given reaction occurs in a sufficiently small time step dt:

$a_j(q,t)dt$ = Probability that reaction R_j will occur between time t and time $t+dt$ given that $X(t) = q$.

The linear dependence on dt relies on the fact that dt is chosen sufficiently small. We will typically assume that a_j does not depend on the time t and write $a_j(q)dt$ for the probability that reaction j occurs in state x.

Using the propensity function, we can compute the distribution of states at time $t+dt$ given the distribution at time t:

$$\begin{aligned} P(q,t+dt \mid q_0,t_0) &= P(q,t \mid q_0,t_0)\Big(1-\sum_{j=1}^{M} a_j(q)dt\Big)+\sum_{j=1}^{M} P(q-\xi_j \mid q_0,t_0)a_j(q-\xi_j)dt \\ &= P(q,t \mid q_0,t_0)+\sum_{j=1}^{M}\Big(a_j(q-\xi_j)P(q-\xi_j,t \mid q_0,t_0)-a_j(q)P(q,t \mid q_0,t_0)\Big)dt. \end{aligned} \tag{4.15}$$

Since dt is small, we can take the limit as $dt \to 0$ and we obtain the *chemical master equation* (CME):

$$\frac{\partial P}{\partial t}(q,t \mid q_0,t_0) = \sum_{j=1}^{M}\Big(a_j(q-\xi_j)P(q-\xi_j,t \mid q_0,t_0)-a_j(q)P(q,t \mid q_0,t_0)\Big) \tag{4.16}$$

This equation is also referred to as the *forward Kolmogorov equation* for a discrete state, continuous time random process.

We will sometimes find it convenient to use a slightly different notation in which we let ξ represent any transition in the system state (without enumerating the reactions). In this case, we write the propensity function as $a(\xi;q,t)$, which represents the incremental probability that we will transition from state q to state $q+\xi$ at time t. When the propensities are not explicitly dependent on time, we simply write $a(\xi;q)$. In this notation, the chemical master equation becomes

$$\frac{\partial P}{\partial t}(q,t \mid q_0,t_0) = \sum_{\xi}\Big(a(\xi;q-\xi_j)P(q-\xi_j,t \mid q_0,t_0)-a(\xi;q)P(q,t \mid q_0,t_0)\Big), \tag{4.17}$$

where the sum is understood to be over all allowable transitions.

Under some additional assumptions, we can rewrite the master equation in differential form as

$$\frac{d}{dt}P(q,t) = \sum_{\xi} a(\xi;q-\xi)P(q-\xi,t) - \sum_{\xi} a(\xi;q)P(q,t), \tag{4.18}$$

where we have dropped the dependence on the initial condition for notational convenience. We see that the master equation is a *linear* differential equation with state $P(q,t)$. However, it is important to note that the size of the state vector can be very large: we must keep track of the probability of every possible microstate of the system. For example, in the case of the ligand-receptor problem discussed earlier, this has a factorial number of states based on the number of possible sites in the

model. Hence, even for very simple systems, the master equation cannot typically be solved either analytically or in a numerically efficient fashion.

Despite its complexity, the master equation does capture many of the important details of the chemical physics of the system and we shall use it as our basic representation of the underlying dynamics. As we shall see, starting from this equation we can then derive a variety of alternative approximations that allow us to answer specific equations of interest.

The key element of the master equation is the propensity function $a(\xi; q, t)$, which governs the rate of transition between microstates. Although the detailed value of the propensity function can be quite complex, its functional form is often relatively simple. In particular, for a unimolecular reaction ξ of the form $\mathrm{A} \to B$, the propensity function is proportional to the number of molecules of A that are present:

$$a(\xi; q, t) = c_\xi n_A. \tag{4.19}$$

This follows from the fact that each reaction is independent and hence the likelihood of a reaction happening depends directly on the number of copies of A that are present.

Similarly, for a bimolecular reaction, we have that the likelihood of a reaction occurring is proportional to the product of the number of molecules of each type that are present (since this is the number of independent reactions that can occur). Hence, for a reaction ξ of the form $\mathrm{A} + \mathrm{B} \longrightarrow \mathrm{C}$ we have

$$a(\xi; q, t) = c_\xi n_A n_B. \tag{4.20}$$

The rigorous verification of this functional form is beyond the scope of this text, but roughly we keep track of the likelihood of a single reaction occurring between A and B and then multiply by the total number of combinations of the two molecules that can react ($n_\mathrm{A} \cdot n_\mathrm{B}$).

A special case of a bimolecular reaction occurs when A = B, so that our reaction is given by $2\,\mathrm{A} \to B$. In this case we must take into account that a molecule cannot react with itself, and so the propensity function is of the form

$$a(\xi; q, t) = c_\xi n_A (n_A - 1). \tag{4.21}$$

Although it is tempting to extend this formula to the case of more than two species being involved in a reaction, usually such reactions actually involve combinations of bimolecular reactions, e.g.:

$$\mathrm{A} + \mathrm{B} + \mathrm{C} \longrightarrow \mathrm{D} \qquad \Longrightarrow \qquad \mathrm{A} + \mathrm{B} \longrightarrow \mathrm{AB} \quad \mathrm{AB} + \mathrm{C} \longrightarrow \mathrm{D}$$

This more detailed description reflects that fact that it is extremely unlikely that three molecules will all come together at precisely the same instant, versus the much more likely possibility that two molecules will initially react, followed be a second reaction involving the third molecule.

Table 4.1: Examples of propensity functions for some common cases [18]. Here we take r_a and r_b to be the effective radii of the molecules, $m^* = m_a m_b/(m_a + m_b)$ is the reduced mass of the two molecules, Ω is the volume over which the reaction occurs, T is temperature, k_B is Boltzmann's constant and n_a, n_b are the numbers of molecules of A and B present.

Reaction type	Propensity function coefficient, c_ξ
Reaction occurs if molecules "touch"	$\Omega^{-1}\left(\frac{8k_BT}{\pi m^*}\right)^{1/2}\pi(r_a+r_b)^2$
Reaction occurs if molecules collide with energy ϵ	$\Omega^{-1}\left(\frac{8k_BT}{\pi m^*}\right)^{1/2}\pi(r_a+r_b)^2\cdot e^{-\epsilon/k_BT}$
Steady state transcription factor	$P_{\text{bound}}k_{\text{oc}}n_{\text{RNAP}}$

The propensity functions for these cases and some others are given in Table 4.1.

Example 4.6 (Transcription of mRNA)**.** Consider the production of mRNA from a single copy of DNA. We have two basic reactions that can occur: mRNA can be produced by RNA polymerase transcribing the DNA and producing an mRNA strand, or mRNA can be degraded. We represent the microstate q of the system in terms of the number of mRNA's that are present, which we write as n for ease of notation. The reactions can now be represented as $\xi = +1$, corresponding to transcription and $\xi = -1$, corresponding to degradation. We choose as our propensity functions

$$a(+1;n,t) = \alpha, \qquad a(-1;n,t) = \gamma n,$$

by which we mean that the probability of that a gene is transcribed in time dt is αdt and the probability that a transcript in time dt is $\gamma n dt$ (proportional to the number of mRNA's).

We can now write down the master equation as described above. Equation (4.15) becomes

$$\begin{aligned} P(n,t+dt) &= P(n,t)\Big(1 - \sum_{\xi=+1,-1} a(\xi;n,t)dt\Big) + \sum_{\xi=+1,-1} P(n-\xi,t)a(\xi;q-\xi)dt \\ &= P(n,t) - a(+1;n,t)P(n,t) - a(-1;n,t)P(n,t) \\ &\quad + a(+1,n-1,t)P(n-1,t) + a(-1;n+1,t)P(n+1) \\ &= P(n,t) + \alpha P(n-1,t)dt - (\alpha - \gamma n)P(n,t)dt + \gamma(n+1)P(n+1,t)dt. \end{aligned}$$

This formula holds for $n > 0$, with the $n = 0$ case satisfying

$$P(0,t+dt) = P(0,t) - \alpha P(0,t)dt + \gamma P(1,t)dt.$$

Notice that we have an infinite number of equations, since n can be any positive integer.

We can write the differential equation version of the master equation by subtracting the first term on the right hand side and dividing by dt:

$$\frac{d}{dt}P(n,t) = \alpha P(n-1,t) - (\alpha+\gamma n)P(n,t) + \gamma(n+1)P(n+1,t), \qquad n > 0$$
$$\frac{d}{dt}P(0,t) = -\alpha P(0,t)dt + \gamma P(1,t).$$

Again, this is an infinite number of differential equations, although we could take some limit N and simply declare that $P(N,t) = 0$ to yield a finite number.

One simple type of analysis that can be done on this equation without truncating it to a finite number is to look for a steady state solution to the equation. In this case, we set $\dot{P}(n,t) = 0$ and look for a constant solution $P(n,t) = p_e(n)$. This yields an algebraic set of relations

$$\begin{aligned}
0 &= -\alpha p_e(0) + \gamma p_e(1) && \Longrightarrow \quad && \alpha p_e(0) = \gamma p_e(1)\\
0 &= \alpha p_e(0) - (\alpha+\gamma)p_e(1) + 2\gamma p_e(2) && && \alpha p_e(1) = 2\gamma p_e(2)\\
0 &= \alpha p_e(1) - (\alpha+2\gamma)p_e(2) + 3\gamma p_e(3) && && \alpha p_e(1) = 3\gamma p_e(3)\\
&\vdots && && \vdots\\
& && && \alpha p(n-1) = n\gamma p(n).
\end{aligned}$$

It follows that the distribution of steady state probabilities is given by the Poisson distribution

$$p(n) = e^{\alpha/\gamma}\frac{(\alpha/\gamma)^n}{n!},$$

and the mean, variance and coefficient of variation are thus

$$\mu = \frac{\alpha}{\gamma}, \qquad \sigma^2 = \frac{\alpha}{\gamma}, \qquad CV = \frac{\mu}{\sigma} = \frac{1}{\sqrt{\mu}} = \sqrt{\frac{\gamma}{\alpha}}.$$

∇

Chemical Langevin equation (CLE)

The chemical master equation gives a complete description of the evolution of the distribution of a system, but it can often be quite cumbersome to work with directly. A number of approximations to the master equation are thus used to provide more tractable formulations of the dynamics. The first of these that we shall consider is known as the *chemical Langevin equation* (CLE).

To derive the chemical Langevin equation, we start by assuming that the number of species in the system is large and that we can therefore represent the system using a vector of real numbers X, with X_i representing the (real-valued) number of molecules in S_i. (Often X_i will be divided by the volume to give a real-valued concentration of species S_i.) In addition, we assume that we are interested in the

dynamics on time scales in which individual reactions are not important and so we can look at how the system state changes over time intervals in which many reactions occur and hence the system state evolves in a smooth fashion.

Let $X(t)$ be the state vector for the system, where we assume now that the elements of X are real-valued rather than integer valued. We make the further approximation that we can lump together multiple reactions so that instead of keeping track of the individual reactions, we can average across a number of reactions over a time τ to allow the continuous state to evolve in continuous time. The resulting dynamics can be described by a stochastic process of the form

$$X_i(t+\tau) = X(t) + \sum_{j=1}^{M} \xi_{ij} a_j(X(t))\tau + \sum_{j=1}^{M} \xi_{ij} a_j^{1/2}(X(t)) \mathcal{N}_j(0, \sqrt{\tau}),$$

where a_j are the propensity functions for the individual reactions, ξ_{ij} are the corresponding changes in the system states X_i and $\mathcal{N}_j$ are a set of independent Gaussian random variables with zero mean and variance τ.

If we assume that τ is small enough that we can use the derivative to approximate the previous equation (but still large enough that we can average over multiple reactions), then we can write

$$\frac{dX_i(t)}{dt} = \sum_{j=1}^{M} \xi_{ji} a_j(X(t)) + \sum_{j=1}^{M} \xi_{ji} a_j^{1/2}(X(t)) \Gamma_j(t) \quad =: A_i(X(t)) + \sum_{j=1}^{M} B_{ij}(X(t)) \Gamma_j(t), \tag{4.22}$$

where Γ_j are white noise processes. This equation is called the *chemical Langevin equation* (CLE).

Example 4.7 (Protein production)**.** Consider a simplified model of protein production in which mRNAs are produced by transcription and proteins by translation. We also include degradation of both mRNAs and proteins, but we do not model the detailed processes of elongation of the mRNA and polypeptide chains.

We can capture the state of the system by keeping track of the number of copies of mRNA and proteins. We further approximate this by assuming that the number of each of these is sufficiently large that we can keep track of its concentration, and hence $X = (n_m, n_p)$ where $n_m \in \mathbb{R}$ is the amount of mRNA and $n_p \in \mathbb{R}$ is the concentration of protein. Letting Ω represent the volume, the reactions that govern the dynamics of the system are given by:

$$\begin{aligned}
&R_1 : \phi \xrightarrow{\alpha} \text{mRNA} && \xi_1 = (1,0) && a_1(X) = \alpha/\Omega \\
&R_2 : \text{mRNA} \xrightarrow{\gamma} \phi && \xi_2 = (-1,0) && a_2(X) = \gamma/\Omega\, n_m \\
&R_3 : \text{mRNA} \xrightarrow{\beta} \text{mRNA} + \text{protein} && \xi_3 = (0,1) && a_3(X) = \beta/\Omega\, n_m \\
&R_4 : \text{protein} \xrightarrow{\delta} \phi && \xi_4 = (0,-1) && a_4(X) = \delta/\Omega\, n_p.
\end{aligned}$$

Substituting these expressions into equation (4.22), we obtain a stochastic differential equation of the form

$$\frac{d}{dt}\begin{pmatrix} n_m \\ n_p \end{pmatrix} = \begin{pmatrix} -\gamma/\Omega & 0 \\ \beta/\Omega & -\delta/\Omega \end{pmatrix}\begin{pmatrix} n_m \\ n_p \end{pmatrix} + \begin{pmatrix} \alpha/\Omega \\ 0 \end{pmatrix} + \begin{pmatrix} \left(\sqrt{\alpha/\Omega} + \sqrt{\gamma n_m/\Omega}\right)\Gamma_m \\ \left(\sqrt{\beta n_m/\Omega} + \sqrt{\delta n_p/\Omega}\right)\Gamma_p \end{pmatrix},$$

where Γ_m and Γ_p are independent white noise processes with unit variance. (Note that in deriving this equation we have used the fact that the sum of two independent Gaussian processes is a Gaussian process.) ∇

Fokker-Planck equations (FPE)

The chemical Langevin equation provides a stochastic ordinary differential equation that describes the evolution of the system state. A slightly different (but completely equivalent) representation of the dynamics is to model how the probability distribution $P(q,t)$ evolves in time. As in the case of the chemical Langevin equation, we will assume that the system state is continuous and write down a formula for the evolution of the density function $p(x,t)$. This formula is known as the *Fokker-Planck equations* (FPE) and is essentially an approximation on the chemical master equation.

Consider first the case of a random process in one dimension. We assume that the random process is in the same form as the previous section:

$$\frac{dX(t)}{dt} = A(X(t)) + B(X(t))\Gamma(t). \tag{4.23}$$

The function $A(X)$ is called the *drift term* and $B(X)$ is the *diffusion term*. It can be shown that the probability density function for X, $p(x,t \mid x_0,t_0)$, satisfies the partial differential equation

$$\frac{\partial p}{\partial t}(x,t \mid x_0,t_0) = -\frac{\partial}{\partial x}(A(x,t)p(x,t \mid x_0,t_0)) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(B^2(x,t)p(x,t \mid x_0,t_0)) \tag{4.24}$$

Note that here we have shifted to the probability density function since we are considering X to be a continuous state random process.

In the multivariate case, a bit more care is required. Using the chemical Langevin equation (4.22), we define

$$D_i(x,t) = \sum_{j=1}^{M} B_{ij}^2(x,t), \qquad C_{ij}(x,t) = \sum_{k=1}^{M} B_{ik}(x,t)B_{jk}(x,t),\ i < j = 1,\ldots,M.$$

The Fokker-Planck equation now becomes

$$\begin{aligned}\frac{\partial p}{\partial t}(x,t \mid x_0,t_0) = &-\sum_{i=1}^{M}\frac{\partial}{\partial x_i}(A_i(x,t)p(x,t \mid x_0,t_0)) \\ &+\frac{1}{2}\sum_{i=1}^{M}\frac{\partial}{\partial x_i}\frac{\partial^2}{\partial x^2}(D_i(x,t)p(x,t \mid x_0,t_0)) \\ &+\sum_{\substack{i,j=1\\ i<j}}^{M}\frac{\partial^2}{\partial x_i \partial x_j}(C_{ij}(x,t)p(x,t \mid x_0,t_0)).\end{aligned} \tag{4.25}$$

Linear noise approximation (LNA)

The chemical Langevin equation and the Fokker-Planck equation provide approximations to the chemical master equation. A slightly different approximation can be obtained by expanding the density function in terms of a size parameter Ω. This approximation is know as the *linear noise approximation* (LNA) or the Ω *expansion* [26].

We begin with a master equation for a continuous random variable X, which we take to be of the form

$$\frac{\partial p}{\partial t}(x,t) = \int (a_\Omega(\xi; x-\xi)p(x-\xi,t) - a_\Omega(\xi;x)p(x,t))\,d\xi,$$

where we have dropped the dependence on the initial condition for notational simplicity. As before, the propensity function $a_\Omega(\xi;x)$ represents the transition probability between a state x and a state $x+\xi$ and we assume that it is a function of a parameter Ω that represents the size of the system (typically the volume). Since we are working with continuous variables, we now have an integral in place of our previous sum.

We assume that the mean of X can be written as $\Omega\phi(t)$ where $\phi(t)$ is a continuous function of time that represents the evolution of the mean of X/Ω. To understand the fluctuations of the system about this mean, we write

$$X = \Omega\phi + \Omega^{\frac{1}{2}}Z,$$

where Z is a new variable representing the perturbations of the system about its mean. We can write the distribution for Z as

$$p_Z(z,t) = p_X(\Omega\phi(t) + \Omega^{\frac{1}{2}}z, t)$$

and it follows that the derivatives of p_Z can be written as

$$\begin{aligned}\frac{\partial^\nu p_Z}{z^\nu} &= \Omega^{\frac{1}{2}\nu}\frac{\partial^\nu p_X}{x^\nu} \\ \frac{\partial p_Z}{\partial t} &= \frac{\partial p_X}{\partial t} + \Omega\frac{d\phi}{dt}\frac{\partial p_X}{\partial x} = \frac{\partial p_X}{\partial t} + \Omega^{\frac{1}{2}}\frac{d\phi}{dt}\frac{\partial p_Z}{\partial z}.\end{aligned}$$

We further assume that the Ω dependence of the propensity function is such that

$$a_\Omega(\xi, \Omega\phi) = f(\Omega)\tilde{a}(\xi; \phi),$$

where $\tilde{a}$ is not dependent on Ω. From these relations, we can now derive the master equation for p_Z in terms of powers of Ω (derivation omitted).

The $\Omega^{1/2}$ term in the expansion turns out to yield

$$\frac{d\phi}{dt} = \int \xi a(\xi, \Omega\phi)\, d\xi, \qquad \phi(0) = \frac{X(0)}{\Omega},$$

which is precisely the equation for the mean of the concentration. It can further be shown that the terms in Ω^0 are given by

$$\frac{\partial p_Z(z,\tau)}{\partial \tau} = -\alpha_1'(\phi)\frac{\partial}{\partial z}(z p_Z(z,t)) + \frac{1}{2}\alpha_2(\phi)\frac{\partial^2 p_Z(z,t)}{\partial z^2}, \tag{4.26}$$

where

$$\alpha_\nu(x) = \int \xi^\nu \tilde{a}(\xi; x)\, d\xi, \qquad \tau = \Omega^{-1} f(\Omega) t.$$

Notice that in the case that $\phi(t) = \phi_0$, this equation becomes the Fokker-Planck equation derived previously.

Higher order approximations to this equation can also be carried out by keeping track of the expansion terms in higher order powers of Ω. In the case where Ω represents the volume of the system, the next term in the expansion is Ω^{-1} and this represents fluctuations that are on the order of a single molecule, which can usually be ignored.

Rate reaction equations (RRE)

As we already saw in Chapter 2, the reaction rate equations can be used to describe the dynamics of a chemical system in the case where there are a large number of molecules whose state can be approximated using just the concentrations of the molecules. We re-derive the results from Section 2.2 here, being more careful to point out what approximations are being made.

We start with the chemical Langevin equations (4.22), from which we can write the dynamics for the average quantity of the each species at each point in time:

$$\frac{d\langle X_i(t)\rangle}{dt} = \sum_{j=1}^{M} \xi_{ji}\langle a_j(X(t))\rangle,$$

where the second order term drops out under the assumption that the Γ_j's are independent processes. We see that the reaction rate equations follow by defining $x_i = \langle X_i\rangle/\Omega$ and *assuming* that $\langle a_j(X(t))\rangle = a_j(\langle X(t)\rangle)$. This relationship is true when a_j is linear (e.g., in the case of a unimolecular reaction), but is an approximation otherwise.

4.3 Simulation of Stochastic sections

4.4 Analysis of Stochastic Systems

4.5 Linearized Modeling and Analysis

In this section we consider the special case of linear stochastic systems that are driven by random processes.

Linear input/output response

We now consider the problem of how to compute the response of a linear system to a random process. We assume we have a linear system described in state space as

$$\dot{X} = AX + FW, \qquad Y = CX \tag{4.27}$$

Given an "input" W, which is itself a random process with mean $\mu(t)$, variance $\sigma^2(t)$ and correlation $\rho(t, t+\tau)$, what is the description of the random process Y?

Let W be a white noise process, with zero mean and noise intensity Q:

$$\rho(\tau) = Q\delta(\tau).$$

We can write the output of the system in terms of the convolution integral

$$Y(t) = \int_0^t h(t-\tau)W(\tau)\,d\tau,$$

where $h(t-\tau)$ is the impulse response for the system

$$h(t-\tau) = Ce^{A(t-\tau)}B + D\delta(t-\tau).$$

We now compute the statistics of the output, starting with the mean:

$$\begin{aligned}\mathbb{E}\{Y(t)\} &= E\{\int_0^t h(t-\eta)W(\eta)\,d\eta\} \\ &= \int_0^t h(t-\eta)E\{W(\eta)\}\,d\eta = 0.\end{aligned}$$

Note here that we have relied on the linearity of the convolution integral to pull the expectation inside the integral.

We can compute the covariance of the output by computing the correlation $\rho(\tau)$ and setting $\sigma^2 = \rho(0)$. The correlation function for y is

$$\begin{aligned}\rho_Y(t,s) = E\{Y(t)Y(s)\} &= E\{\int_0^t h(t-\eta)W(\eta)\,d\eta \cdot \int_0^s h(s-\xi)W(\xi)\,d\xi\} \\ &= E\{\int_0^t \int_0^s h(t-\eta)W(\eta)W(\xi)h(s-\xi)\,d\eta d\xi\}\end{aligned}$$

Once again linearity allows us to exchange expectation and integration

$$\begin{aligned}\rho_Y(t,s) &= \int_0^t \int_0^s h(t-\eta)E\{W(\eta)W(\xi)\}h(s-\xi)\,d\eta d\xi \\ &= \int_0^t \int_0^s h(t-\eta)Q\delta(\eta-\xi)h(s-\xi)\,d\eta d\xi \\ &= \int_0^t h(t-\eta)Qh(s-\eta)\,d\eta\end{aligned}$$

Now let $\tau = s - t$ and write

$$\begin{aligned}\rho_Y(\tau) = \rho_Y(t,t+\tau) &= \int_0^t h(t-\eta)Qh(t+\tau-\eta)\,d\eta \\ &= \int_0^t h(\xi)Qh(\xi+\tau)\,d\xi \qquad (\text{setting } \xi = t-\eta)\end{aligned}$$

Finally, we let $t \to \infty$ (steady state)

$$\lim_{t\to\infty} \rho_Y(t,t+\tau) = \bar{\rho}_Y(\tau) = \int_0^\infty h(\xi)Qh(\xi+\tau)d\xi \tag{4.28}$$

If this integral exists, then we can compute the second order statistics for the output Y.

We can provide a more explicit formula for the correlation function ρ in terms of the matrices A, F and C by expanding equation (4.28). We will consider the general case where $W \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ and use the correlation matrix $R(t,s)$ instead of the correlation function $\rho(t,s)$. Define the *state transition matrix* $\Phi(t,t_0) = e^{A(t-t_0)}$ so that the solution of system (4.27) is given by

$$x(t) = \Phi(t,t_0)x(t_0) + \int_{t_0}^t \Phi(t,\lambda)Fw(\lambda)d\lambda$$

Proposition 4.2 (Stochastic response to white noise)**.** *Let $E\{X(t_0)X^T(t_0)\} = P(t_0)$ and W be white noise with $E\{W(\lambda)W^T(\xi)\} = R_W\delta(\lambda-\xi)$. Then the correlation matrix for X is given by*

$$R_X(t,s) = P(t)\Phi^T(s,t)$$

where $P(t)$ satisfies the linear matrix differential equation

$$\dot{P}(t) = AP + PA^T + FR_WF, \qquad P(0) \quad = P_0.$$

Proof. Using the definition of the correlation matrix, we have

$$\begin{aligned}
E\{X(t)X^T(s)\} &= E\Big\{\Phi(t,0)X(0)X^T(0)\Phi^T(t,0) + \text{cross terms} \\
&\qquad + \int_0^t \Phi(t,\xi)FW(\xi)\,d\xi \int_0^s W^t(\lambda)F^T\Phi(s,\lambda)\,d\lambda\Big\} \\
&= \Phi(t,0)E\{X(0)X^T(0)\}\Phi(s,0) \\
&\qquad + \int_0^t \int_0^s \Phi(t,\xi)FE\{W(\xi)W^T(\lambda)\}F^T\Phi(s,\lambda)\,d\xi\,d\lambda \\
&= \Phi(t,0)P(0)\phi^T(s,0) + \int_0^t \Phi(t,\lambda)FR_W(\lambda)F^T\Phi(s,\lambda)\,d\lambda.
\end{aligned}$$

Now use the fact that $\Phi(s,0) = \Phi(s,t)\Phi(t,0)$ (and similar relations) to obtain

$$R_X(t,s) = P(t)\Phi^T(s,t)$$

where

$$P(t) = \Phi(t,0)P(0)\Phi^T(t,0) + \int_0^T \Phi(t,\lambda)FR_WF^T(\lambda)\Phi^T(t,\lambda)d\lambda$$

Finally, differentiate to obtain

$$\dot{P}(t) = AP + PA^T + FR_WF, \qquad P(0) \quad = P_0$$

(see Friedland for details). □

The correlation matrix for the output Y can be computing using the fact that $Y = CX$ and hence $R_Y = C^TR_XC$. We will often be interested in the steady state properties of the output, which given by the following proposition.

Proposition 4.3 (Steady state response to white noise)**.** *For a time-invariant linear system driven by white noise, the correlation matrices for the state and output converge in steady state to*

$$R_X(\tau) = R_X(t,t+\tau) = Pe^{A^T\tau}, \qquad R_Y(\tau) = CR_X(\tau)C^T$$

where P satisfies the algebraic equation

$$AP + PA^T + FR_WF^T = 0 \qquad P > 0. \tag{4.29}$$

Equation (4.29) is called the *Lyapunov equation* and can be solved in MATLAB using the function `lyap`.

Example 4.8 (First-order system)**.** Consider a scalar linear process

$$\dot{X} = -aX + W, \qquad Y = cX,$$

where W is a white, Gaussian random process with noise intensity σ^2. Using the results of Proposition 4.2, the correlation function for X is given by

$$R_X(t, t+\tau) = p(t)e^{-a\tau}$$

where $p(t) > 0$ satisfies

$$p(t) = -2ap + \sigma^2.$$

We can solve explicitly for $p(t)$ since it is a (non-homogeneous) linear differential equation:

$$p(t) = e^{-2at}p(0) + (1 - e^{-2at})\frac{\sigma^2}{2a}.$$

Finally, making use of the fact that $Y = cX$ we have

$$\rho(t, t+\tau) = c^2(e^{-2at}p(0) + (1 - e^{-2at})\frac{\sigma^2}{2a})e^{-a\tau}.$$

In steady state, the correlation function for the output becomes

$$\rho(\tau) = \frac{c^2\sigma^2}{2a}e^{-a\tau}.$$

Note correlation function has the same form as the Ornstein-Uhlenbeck process in Example 4.7 (with $Q = c^2\sigma^2$). ∇

Random Processes in the Frequency Domain

As in the case of deterministic linear systems, we can analyze a stochastic linear system either in the state space or the frequency domain. The frequency domain approach provides a very rich set of tools for modeling and analysis of interconnected systems, relying on the frequency response and transfer functions to represent the flow of signals around the system.

Given a random process $X(t)$, we can look at the frequency content of the properties of the response. In particular, if we let $\rho(\tau)$ be the correlation function for a (scalar) random process, then we define the *power spectral density function* as the Fourier transform of ρ:

$$S(\omega) = \int_{-\infty}^{\infty} \rho(\tau)e^{-j\omega\tau}\,d\tau, \qquad \rho(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)e^{j\omega\tau}\,d\tau.$$

The power spectral density provides an indication of how quickly the values of a random process can change through the frequency content: if there is high frequency content in the power spectral density, the values of the random variable can change quickly in time.

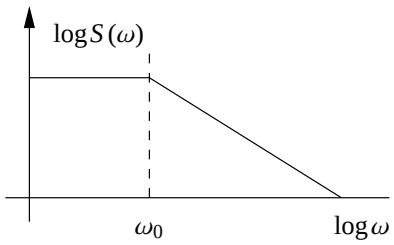


Figure 4.5: Spectral power density for a first-order Markov process.

.

Example 4.9 (First-order Markov process)**.** To illustrate the use of these measures, consider a first-order Markov process as defined in Example 4.7. The correlation function is

$$\rho(\tau) = \frac{Q}{2\omega_0} e^{-\omega_0(\tau)}.$$

The power spectral density becomes

$$\begin{aligned} S(\omega) &= \int_{-\infty}^{\infty} \frac{Q}{2\omega_0} e^{-\omega|\tau|} e^{-j\omega\tau}\, d\tau \\ &= \int_{-\infty}^{0} \frac{Q}{2\omega_0} e^{(\omega - j\omega)\tau}\, d\tau + \int_{0}^{\infty} \frac{Q}{2\omega_0} e^{(-\omega - j\omega)\tau}\, d\tau = \frac{Q}{\omega^2 + \omega_0^2}. \end{aligned}$$

We see that the power spectral density is similar to a transfer function and we can plot $S(\omega)$ as a function of ω in a manner similar to a Bode plot, as shown in Figure 4.5. Note that although $S(\omega)$ has a form similar to a transfer function, it is a real-valued function and is not defined for complex s. ∇

Using the power spectral density, we can more formally define "white noise": a *white noise process* is a zero-mean, random process with power spectral density $S(\omega) = W$ = constant for all ω. If $X(t) \in \mathbb{R}^n$ (a random vector), then $W \in \mathbb{R}^{n\times n}$. We see that a random process is white if all frequencies are equally represented in its power spectral density; this spectral property is the reason for the terminology "white". The following proposition verifies that this formal definition agrees with our previous (time domain) definition.

Proposition 4.4. *For a white noise process,*

$$\rho(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) e^{j\omega\tau}\, d\tau = W\delta(\tau),$$

where $\delta(\tau)$ is the unit impulse function.

Proof. If $\tau \neq 0$ then

$$\rho(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} W(\cos(\omega\tau) + j\sin(\omega\tau)\, d\tau = 0$$

If $\tau = 0$ then $\rho(\tau) = \infty$. Can show that

$$\rho(0) = \lim_{\epsilon \to 0} \int_{-\epsilon}^{\epsilon} \int_{-\infty}^{\infty} (\cdots)\, d\omega d\tau = W\delta(0)$$

□

Given a linear system

$$\dot{X} = AX + FW, \qquad Y = CX,$$

with W given by white noise, we can compute the spectral density function corresponding to the output Y. We start by computing the Fourier transform of the steady state correlation function (4.28):

$$\begin{aligned} S_Y(\omega) &= \int_{-\infty}^{\infty} \left[\int_0^{\infty} h(\xi) Q h(\xi + \tau) d\xi \right] e^{-j\omega\tau}\, d\tau \\ &= \int_0^{\infty} h(\xi) Q \left[\int_{-\infty}^{\infty} h(\xi + \tau) e^{-j\omega\tau}\, d\tau \right] d\xi \\ &= \int_0^{\infty} h(\xi) Q \left[\int_0^{\infty} h(\lambda) e^{-j\omega(\lambda - \xi)}\, d\lambda \right] d\xi \\ &= \int_0^{\infty} h(\xi) e^{j\omega\xi}\, d\xi \cdot QH(j\omega) = H(-j\omega) Q_u H(j\omega) \end{aligned}$$

This is then the (steady state) response of a linear system to white noise.

As with transfer functions, one of the advantages of computations in the frequency domain is that the composition of two linear systems can be represented by multiplication. In the case of the power spectral density, if we pass white noise through a system with transfer function $H_1(s)$ followed by transfer function $H_2(s)$, the resulting power spectral density of the output is given by

$$S_Y(\omega) = H_1(-j\omega) H_2(-j\omega) Q_u H_2(j\omega) H_1(j\omega).$$

As stated earlier, white noise is an idealized signal that is not seen in practice. One of the ways to produced more realistic models of noise and disturbances it to apply a filter to white noise that matches a measured power spectral density function. Thus, we wish to find a covariance W and filter $H(s)$ such that we match the statistics $S(\omega)$ of a measured noise or disturbance signal. In other words, given $S(\omega)$, find $W > 0$ and $H(s)$ such that $S(\omega) = H(-j\omega) W H(j\omega)$. This problem is know as the *spectral factorization problem*.

Figure 4.6 summarizes the relationship between the time and frequency domains.

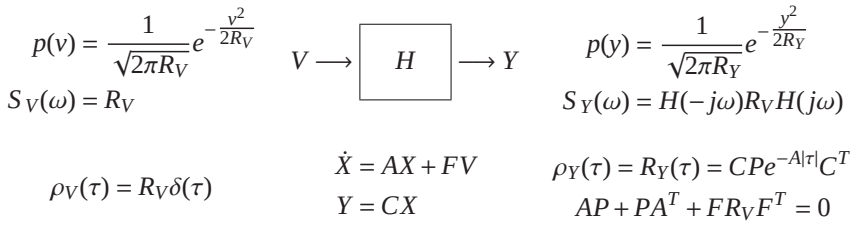


Figure 4.6: Summary of steady state stochastic response.

Application to Biomolecular Systems

4.6 Markov chain modeling and analysis

4.7 System identification techniques

4.8 Model Reduction

Chapter 5

Feedback Examples

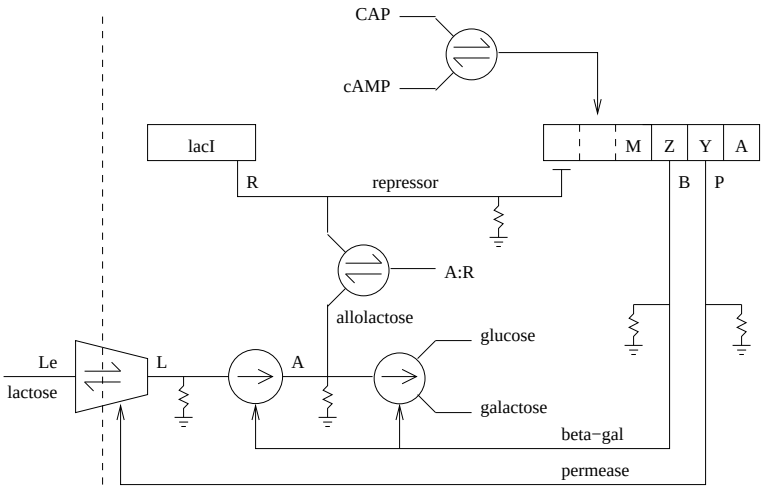


Figure 5.1: Schematic diagram for the *lac* system.

5.1 The *lac* Operon

Modeling

The *lac* operon is one of the most studied regulatory networks in molecular biology. Its function is to determine when the cell should produce the proteins and enzymes necessary to import and metabolize lactose from its external environment. Since glucose is a more efficient source of carbon, the lactose machinery is not produced unless lactose is present and glucose is not present. The *lac* control system implements this computation.

In constructing a model for the *lac* system, we need to decide what questions we wish to answer. Here we will attempt to develop a model that allows us to understand what levels of lactose are required for the *lac* system to become active in the absence of glucose. We will focus on the so-called "bi-stability" of the *lac* operon: there are two steady operating conditions—at low lactose levels the machinery is off and at high lactose levels the machinery is on. The system has hysteresis, so once the operon is actived, it remains active even if the lactose concentration descreases. We will construct a differential equation model of the system, with various simplifying assumptions along the way.

A schematic diagram of the *lac* control system is shown in Figure 5.1. Starting at the bottom of the figure, lactose permease is an integral membrane protein that helps transport lactose into the cell. Once in the cell, lactose is converted to allolactose, and allolactose is then broken down into glucose and galactose, both with the assistance of the enzyme β-galactosidase (β-gal for short). From here, the glucose is processed using the usual glucose metabolic pathway and the galactose.

The control circuitry is implemented via the reactions and transcriptional reg-

ulation shown in the top portion of the diagram. The *lac* operon, consisting of the genes *lacZ* (coding for β-gal), *lacY* (coding for lactose permease) and *lacA* (coding for a transacetylase), has a combinatorial promoter. Normally, lac repressor (*lacI*) is present and the operon is off. The activator for the operon is CAP, which has a positive inducer cAMP. The concentration of cAMP is controlled by glucose: when glucose is present, there is very little cAMP available in the cell (and hence CAP is not active).

The bistable switching behavior in the *lac* control system is implemented with a feedback circuit involving the *lac* repressor. Allolactose binds *lac* repressor and so when lactose is being metabolized, then the repressor is sequestered by allolactose and the *lac* operon is no longer repressed.

To model this circuit, we need to write down the dynamics of all of the reactions and protein production for the circuitry shown in Figure 5.1. We will denote the concentration of the β-gal mRNA and protein as m_b and B. We assume that the internal concentration of lactose is given by L, ignoring the dynamics of lactose permease and transport of lactose into the cell. Similarly, we assume that the concentration of repressor protein, denoted R, is constant.

We start by keeping track of the concentration of free allolactose A. The relevant reactions are given by the transport of lactose into the cell, the conversion of lactose into allolactose and then into glucose and lactose and finally the sequestration of repressor R by allolactose:

$$\begin{aligned} \mathrm{L^e + P} &\rightleftharpoons \mathrm{L^e{:}P} \rightleftharpoons \mathrm{L + P} && \text{Transport} \\ \mathrm{L + B} &\rightleftharpoons \mathrm{L{:}B} \longrightarrow \mathrm{A + B} && \text{Conversion} \\ \mathrm{A + B} &\rightleftharpoons \mathrm{A{:}B} \longrightarrow \mathrm{Glu + Gal + B} && \text{Conversion} \\ \mathrm{A + R} &\rightleftharpoons \mathrm{A{:}R} && \text{Sequestration.} \end{aligned}$$

We see that the dynamics involve a number of enzymatic reactions and hence we can use Michaelis-Menten kinetics to model the response at a slightly reduced level of detail. The differential equation for the internal lactose concentration L becomes

$$\frac{dL}{dt} = \alpha_{\mathrm{LL^e}} P \frac{L^{\mathrm{e}}}{K_{\mathrm{L^e}} + L^{\mathrm{e}}} - \alpha_{\mathrm{PL}} B \frac{L}{K_{\mathrm{PL}} + L} - \alpha_{\mathrm{AL}} B \frac{L}{K_{\mathrm{AL}} + L} - \delta_L L, \tag{5.1}$$

where the first two terms arise from the transport of lactose into and out of the cell, the third term is the conversion of lactose to allolactose and the final term is due to degradation and dilution. Similarly, the dynamics for the allolactose concentration can be modeled as

$$\frac{dA}{dt} = \alpha_{\mathrm{AL}} B \frac{L}{K_{\mathrm{AL}} + L} - \alpha_{\mathrm{AB}} B \frac{A}{K_A + A} + k^{\mathrm{r}}_{\mathrm{AR}}[\mathrm{A{:}R}] - k^{\mathrm{f}}_{\mathrm{AR}}[\mathrm{A}][\mathrm{R}] - \delta_A A.$$

The dynamics of the production of β-gal and lactose permease are given by the transcription and translational dynamics of protein production. These genes

are both part of the same operon (along with *lacA*) and hence the use a single mRNA strand for translation. To determine the production rate of mRNA, we need to determine the amount of repression that is present as a function of the amount of repressor, which in turn depends on the amount of allolactose that is present. We make the simplifying assumption that the sequestration reaction is fast, so that it is in equilibrium and hence

$$[\text{A:R}] = k_{\text{AR}}[\text{A}][\text{R}], \qquad k_{\text{AR}} = k^{\text{f}}_{\text{AR}}/k^{\text{r}}_{\text{AR}}.$$

We also assume that the total repressor concentration is constant (production matches degradation and dilution). Letting $R_T = [\text{R}] + [\text{A:R}]$ represent the total repressor concentration, we can write

$$[\text{R}] = R_T - k_{\text{AR}}[\text{A}][\text{R}] \qquad \Longrightarrow \qquad [\text{R}] = \frac{R_T}{1 + k_{\text{AR}}[\text{A}]}. \tag{5.2}$$

The simplification that the sequestration reaction is in equilibrium also simplifies the reaction dynamics for allolactose, which becomes

$$\frac{dA}{dt} = \alpha_{\text{AL}} B \frac{L}{K_{\text{AL}} + L} - \alpha_{\text{A}} B \frac{A}{K_A + A} - \delta_A A. \tag{5.3}$$

We next need to compute the effect of the repressor on the production of β-gal and lactose permease. It will be useful to express the promoter state in terms of the allolactose concentration A rather than R, using equation (5.2). We model this using a Hill function of the form

$$F_{\text{BA}}(A) = \frac{\alpha_{\text{R}}}{K_{\text{R}} + R^n} = \frac{\alpha_{\text{R}}(1 + K_{\text{AR}}A)^n}{K_{\text{R}}(1 + K_{\text{AR}}A)^n + R_T}$$

Letting M represent the concentration of the (common) mRNA, the resulting form of the protein production dynamics becomes

$$\begin{aligned}
\frac{dM}{dt} &= e^{-\mu\tau_M} F_{\text{BA}}(A(t - \tau_m)) - \bar{\gamma}_M M, \\
\frac{dB}{dt} &= \beta_B e^{-\mu\tau_B} M(t - \tau_B) - \bar{\delta}_B B, \\
\frac{dP}{dt} &= \beta_P e^{-\mu(\tau_M + \tau_P)} M(t - \tau_M - \tau_P) - \bar{\delta}_P P.
\end{aligned} \tag{5.4}$$

This model includes the degradation and dilution of mRNA ($\bar{\gamma}_M$), the transcriptional delays β-gal mRNA (τ_M), the degradation and dilution of the proteins ($\bar{\delta}_B$, $\bar{\delta}_P$) and the delays in the translation and folding of the final proteins (τ_B, τ_P).

Table 5.1: Parameter values for *lac* dynamics (from [**?**]).

Parameter	Value	Description
$\bar{\mu}$	3.03×10^{-2} min^{-1}	dilution rate
α_M	997 nMmin^{-1}	production rate of β-gal mRNA
β_B	1.66×10^{-2} min^{-1}	production rate of β-galactosidase
β_P	??? min^{-1}	production rate of lactose permease
α_A	1.76×10^{4} min^{-1}	production rate of allolactose
$\bar{\gamma}_M$	0.411 min^{-1}	degradation and dilution of β-gal mRNA
$\bar{\delta}_B$	8.33×10^{-4} min^{-1}	degradation and dilution of β-gal
$\bar{\delta}_P$	?? min^{-1}	degradation and dilution of lactose permease
$\bar{\delta}_A$	1.35×10^{-2} min^{-1}	degradation and dilution of allolactose
n	2	Hill coefficient for repressor
K	7200	
K_1	2.52×10^{-2} (μM)$^{-2}$	
K_L	0.97 μM	
K_A	1.95 μM	
β_A	2.15×10^{4} min^{-1}	
τ_M	0.10 min	
τ_M	2.00 min	

Bifurcation analysis

Sensitivity analysis

Consider the model of the *lac* operon introduced in Section **??**. For the gene *lacZ* (which encodes the protein β-galactosidase), we let B represent the protein concentration and M represent the mRNA concentration. We also consider the concentration of the lactose L inside the cell, which we will treat as an external input, and the concentration of allolactose, A. Assuming that the time delays considered previously can be ignored, the dynamics in terms of these variables are

$$\begin{aligned}
\frac{dM}{dt} &= F_{BA}(A,\theta) - \gamma_b M, & F_{BA}(A,\theta) &= \alpha_{AB}\frac{1+k_1A^n}{K+k_1A^n},\\
\frac{dB}{dt} &= \beta_B M - \delta_B B, & F_{AL}(L,\theta) &= \alpha_A\frac{L}{k_L+L},\\
\frac{dA}{ddt} &= BF_{AL}(L,\theta) - BF_{AA}(A,\theta) - \gamma_A A, & F_{AA}(A,\theta) &= \beta_A\frac{A}{k_A+A}.
\end{aligned} \tag{5.5}$$

Here the state is $x = (M,B,A) \in \mathbb{R}^3$, the input is $w = L \in \mathbb{R}$ and the parameters are $\theta = (\alpha_B,\beta_B,\alpha_A,\gamma_B,\delta_B,\gamma_A,n,k,k_1,k_L,k_A,\beta_A) \in \mathbb{R}^{12}$. The values for the parameters are listed in Table **??**.

We investigate the dynamics around one of the equilibrium points, corresponding to an intermediate input of $L = 40$ μM. There are three equilibrium points at

this value of the input:

$$x_{1,e} = (0.000393, 0.000210, 3.17), \qquad x_{2,e} = (0.00328, 0.00174, 19.4), \qquad x_{3,e} = (0.0142, 0.00758, 42.1).$$

We choose the third equilibrium point, corresponding to the lactose metabolic machinery being activitated and study the sensitivity of the steady state concentrations of allolactose (A) and β-galactosidase (B) to changes in the parameter values.

The dynamics of the system can be represented in the form $\dot{x} = f(x, \theta, L)$ with

$$f(x,\theta,L) = \begin{pmatrix} F_{\mathrm{BA}}(A) - \gamma_B M - \mu M \\ \beta_B M - \delta_B B - \mu B \\ F_{\mathrm{AL}}(L)B - F_{\mathrm{AA}}(A)B - \delta_A A - \mu A \end{pmatrix}.$$

To compute the sensitivity with respect to the parameters, we compute the derivatives of f with respect to the state x,

$$\frac{\partial f}{\partial x} = \begin{pmatrix} -\gamma_B - \mu & 0 & \frac{\partial F_{BA}}{\partial A} \\ \beta_B & -\delta_B - \mu & 0 \\ 0 & F_{AL} - F_{\mathrm{AA}} & -B\frac{\partial F_{AA}}{\partial A} \end{pmatrix}$$

and the parameters θ,

$$\frac{\partial f}{\partial \theta} = \begin{pmatrix} F_{BA} & 0 & 0 & -M & 0 & 0 & \frac{\partial F_{BA}}{\partial n} & \frac{\partial F_{BA}}{\partial k} & \frac{\partial F_{BA}}{\partial k_1} & 0 & 0 & 0 \end{pmatrix}.$$

Carrying out the relevant computations and evaluating the resulting expression numerically, we obtain

$$\frac{\partial}{\partial \theta}\begin{pmatrix} B_e \\ A_e \end{pmatrix} = \begin{pmatrix} -1.21 & 0.0243 & -3.35\times 10^{-6} & 0.935 & 1.46 & \dots & 0.00115 \\ -2720. & 47.7 & -0.00656 & 1830. & 2860. & \dots & 3.27 \end{pmatrix}.$$

We can also normalize the sensitivity computation:

$$\bar{S}_{x_e\theta} = \frac{\partial x_e / x_e}{\partial \theta / \theta_0} = D^{-1}(x_e) S_{x_e\theta} D^{-1}(\theta_0)$$

which yields

$$\bar{S}_{y_e\theta} = \begin{pmatrix} -4.85 & 3.2 & -3.18 & 3.11 & 3.2 & 6.3 & -6.05 & -4.1 & 4.02 & 6.05 \\ -1.96 & 1.13 & -1.12 & 1.1 & 1.13 & 3.24 & -3.11 & -2.11 & 2.07 & 3.11 \end{pmatrix}$$

where

$$\theta = \begin{pmatrix} \mu & \alpha_M & K & K_1 & \beta_B & \alpha_A & K_L & \beta_A & K_A & L \end{pmatrix}.$$

We see from this computation that increasing the growth rate decreases the equilibrium concentation of B and A, while increasing the lactose concentration by 2-fold increases the equilibrium β-gal concentration 12-fold (6X) and the allolactose concentration by 6-fold (3X).

5.2 Heat Shock Response in Bacteria

5.3 Bacteriophage λ

Bacteriophage λ (also called λ phage or phage λ) is a virus that infects *E. coli* and propogates itself by integrating its DNA into the genome of the infected cell. The virus includes a decision "switch" that determines whether the virus should propogate itself by DNA integration (the *lysogenic* phase) or whether it should destroy the host cell and spread to other nearby bacteria (the *lytic* phase). In this section we describe what is known about the modeling of the lysis/lysogeny decision-making circuitry and explore some of the properties of its dynamics.

The material in this section is based on the work of Ptashne [**?**], Arkin et al. [**?**] and St. Pierre et al. [**?**]. The models used to create the plots in this section are available on the companion web site for the text.

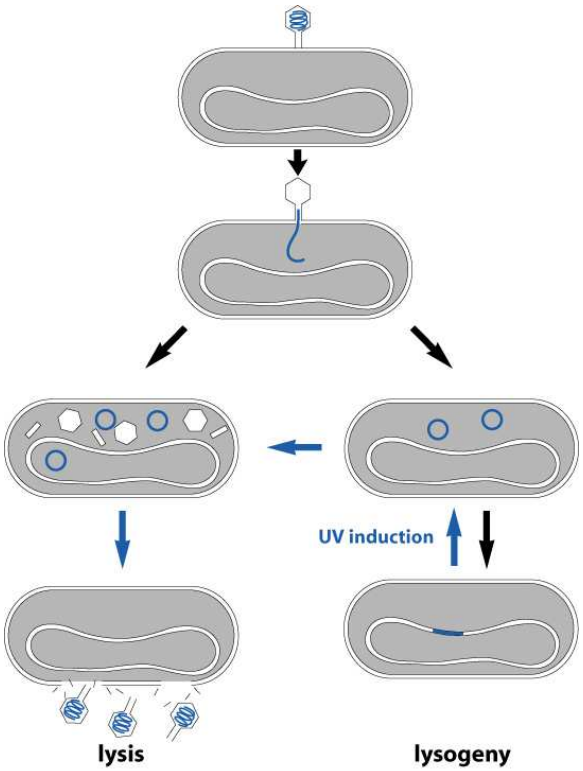
Phage λ lifecycle

A detailed model for λ

Reduced order models for λ

Dynamic analysis

Open issues



A Genetic Switch, 3rd edition, 2004
© Cold Spring Harbor Laboratory Press
Chapter 1, Figure 2

Figure 5.2: Growth cycle of phage λ. From Ptashne.

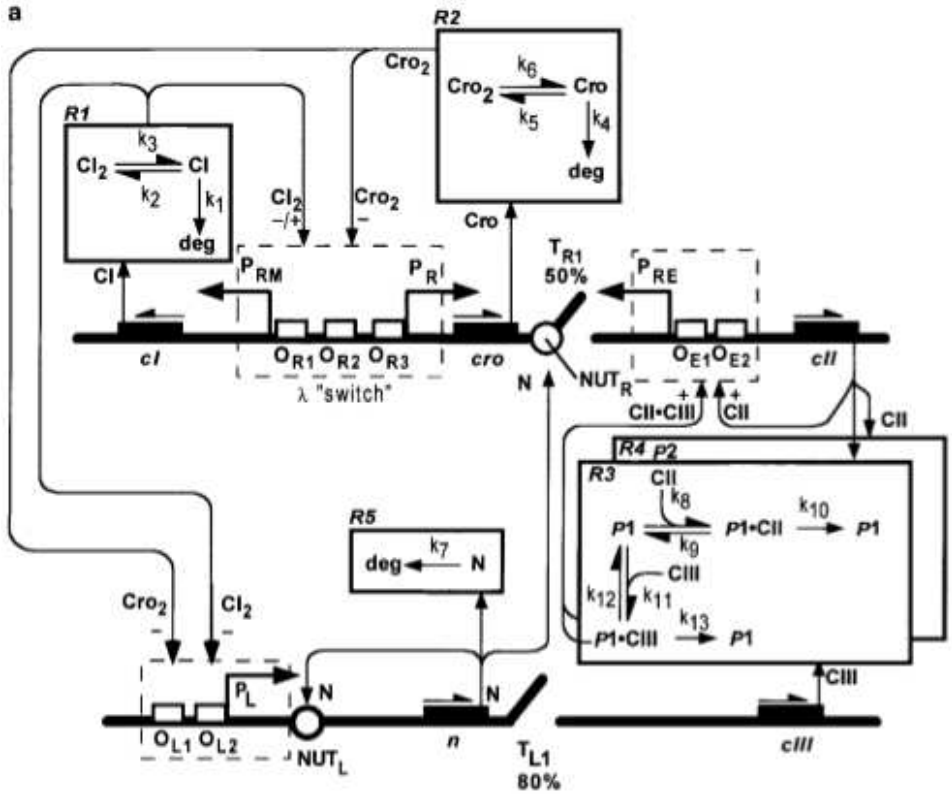
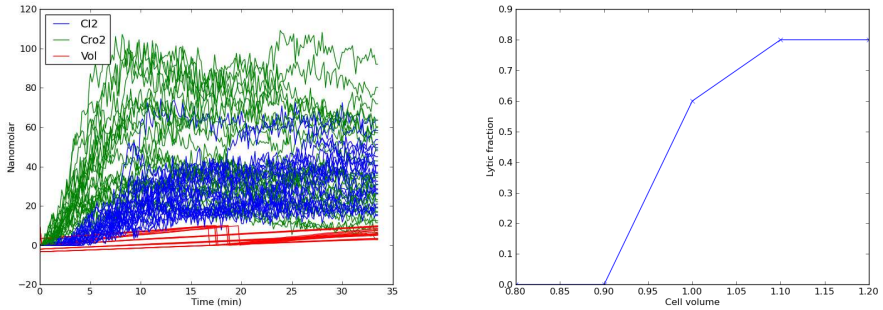


Figure 5.3: A detailed circuit diagram for the λ decision-making circuit. From Arkin, Ross and McAdams (1998).



(a) (b)

Figure 5.4: Simulation results using the detailed model.

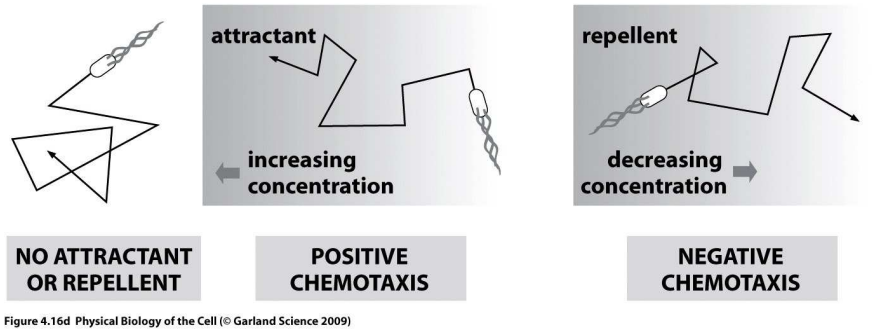


Figure 5.5: Examples of chemotaxis. Figure from Phillips, Kondev and Theriot [35]; used with permission of Garland Science.

5.4 Bacterial Chemotaxis

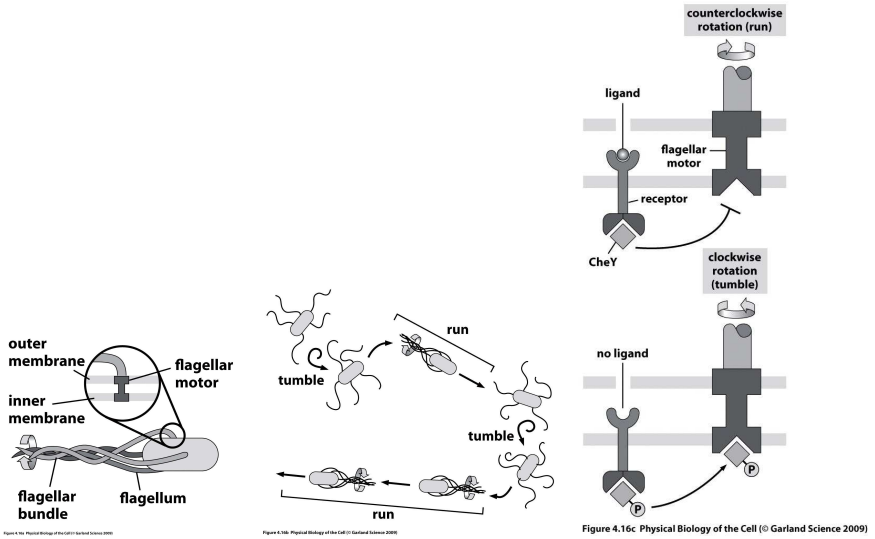
Chemotaxis refers to the process by which micro-organisms move in response to chemical stimuli. Examples of chemotaxis include the ability of organisms to move in the direction of nutrients or move away from toxins in the environment. Chemotaxis is called *positive chemotaxis* if the motion is in the direction of the stimulus and *negative chemotaxis* if the motion is away from the stimulant, as shown in Figure 5.5. Many chemotaxis mechanisms are stochastic in nature, with biased random motions causing the average behavior to be either positive, negative or neutral (in the absence of stimuli).

In this section we look in some detail at bacterial chemotaxis, which *E. coli* use to move in the direction of increasing nutrients. The material in this section is based primarily on the work of Barkai and Leibler [8] and Rao, Kirby and Arkin [37].

Control system overview

The chemotaxis system in *E. coli* consists of a sensing system that detects the presence of nutrients, and actuation system that propels the organism in its environment, and control circuitry that determines how the cell should move in the presence of chemicals that stimulate the sensing system. The approximate location of these elements are shown in Figure **??**.

The actuation system in the *E. coli* consists of a set of flagella that can be spun using a flagellar motor embedded in the outer membrane of the cell, as shown in Figure 5.6a. When the flagella all spin in the counter clockwise direction, the individual flagella form a bundle and cause the organism to move roughly in a straight line. This behavior is called a “run” motion. Alternatively, if the flagella spin in the clockwise direction, the individual flagella do not form a bundle and the organism “tumbles”, causing it to rotate (Figure 5.6b). The selection of the motor direction is controlled by the protein CheY: if phosphorylated CheY binds to the



(a) (b) (c)

Figure 5.6: Bacterial chemotaxis. Figures from Phillips, Kondev and Theriot [35]; used with permission of Garland Science.

motor complex, the motor spins clockwise (tumble), otherwise it spins counter-clockwise (run).

Because of the size of the organism, it is not possible for a bacterium to sense gradients across its length. Hence, a more sophisticated strategy is used, in which the organism undergoes a combination of run and tumble motions. The basic idea is illustrated in Figure 5.6c: when high concentration of ligand (nutrient) is present, the CheY protein is left unphosphorylated and does not bind to the actuation complex, resulting in a counter-clockwise rotation of the flagellar motor (run). Conversely, if the ligand is present then the molecular machinery of the cell causes CheY to be phosphorylated and this modifies the flagellar motor dynamics so that a clockwise rotation occurs (tumble). The net effect of this combination of behaviors is that when the organism is traveling through regions of higher nutrient concentration, it continues to move in a straight line for a longer period before tumbling, causing it to move in directions of increasing nutrient concentration.

A simple model for the molecular control system that regulates chemotaxis is shown in Figure 5.7. We start with the basic sensing and and actuation mechanisms. A membrane bound protein MCP (methyl-accepting chemotaxis protein) that is capable of binding to the external ligand serves as a signal transducing element from the cell exterior to the cytoplasm. Two other proteins, CheW and CheA, form a complex with MCP. This complex can either be in an active or inactive state. In the active state, CheA is autophosphorylated and serves as a phosphotransferase

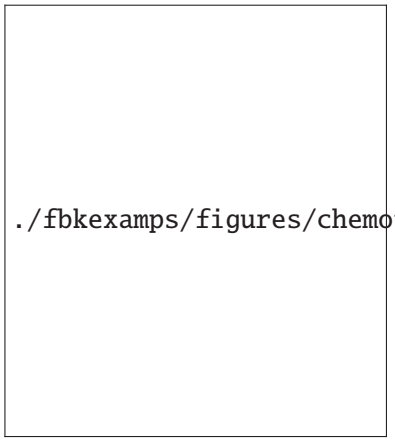


Figure 5.7: Control system for chemotaxis. Figure from Rao *et al.* [37] (Figure 1A).

for two additional proteins, CheB and CheY. The phosphorylated form of CheY then binds to the motor complex, causing clockwise rotation of the motor.

The activity of the receptor complex is governed by two primary factors: the binding of a ligand molecule to the MCP protein and the presence or absence of up to 4 methyl groups on the MCP protein. The specific dependence on each of these factors is somewhat complicated. Roughly speaking, when the ligand L is bound to the receptor then the complex is less likely to be active. Furthermore, as more methyl groups are present, the ligand binding probability increases, allowing the gain of the sensor to be adjusted through methylation. Finally, even in the absence of ligand the receptor complex can be active, with the probability increasing with increased methylation. Figure 5.8 summarizes the possible states, their free energies and the probability of activity.

Several other elements are contained in the chemotaxis control circuit. The most important of these are implemented by the proteins CheR and CheB, both of which affect the receptor complex. CheR, which is constitutively produced in the cell, methylates the receptor complex at one of the four different methylation sites. Conversely, the phosphorylated form of CheB demethylates the receptor complex. As described above, the methylation patterns of the receptor complex affect its activity, which affects the phosphorylation of CheA and, in turn, phosphorylation of CheY and CheB. The combination of CheA, CheB and the methylation of the receptor complex forms a negative feedback loop: if the receptor is active, then CheA phosphorylates CheB, which in turn demethylates the receptor complex, making it less active. As we shall see when we investigate the detailed dynamics below, this feedback loop corresponds to a type of integral feedback law. This integral action allows the cell to adjust to different levels of ligand concentration, so that the behavior of the system is invariant to the absolute nutrient levels.

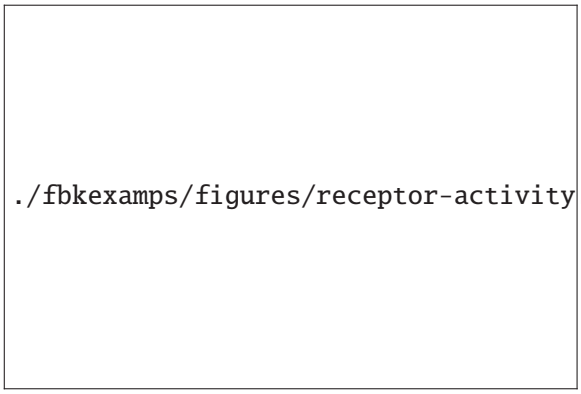


Figure 5.8: Receptor complex states. The probability of a given state being in an active configuration is given by p. Figure obtained from [30].

Modeling

The detailed reactions that implement chemotaxis are illustrated in Figure 5.9. Letting T represent the receptor complex and T^A represent an active form, the basic reactions can be written as

$$\begin{aligned} T^A + A &\rightleftharpoons T^A{:}A \longrightarrow A^p + T^A \\ A^p + B &\rightleftharpoons A^p{:}B \longrightarrow A + B^p \qquad B^p + P \rightleftharpoons B^p{:}P \longrightarrow B + P \\ A^p + Y &\rightleftharpoons A^p{:}Y \longrightarrow A + Y^p \qquad Y^p + Z \rightleftharpoons Y^p{:}Z \longrightarrow Y + Z \end{aligned} \tag{5.6}$$

where CheA, CheB, CheY and CheZ are written simply as A, B, Y and Z for simplicity and P is a non-specific phosphotase. We see that these are basically three linked sets of phosphorylation and dephosphorylation reactions, with CheA serving as a phosphotransferase and P and CheZ serving as phosphotases.

The description of the methylation of the receptor complex is a bit more complicated. Each receptor complex can have multiple methyl groups attached and the activity of the receptor complex depends on both the amount of methylation and whether a ligand is attached to the receptor site. Furthermore, the binding probabilities for the receptor also depend on the methylation pattern. To capture this, we use the set of reactions that are illustrated in Figures 5.7 and 5.9. In this diagram, T_i^s represents a receptor that has i methylation sites filled and ligand state s (which can be either u if unoccupied or o if occupied). We let M represent the maximum number of methylation sites ($M = 4$ for *E. coli*).

Using this notation, the transitions between the states correspond to the reactions shown in Figure 5.10:

$$\begin{aligned} T_i^x + B^p &\rightleftharpoons T_i^x{:}B^p \longrightarrow T_{i-1}^x + B^p \qquad i > 0 \\ T_i^x + R &\rightleftharpoons T_i^x{:}R \longrightarrow T_{i+1}^x + R \qquad i < M \\ T_i^u + L &\rightleftharpoons T_i^o \end{aligned}$$

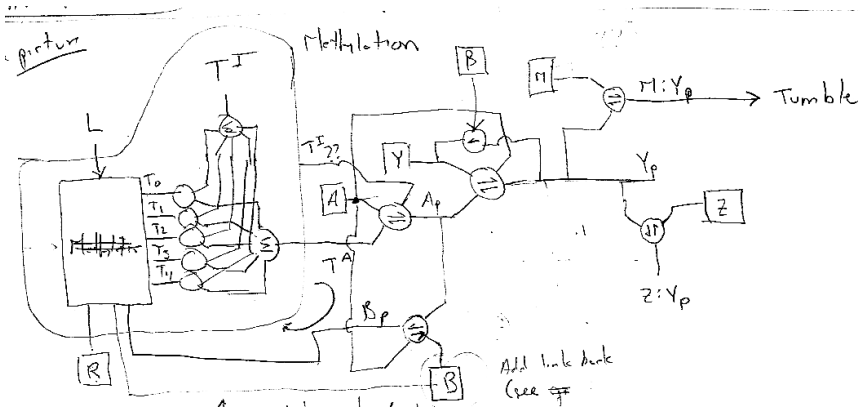


Figure 5.9: Circuit diagram for chemotaxis.

We now must write reactions for each of the receptor complexes with CheA. Each form of the receptor complex has a different activity level and so the most complete description is to write a separate reaction for each T_i^o and T_i^u species:

$$T_i^x + A \underset{k_i^{r,o}}{\overset{k_i^{f,o}}{\rightleftharpoons}} T_i^x{:}A \xrightarrow{k_i^{c,o}} A^p + T_i^x,$$

where $x \in \{o, u\}$ and $i = 0, \ldots, M$. This set of reactions replaces the placeholder reaction $T^A + A \rightleftharpoons T^A{:}A \longrightarrow A^p + T^A$ used earlier.

Approximate Model

The detailed model described above is sufficiently complicated that it can be difficult to analyze. In this section we develop a slightly simpler model that can be used to explore the adaptation properties of the circuit, which happen on a slower time-scale.

./fbkexamps/figures/chemotaxis-methylation.eps

Figure 5.10: Methylation model for chemotaxis. Figure from Barkai and Leibler [8] (Box 1). Note: the figure uses the notation E_i^s for the receptor complex instead of T_i^s.

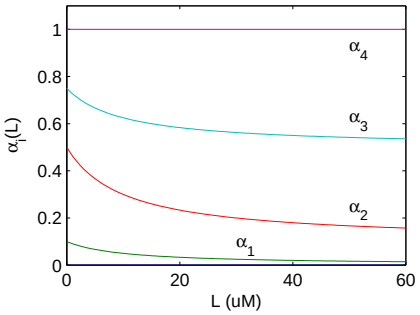


Figure 5.11: Probability of activity.

We begin by simplifying the representation of the receptor complex and its methylation pattern. Let $L(t)$ represent the ligand concentration and T_i represent the concentration of the receptor complex with i sides methylated. If we assume that the binding reaction of the ligand L to the complex is fast, we can write the probability that a receptor complex with i sites methylated is in its active state as a static function $\alpha_i(L)$, which we take to be of the form

$$\alpha_i(L) = \frac{\alpha_i^o L}{K_L + L} + \frac{\alpha_i K_L}{K_L + L}.$$

The coefficients α_i^o and α_i capture the effect of presence or absence of the ligand on the activity level of the complex. Note that α_i has the form of a Michaelis-Menten function, reflecting our assumption that ligand binding is fast compared to the rest of the dynamics in the model. Following [37], we take the coefficients to be

$$\begin{aligned} a_0 &= 0, & a_1 &= 0.1, & a_2 &= 0.5, & a_3 &= 0.75, & a_4 &= 1, \\ a_0^o &= 0, & a_1^o &= 0, & a_2^o &= 0.1, & a_3^o &= 0.5, & a_4^o &= 1. \end{aligned}$$

and choose $K_L = 10$ μM. Figure 5.11 shows how each α_i varies with L.

The total concentration of active receptors can now be written in terms of the receptor complex concentrations T_i and the activity probabilities $\alpha_i(L)$. We write the concentration of activated complex $\mathrm{T^A}$ and inactivated complex $\mathrm{T^I}$ as

$$T^A = \sum_{i=0}^{4} \alpha_i(L) T_i, \qquad T^I = \sum_{i=0}^{4} (1 - \alpha_i(L)) T_i.$$

These formulas can now be used in our dynamics as an effective concentration of active or inactive receptors, justifying the notation that we used in equation (5.6).

We next model the transition between the methylation patterns on the receptor. We assume that the rate of methylation depends on the activity of the receptor complex, with active receptors less likely to be demethylated and inactive receptors

less likely to be methylated [37, 30]. Let

$$r_B = k_B \frac{B^p}{K_B + T^A}, \qquad r_R = k_R \frac{R}{K_R + T^I},$$

represent rates of the methylation and demethylation reactions. We choose the coefficients as

$$k_B = 0.5, \quad K_B = 5.5, \quad k_R = 0.255, \quad K_R = 0.251,$$

We can now write the methylation dynamics as

$$\frac{d}{dt}T_i = r_R(1-\alpha_{i+1}(L))T_{i-1} + r_B\alpha_{i+1}(L)T_{i+1} - r_R(1-\alpha_i(L))T_i - r_B\alpha_i(L)T_i,$$

where the first and second terms represent transitions into this state via methylation or demethylation of neighboring states (see Figure 5.10) and the last two terms represent transitions out of the current state by methylation and demethylation, respectively. Note that the equations for T_0 and T_4 are slightly different since the demethylation and methylation reactions are not present, respectively.

Finally, we write the dynamics of the phosphorylation and dephosphorylation reactions, and the binding of $\mathrm{CheY^p}$ to the motor complex. Under the assumption that the concentrations of the phosphorylated proteins are small relative to the total protein concentrations, we can approximate the reaction dynamics as

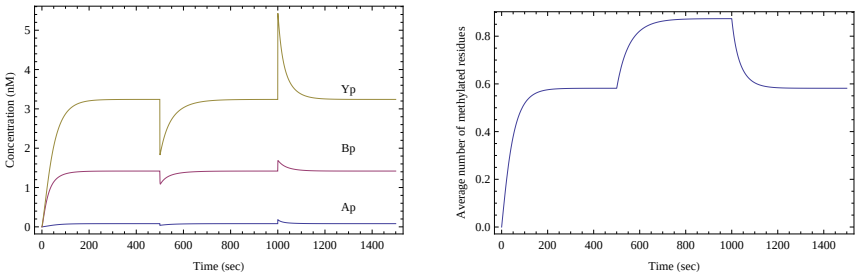
$$\begin{aligned}
\frac{d}{dt}A^p &= 50T^A A - 100A^pY - 30A^pB, \\
\frac{d}{dt}Y^p &= 100A^pY - 0.1Y^p - 5[\mathrm{M}]Y^p + 19[\mathrm{M{:}Y^p}] - 30Y^p, \\
\frac{d}{dt}B^p &= 30A^pB - B^p, \\
\frac{d}{dt}[\mathrm{M{:}Y^p}] &= 5[\mathrm{M}]Y^p - 19[\mathrm{M{:}Y^p}].
\end{aligned}$$

The total concentrations of the species are given by

$$A + A^p = 5 \text{ nM}, \qquad B + B^p = 2 \text{ nM}, \qquad Y + Y^p + [\mathrm{M{:}Y^p}] = 17.9 \text{ nM},$$
$$[\mathrm{M}] + [\mathrm{M{:}Y^p}] = 5.8 \text{ nM}, \qquad R = 0.2 \text{ nM}, \qquad \sum_{i=0}^{4} T_i = 5 \text{ nM}.$$

The reaction coefficients and concentrations are taken from Rao *et al.* [37].

Figure 5.12a shows a the concentration of the phosphorylated proteins based on a simulation of the model. Initially, all species are started in their unphosphorylated and demethylated states. At time $T = 500$ s the ligand concentration is increased to $L = 10$ μM and at time $T = 1000$ it is returned to zero. We see that immediately after the ligand is added, the $\mathrm{CheY^p}$ concentration drops, allowing longer runs between tumble motions. After a short period, however, the $\mathrm{CheY^p}$ concentration adapts to



(a) (b)

Figure 5.12: Simulation and analysis of reduced-order chemotaxis model.

the higher concentration and the nominal run versus tumble behavior is restored. Similarly, after the ligand concentration is decreased the concentration of $\mathrm{CheY^p}$ increases, causing a larger fraction of tumbles (and subsequent changes in direction). Again, adaptation over a longer time scale returns that CheY concentration to its nominal value.

Figure 5.12b helps explain the adaptation response. We see that the average amount of methylation of the receptor proteins increases when the ligand concentration is high, which decreases the activity of CheA (and hence decreases the phosphorylation of CheY).

Integral action

The perfect adaptation mechanism in the chemotaxis control circuitry has the same function as the use of integral action in control system design: by including a feedback on the integral of the error, it is possible to provide exact cancellation to constant disturbances. In this section we demonstrate that a simplified version of the dynamics can indeed be regarded as integral action of an appropriate signal. This interpretation was first pointed out by Yi *et al* [48].

We begin by formulating an even simpler model for the system dynamics that captures the basic features required to understand the integral action. Let X represent the receptor complex and assume that it is either methylated or not. We let $\mathrm{X_m}$ represent the methylated state and we further assume that this methylated state can be activated, which we write as $\mathrm{X_m^*}$. This simplified description replaces the multiple states T_i and probabilities $\alpha_i(L)$. We also ignore the additional phosphorylation dynamics of CheY and simply take the activated receptor concentration X_m^* as our measure of overall activity.

Figure 5.13 shows the transitions between the various forms X. As before, CheR methylates the receptor and $\mathrm{CheB^p}$ demethylates it. We simplify the picture by only allowing $\mathrm{CheB^p}$ to act on the active state $\mathrm{X_m^*}$ and CheR to act on the inactive state. We take the ligand into account by assuming that the transition between the active

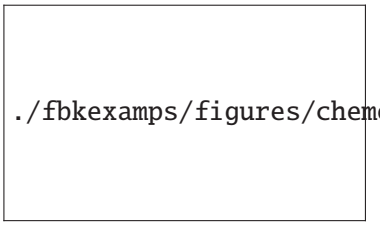


Figure 5.13: Reduced order model of receptor activity. Obtained from [3], Figure 7.9.

form X_m^* and the inactive form X_m depends on the ligand concentration: higher ligand concentration will increase the rate of transition to the inactive state.

This model is a considerable simplification from the ligand binding model that is illustrated in Figures 5.8 and 5.10. In the previous models, there is some probability of activity with or without methylation and with or without ligand. In this simplified model, we assume that only three states are of interest: demethylated, methylated/inactive and methylated/active. We also modify the way that that ligand binding is captured and instead of keeping track of all of the possibilities in Figure 5.8, we assume that the ligand transitions us from an active state X_m^* to an inactive X_m. These states and transitions are roughly consistent with the different energy levels and probabilities in Figure 5.8, but it is clearly a much coarser model.

Accepting these approximations, the model illustrated in Figure 5.13 results in a set of chemical reactions of the form

$$\begin{aligned}
&\mathrm{X} + \mathrm{R} \rightleftharpoons \mathrm{X{:}R} \longrightarrow \mathrm{X_m} + \mathrm{R} && \text{methylation} \\
&\mathrm{X_m^*} + \mathrm{B^p} \rightleftharpoons \mathrm{X_m^*{:}B^p} \longrightarrow \mathrm{X} + \mathrm{B^p} && \text{demethylation} \\
&\mathrm{X_m^*} \underset{k^r}{\overset{k^f(L)}{\rightleftharpoons}} \mathrm{X_m} && \text{activation/deactivation}
\end{aligned}$$

For simplicity we take both R and $\mathrm{B^p}$ to have constant concentration.

Approximating the first two reactions by their Michaelis-Menten forms and assuming that $X \gg 1$, we can write the resulting dynamics for the system as

$$\begin{aligned}
\frac{d}{dt} X_m &= k_R R + k^f(L) X_m^* - k^r X_m \\
\frac{d}{dt} X_m^* &= -k_B B^p \frac{X_m^*}{K_{X_m^*} + X_m^*} - k^f(L) X_m^* + k^r X_m.
\end{aligned}$$

We wish to use this model to understand how the steady state activity level X_m^* depends on the ligand concentration L (which enters through the deactivation rate $k^f(L)$). Starting with the first equation, we see that at equilibrium we have

$$X_{m,e} = (K_R/k^r)R.$$

To find $X^*_{m,e}$, we note that at equilibrium

$$0 = \frac{d}{dt}(X_{m,e} + X^*_{m,e}) = -k_B B^p \frac{X^*_{m,e}}{K_{X^*_m} + X^*_{m,e}} + k_R R.$$

From this equation we can solve for $X^*_{m,e}$ as a function of the CheR concentration:

$$X^*_{m,e} = \frac{K_{X^*_m} k_R R}{k_B B^p - k_R R}$$

Note that this solution does not depend on $k^f(L)$ or k^r and hence we see that the steady state solution is independent of the ligand concentration.

To see the integral action more directly, we write the dynamics in terms of a new variable $z = X^*_m - X^*_{m,e}$.

Further reading

5.5 Yeast mating response

Part II

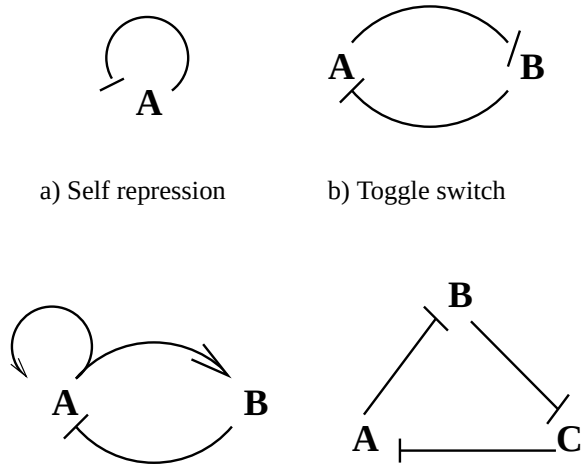
Design and Synthesis

Chapter 6
Biological Circuit Components

6.1 Biology Circuit Design

One of the fundamental building blocks employed in synthetic biology is the process of transcriptional regulation, which is found in natural transcriptional networks. A transcriptional network is composed of a number of genes that express proteins that then act as transcription factors for other genes. The rate at which a gene is transcribed is controlled by the *promoter*, a regulatory region of DNA that precedes the gene. RNA polymerase binds a defined site (a specific DNA sequence) on the promoter. The quality of this site specifies the transcription rate of the gene (the sequence of the site determines the chemical affinity of RNA polymerase to the site). RNA polymerase acts on all of the genes. However, each transcription factor modulates the transcription rate of a set of target genes. Transcription factors affect the transcription rate by binding specific sites on the promoter region of the regulated genes. When bound, they change the probability per unit time that RNA polymerase binds the promoter region. Transcription factors thus affect the rate at which RNA polymerase initiates transcription. A transcription factor can act as a *repressor* when it prevents RNA polymerase from binding to the promoter site. A transcription factor acts as an *activator* if it facilitates the binding of RNA polymerase to the promoter. Such interactions can be generally represented as nodes connected by directed edges. Synthetic bio-molecular circuits are fabricated typically in bacteria *E. coli*, by cutting and pasting together according to a desired sequence genes and promoter sites (natural and engineered). Since the expression of a gene is under the control of the upstream promoter region, one can this way create a desired circuit of activation and repression interactions among genes. Early examples of such circuits include an activator-repressor system that can display toggle switch or clock behavior [6], a loop oscillator called the repressilator obtained by connecting three inverters in a ring topology [14], a toggle switch obtained connecting two inverters in a ring fashion [16], and an autorepressed circuit [9] (Figure 6.1). Several scientific and technological developments accumulating over the past four decades have set the stage for the design and fabrication of early synthetic bio-molecular circuits (Figure 6.2).

An early milestone in the history of synthetic biology can be traced back to the discovery of mathematical logic in gene regulation. In their 1961 paper, Jacob and Monod introduced for the first time the idea of gene expression regulation through transcriptional feedback [25]. Only a few years later (1969), special enzymes that



a) Self repression b) Toggle switch

c) Activator-repressor clock d) Repressilator

Figure 6.1: Early transcriptional circuits that have been fabricated in bacteria *E. coli*: the self-repression circuit [9], the toggle switch [16], the activator-repressor clock [6], and the repressilator [14]. Each node represents a gene and each arrow from node Z to node X indicates that the transcription factor encoded in z, denoted Z, regulates gene x [3]. If z represses the expression of x, the interaction is represented by Z⊣X. If z activates the expression of x, the interaction is represented by Z→X [3].

can cut double-stranded DNA at specific recognition sites (known as restriction sites) were discovered by Arber and co-workers [4]. These enzymes, called restriction enzymes, were major enabler of recombinant DNA technology. One of the most celebrated products of such a technology is the large scale production of insulin by employing *E. coli* bacteria as a cell factory [47]. The development of recombinant DNA technology along with the demonstration in 1970 that genes can be artificially synthesized, provided the ability to cut and paste natural or synthetic promoters and genes in almost any fashion on size-wise compatible plasmids. This "cut and paste" procedure is called *cloning* [2]. Cloning of any DNA fragment involves four steps: *fragmentation, ligation, transfection*. The DNA of interest is first isolated. Then, a ligation procedure is employed in which the amplified fragment is inserted into a vector. The vector (which is frequently circular) is linearized by means of restriction enzymes that cleave it at target sites called restriction sites. It is then incubated with the fragment of interest with an enzyme called *DNA ligase*. Polymerase chain reaction (PCR), devised in the 1980s, allows then to exponentially amplify a small amount of DNA in amounts large enough to be used for transfection and transformation in living cells [2]. Today, commercial synthesis of DNA sequences and genes has become cheaper and faster with a price often below $ 1 per base pair [7].

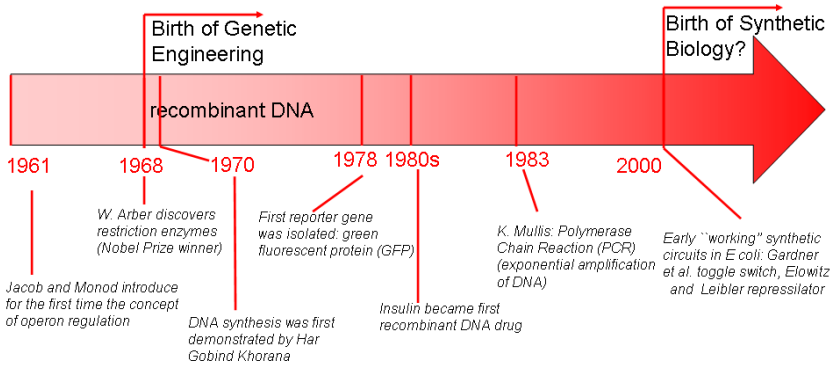


Figure 6.2: Milestones in the history of synthetic biology.

Another key enabling technology has been the development of *in vivo* measurement techniques that allow to measure the amount of protein produced by a target gene x. For instance, green fluorescent protein (GFP) is a protein with the property that it fluoresces in green when exposed to UV light. It is produced by the jellyfish *Aequoria victoria*, and its gene has been isolated so that it can be used as a reporter gene. The GFP gene is inserted (cloned) into the chromosome, adjacent to or very close to the location of gene x, so both are controlled by the same promoter region. Thus, gene x and GFP are transcribed simultaneously and then translated, so by measuring the intensity of the GFP light emitted one can estimate how much of x is being expressed. Other fluorescent proteins, such as yellow fluorescent protein (YFP) and red fluorescent protein (RFP) are genetic variations of the GFP.

Just as fluorescent proteins can be used as a read out of a circuit, inducers function as external inputs that can be used to probe the system. Inducers function by disabling repressor proteins. Repressor proteins bind to the DNA strand and prevent RNA polymerase from being able to attach to the DNA and synthesize mRNA. Inducers bind to repressor proteins, causing them to change shape and making them unable to bind to DNA. Therefore, they allow transcription to take place.

Inset (Electronic circuits)**.** One of the current directions of the field is to create circuitry with more complex functionalities by assembling simpler circuits, such as those in Figure 6.1. This tendency is consistent with what has been observed in the history of electronics: after the bipolar junction transistor (BJT) was invented in 1947 by William Shockley and co-workers, the transistor era started. A major breakthrough in the transistor era occurred in 1964 with the invention of the first operational amplifier (op amp), which led the way to standardized modular and integrated circuit design. By comparison, synthetic biology may be directing toward a similar development, in which modular and integrated circuit design becomes a reality. This is witnessed by several recent efforts toward formally characteriz-

ing interconnection mechanisms between modules, impedance-like effects, and op amp-like devices to counteract impedance problems [21, 40, 39, 13, 38, 42, 41]. ◊

Chapter 7
Interconnecting Components

7.1 Input/Output Modeling and the Modularity Assumption

Each node y of a transcriptional circuitry is usually modeled as an input/output module taking as input the concentrations of transcription factors that regulate gene y and giving as output the concentration of protein expressed by gene y, denoted Y. This is not the only possible choice for delimiting a module: one could in fact let the messenger RNA (mRNA) or the RNA polymerase flow along the DNA (as suggested by [15]) play the role of input and output signals. The transcription factor enters as input of the transcriptional module through the binding and unbinding dynamics of the transcription factors with the DNA promoter sites upstream of gene y. The internal dynamics of the transcriptional component is determined by the transcription and translation dynamics. The processes of transcription and translation are much slower than the binding dynamics of the transcription factor to the promoter sites on the DNA [3]. Thus, the binding of the transcription factor to the DNA promoter site reaches the equilibrium in seconds, while transcription and translation of the target gene takes minutes to hours. This time scale separation, a key feature of transcriptional circuits, leads to the following central modeling simplification.

> **Modularity assumption.** The dynamics of transcription factor/DNA binding are considered at the equilibrium and each transcription factor concentration enters the input/output transcriptional module through *static* input functions that drive the transcription/translation dynamics (Figure 7.1).

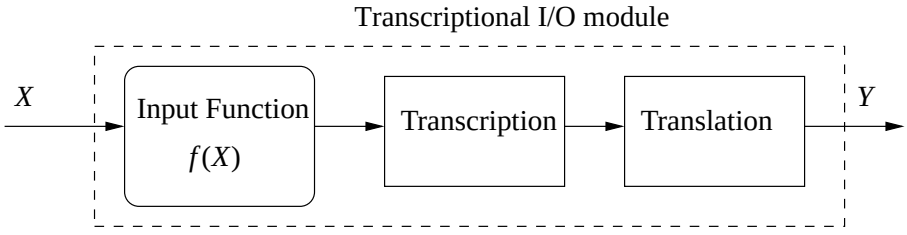


Figure 7.1: A transcriptional module is modeled as an input/output component with input function given by the transcription regulation function $f(X)$ and with internal dynamics established by the transcription and translation processes.

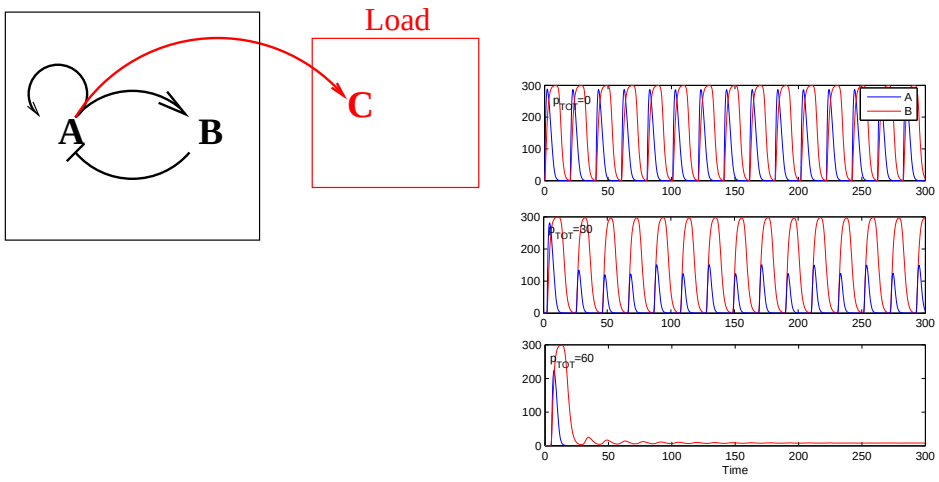


Figure 7.2: The clock behavior can be destroyed by a load. As the number of downstream binding sites for A, p_{TOT}, is increased in the load, the activator and repressor dynamics loose their synchronization and ultimately the oscillations disappear.

For engineering a system with prescribed behavior, one has to be able to change the physical features so as to change the values of the parameters of the model. This is often possible. For example, the binding affinity ($1/K$ in the Hill function model) of a transcription factor to its site on the promoter can be affected by single or multiple base pairs substitutions. The protein decay rate (constant α_2 in equation (2.16)) can be increased by adding degradation tags at the end of the gene expressing protein Y (*http://parts.mit.edu/registry/index.php/Help:Tag*). (Degradation) Tags are genetic additions to the end of a sequence which modify expressed proteins in different ways such as marking the protein for faster degradation. Promoters that can accept multiple input transcription factors (called combinatorial promoters) to implement regulation functions that take multiple inputs can be realized by combining the operator sites of several simple promoters [**?**]. For example, the operators $O_{R1}-O_{R2}$ from the λ promoter of the λ bacteriophage can be used as binding sites for the λ transcription factor [36]. Then, the pair $O_{R2}-O_{R1}$ from the 434 promoter from the 434 bacteriophage [10] can be placed at the end of the $O_{R1}-O_{R2}$ sequence from the λ promoter. Depending on the relative positions of these sites and on their distance from the RNA polymerase binding site, the 434 transcription factor may act as a repressor as when this protein is bound to its $O_{R2}-O_{R1}$ sites it prevents the polymerase to bind, while the λ transcription factor may act as an activator.

7.2 Beyond the Modularity Assumption: Retroactivity

In the previous sections, we have outlined a circuit design process, often used in synthetic biology, that relies on the interconnection of well characterized input/output transcriptional modules through suitable static input functions. Examples of designs performed through this process can be found in Chapter 9. It deeply relies on the modularity assumption, by virtue of which the behavior of the obtained circuit topology can be directly predicted by the properties of the composing units. For example, the monotonicity of the input functions of the transcriptional modules composing the repressilator have been a key feature to formally show the existence of periodic solutions. The form of the input functions in the activator-repressor clock design have been key enablers to easily predict the location and number of equilibria as the parameters are changed. The modularity assumption implies that when two modules are connected together, their behavior does not change because of the interconnection. However, a fundamental systems-engineering issue that arises when interconnecting subsystems is how the process of transmitting a signal to a "downstream" component affects the dynamic state of the sending component. Indeed, after designing, testing, and characterizing the input/output behavior of an individual component in isolation, it is certainly desirable if its characteristics do not change substantially when another component is connected to its output channel. This issue, the effect of "loads" on the output of a system, is well-understood in many fields of engineering, for example in electrical circuit design. It has often been pointed out that similar issues arise for biological systems. Alon states that "modules in engineering, and presumably also in biology, have special features that make them easily embedded in almost any system. For example, output nodes should have 'low impedance,' so that adding on additional downstream clients should not drain the output to existing clients (up to some limit)." An extensive review on problems of loads and modularity in signaling networks can be found in [43, 41, 42], where the authors propose concrete analogies with similar problems arising in electrical circuits.

These questions are even more delicate in *synthetic* biology. For example, suppose that we have built a timing device, a clock made up of a network of activation and/or repression interactions among certain genes and proteins, such as the one of diagram c) of Figure 6.1. Next, we want to employ this clock (upstream system) in order to drive one or more components (downstream systems), by using as its *output* signal the oscillating concentration $A(t)$ of the activator. From a systems/signals point of view, $A(t)$ becomes an *input* to the second system (Figure 7.2). The terms "upstream" and "downstream" reflect the direction in which we think of signals as traveling, *from* the clock *to* the systems being synchronized. However, this is only an idealization, because the binding and unbinding of A to promoter sites in a downstream system competes with the biochemical interactions that constitute the upstream block (retroactivity) and may therefore disrupt the operation of the clock

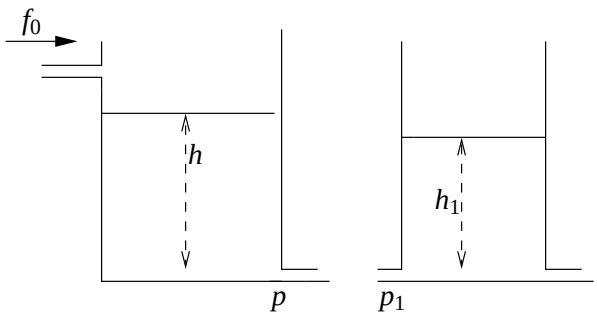


Figure 7.3: On the left, we represent a tank system that takes as input the constant flow f_0 and gives as output the pressure p at the output pipe. On the right, we show a downstream tank.

itself (Figure 7.2). One possible approach to avoid disrupting the behavior of the clock, motivated by the approach used with reporters such as GFP, is to introduce a gene coding for a new protein X, placed under the control of the same promoter as the gene for A, and using the concentration of X, which presumably mirrors that of A, to drive the downstream system. This approach, however, has still the problem that the behavior of the X concentration in time may be altered and even disrupted by the addition of downstream systems that drain X. The net result is still that the downstream systems are not properly timed.

Modeling retroactivity

We broadly call retroactivity the phenomenon by which the behavior of an upstream system is changed upon interconnection to a downstream system. As a simple example, which may be more familiar to an engineering audience, consider the one-tank system shown on the left of Figure 7.3. We consider a constant input flow f_0 as input to the tank system and the pressure p at the output pipe is considered the output of the tank system. The corresponding output flow is given by $k\sqrt{p}$, in which k is a positive constant depending on the geometry of the system. The pressure p is given by (neglecting the atmospheric pressure for simplicity) $p = \rho h$, in which h is the height of the water level in the tank and ρ is water density. Let A be the cross section of the tank, then the tank system can be represented by the equation

$$A\frac{dp}{dt} = \rho f_0 - \rho k\sqrt{p}. \tag{7.1}$$

Let us now connect the output pipe of the same tank to the input pipe of a downstream tank shown on the right of Figure 7.3. Let $p_1 = \rho h_1$ be the pressure generated by the downstream tank at its input and output pipes. Then, the flow at the output of the upstream tank will change and will now be given by $g(p, p_1) = k\sqrt{|p - p_1|}$ if $p > p_1$ and by $g(p, p_1) = -k\sqrt{|p - p_1|}$ if $p \leq p_1$. As a consequence, the time behav-

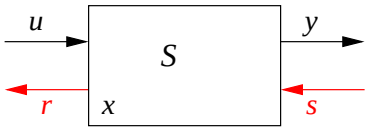


Figure 7.4: A system S input and output signals. The red signals denote signals originating by retroactivity upon interconnection.

ior of the pressure p generated at the output pipe of the upstream tank will change to

$$\begin{aligned} A\frac{dp}{dt} &= \rho f_0 - \rho g(p, p_1) \\ A_1\frac{dp_1}{dt} &= \rho g(p, p_1) - \rho k_1 \sqrt{p_1}, \end{aligned} \tag{7.2}$$

in which A_1 is the cross section of the downstream tank and k_1 is a positive parameter depending on the geometry of the downstream tank. Thus, the input/output response of the tank measured in isolation (equation (7.1)) does not stay the same when the tank is connected through its output pipe to another tank (equation (7.2)). We will model this phenomenon by a signal that travels from downstream to upstream, which we call *retroactivity*. The amount of such a retroactivity will change depending on the features of the interconnection and of the downstream system. For example, if the aperture of the pipe connecting the two tanks is very small compared to the aperture of an output pipe of the downstream tank, the pressure p at the output of the upstream tank will not change much when the downstream tank is connected.

We thus model a system by adding an additional input, called s, to the system to model any change in its dynamics that may occur upon interconnection with a downstream system. Similarly, we add to a system a signal r as another output to model the fact that when such a system is connected downstream of another system, it will send upstream a signal that will alter the dynamics of the upstream system. More generally, we define a system S to have internal state x, two types of inputs (I), and two types of outputs (O): an input "u" (I), an output "y" (O), a *retroactivity to the input* "r" (O), and a *retroactivity to the output* "s" (I) (Figure 7.4). We will thus represent a system S by the equations

$$\dot{x} = f(x,u,s),\ y = Y(x,u,s),\ r = R(x,u,s), \tag{7.3}$$

in which f, Y, R are arbitrary functions and the signals x, u, s, r, y may be scalars or vectors. In such a formalism, we define the input/output model of the isolated system as the one in equations (7.3) without r in which we have also set $s = 0$. Let S_i be a system with inputs u_i and s_i and with outputs y_i and r_i. Let S_1 and S_2 be two systems with disjoint sets of internal states. We define the interconnection of

an upstream system S_1 with a downstream system S_2 by simply setting $y_1 = u_2$ and $s_1 = r_2$. For interconnecting two systems, we require that the two systems do not have internal states in common.

Retroactivity in gene transcriptional circuits

In the previous section, we have defined retroactivity as a general concept modeling the fact that when an upstream system is input/output connected to a downstream one, its dynamic behavior can change. In this section, we focus on transcriptional circuits and show what form the retroactivity takes.

We denote by X the protein, by X (italics) the average protein concentration, and by x (lower case) the gene expressing protein X. A transcriptional component that takes as input protein Z and gives as output protein X is shown in Figure 7.5 in the dashed box. The activity of the promoter controlling gene x depends on the

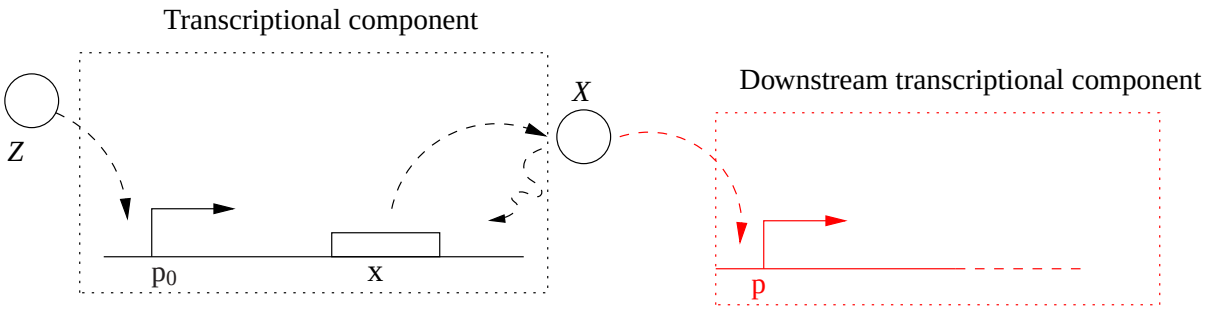


Figure 7.5: The transcriptional component takes as input u protein concentration Z and gives as output y protein concentration X. The transcription factor Z binds to operator sites on the promoter. The red part belongs to a downstream transcriptional block that takes protein concentration X as its input.

amount of Z bound to the promoter. If $Z = Z(t)$, such an activity changes with time. We denote it by $k(t)$. By neglecting the mRNA dynamics, which are not relevant for the current discussion, we can write the dynamics of X as

$$\frac{dX}{dt} = k(t) - \delta X, \tag{7.4}$$

in which δ is the decay rate of the protein. We refer to equation (7.4) as the isolated system dynamics. For the current study, the mRNA dynamics can be neglected because we focus on how the dynamics of X changes when we add downstream systems to which X binds. As a consequence, also the specific form of $k(t)$ is not relevant. Now, assume that X drives a downstream transcriptional module by binding to a promoter p with concentration p (the red part of Figure 7.5). The reversible binding reaction of X with p is given by

$$\mathrm{X}+\mathrm{p} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \mathrm{C},$$

in which C is the complex protein-promoter and k_{on} and k_{off} are the binding and dissociation rates of the protein X to the promoter site p. Since the promoter is not subject to decay, its total concentration p_{TOT} is conserved so that we can write $p + C = p_{TOT}$. Therefore, the new dynamics of X is governed by the equations

$$\begin{aligned} \frac{dX}{dt} &= k(t) - \delta X + \boxed{k_{off}C - k_{on}(p_{TOT} - C)X}, \qquad s = k_{off}C - k_{on}(p_{TOT} - C)X \\ \frac{dC}{dt} &= -k_{off}C + k_{on}(p_{TOT} - C)X, \end{aligned} \tag{7.5}$$

in which the terms in the box represent the signal s, that is, the retroactivity to the output, while the second of equations (7.5) describes the dynamics of the input stage of the downstream system driven by X. Then, we can interpret s as being a mass flow between the upstream and the downstream system. When $s = 0$, the first of equations (7.5) reduces to the dynamics of the isolated system given in equation (7.4). Here, we have assumed that X binds directly to the promoter p. The case in which a signal molecule is needed to transform X to the active form that then binds to p, can be treated in a similar way by considering the additional reversible reaction of X binding to the signal molecule. The end result of adding this reaction is the one of having similar terms in the box of equation (7.5) involving also the signaling molecule concentration.

How large is the effect of the retroactivity s on the dynamics of X and what are the biological parameters that affect it? We focus on the retroactivity to the output s. We can analyze the effect of the retroactivity to the input r on the upstream system by simply analyzing the dynamics of Z in the presence of its binding sites p_0 in Figure 7.5 in a way similar to how we analyze the dynamics of X in the presence of the downstream binding sites p. The effect of the retroactivity s on the behavior of X can be very large (Figure **??**). This is undesirable in a number of situations in which we would like an upstream system to "drive" a downstream one as is the case, for example, when a biological oscillator has to time a number of downstream processes. If, due to the retroactivity, the output signal of the upstream process becomes too low and/or out of phase with the output signal of the isolated system (as in Figure 7.6), the coordination between the oscillator and the downstream processes will be lost. We next propose a procedure to obtain an operative quantification of the effect of the retroactivity on the dynamics of the upstream system.

Quantification of the retroactivity to the output

In this section, we propose a general approach for providing an operative quantification of the retroactivity to the output on the dynamics of the upstream system.

This approach can be generally applied whenever there is a separation of time-scales between the dynamics of the output of the upstream module and the dynamics of the input stage of the downstream module. This separation of time-scales is

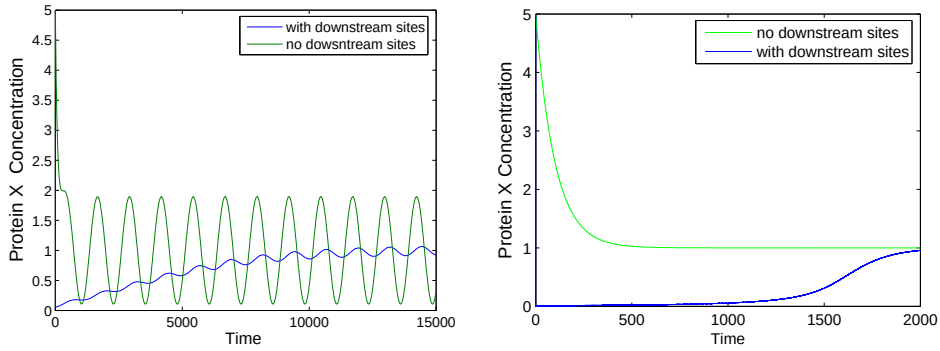


Figure 7.6: The dramatic effect of interconnection. Simulation results for the system in equations (7.5). The green plot (solid line) represents $X(t)$ originating by equations (7.4), while the blue plot (dashed line) represents $X(t)$ obtained by equation (**??**). Both transient and permanent behaviors are different. Here, $k(t) = 0.01(1 + sin(\omega t))$ with $\omega = 0.005$ in the left side plots and $\omega = 0$ in the right side plots, $k_{on} = 10$, $k_{off} = 10$, $\delta = 0.01$, $p_{TOT} = 100$, $X(0) = 5$. The choice of protein decay rate (in min^{-1}) corresponds to a half life of about one hour. The frequency of oscillations is chosen to have a period of about 12 times the protein half life in accordance to what is experimentally observed in the synthetic clock of [6].

always encountered in transcriptional circuits. In fact, the dynamics of the input stage of a downstream system is governed by the reversible binding reaction of the transcription factor with the operator sites. These reactions are often on the time scales of a second and thus are fast compared to the time scales of transcription and translation (often of several minutes) [3]. These determine, in turn, the dynamics of the output of a transcriptional module. Such a separation of time-scales is encountered even when we extend a transcriptional network to include as interconnection mechanisms between transcriptional modules protein-protein interactions (often with a subsecond timescale [46]), as encountered in signal transduction networks.

We quantify the difference between the dynamics of X in the isolated system (equation (7.4)) and the dynamics of X in the connected system (equations (7.5)) by establishing conditions on the biological parameters that make the two dynamics close to each other. This is achieved by exploiting the difference of time scales between the protein production and decay processes and its binding and unbinding process to the promoter p. By virtue of this separation of time scales, we can approximate system (7.5) by a one dimensional system describing the evolution of X on the slow manifold [27]. This reduced system takes the form:

$$\frac{d\bar{X}}{dt} = k(t) - \delta\bar{X} + \bar{s},$$

where $\bar{X}$ is an approximation of X and $\bar{s}$ is an approximation of s, which can be written as $\bar{s} = -\mathcal{R}(\bar{X})(k(t) - \delta\bar{X})$. If $\mathcal{R}(\bar{X})$ is zero, then also $\bar{s} = 0$ and the dynamics of $\bar{X}$ becomes the same as the one of the isolated system (7.4). Since $\bar{X}$ approximates X, the dynamics of X in the full system (7.5) is also close to the dynamics of the isolated system (7.4) whenever $\mathcal{R}(\bar{X}) = 0$. The factor $\mathcal{R}(\bar{X})$ provides then a measure of the retroactivity on the dynamics of X. It is also computable as a function of measurable biochemical parameters and of the signal X traveling across the interconnection, as we next illustrate.

Consider again the full system in equations (7.5), in which the binding and unbinding dynamics is much faster than protein production and decay, that is, $k_{off} \gg k(t)$, $k_{off} \gg \delta$ [3], and $k_{on} = k_{off}/k_d$ with $k_d = O(1)$. Even if the second equation goes to equilibrium very fast compared to the first one, the above system is not in "standard singular perturbation form" [27]. To explicitly model the difference in time scales between the two equations of system (7.5), we introduce a parameter ϵ, which we define as $\epsilon = \delta/k_{off}$. Since $k_{off} \gg \delta$, we also have that $\epsilon \ll 1$. Substituting $k_{off} = \delta/\epsilon$, $k_{on} = \delta/(\epsilon k_d)$, and letting $y = X + C$ (the *total* protein concentration), we obtain the system in singular perturbation form

$$\begin{aligned} \frac{dy}{dt} &= k(t) - \delta(y - C) \\ \epsilon\frac{dC}{dt} &= -\delta C + \frac{\delta}{k_d}(p_{TOT} - C)(y - C). \end{aligned} \tag{7.6}$$

This means, as some authors proposed [**?**], that y (total concentration of protein) is the slow variable of the system (7.5) as opposed to X (concentration of free protein). We can then obtain an approximation of the dynamics of X in the limit in which ϵ is very small, by setting $\epsilon = 0$. This leads to (see [13] for details) the approximated X dynamics

$$\frac{d\bar{X}}{dt} = k(t) - \delta\bar{X} - (k(t) - \delta\bar{X})\frac{d\gamma(\bar{y})}{d\bar{y}}. \tag{7.7}$$

The smaller ϵ, the better is the approximation. Since $\bar{X}$ well approximates X for ϵ small, conditions for which the dynamics of equation (7.7) is close to the dynamics of the isolated system (7.4) also guarantee that the dynamics of X given in system (7.5) is close to the dynamics of the isolated system.

The difference between the dynamics in equation (7.7) (the connected system after a fast transient) and the dynamics in equation (7.4) (the isolated system) is zero when the term $\frac{d\gamma(\bar{y})}{d\bar{y}}$ in equation (7.7) is also zero. We thus consider the factor $\frac{d\gamma(\bar{y})}{d\bar{y}}$ as a quantification of the retroactivity s after a fast transient in the approximation in which $\epsilon \approx 0$. We can also interpret the factor $\frac{d\gamma(\bar{y})}{d\bar{y}}$ as a percentage variation of the dynamics of the connected system with respect to the dynamics of the isolated system at the quasi steady state. We next determine the physical meaning of

such a factor by calculating a more useful expression that is a function of key biochemical parameters. By using the implicit function theorem, one can compute the following expression for $\frac{d\gamma(\bar{y})}{d\bar{y}}$:

$$\frac{d\gamma(\bar{y})}{d\bar{y}} = \frac{1}{1+\frac{(1+\bar{X}/k_d)^2}{p_{TOT}/k_d}} =: \mathcal{R}(\bar{X}), \tag{7.8}$$

in which one can verify that $\mathcal{R}(\bar{X}) < 1$ (see [13] for details). The expression $\mathcal{R}(\bar{X})$ quantifies the retroactivity to the output on the dynamics of X after a fast transient, when we approximate X with $\bar{X}$ in the limit in which $\epsilon \approx 0$. The retroactivity measure is thus low if the affinity of the binding sites p is small (k_d large) or if the signal $X(t)$ is large enough compared to p_{TOT}. Thus, the expression of $\mathcal{R}(\bar{X})$ provides an operative quantification of the retroactivity: such an expression can in fact be evaluated once the association and dissociation constants of X to p are known, the concentration of the binding sites p_{TOT} is known, and the range of operation of the signal $\bar{X}(t)$ that travels across the interconnection is also known.

Therefore, the modularity assumption introduced in Section **??** holds if the value of $\mathcal{R}(\bar{X})$ is low enough. As a consequence, the design of a simple circuit motif such as the ones of Figure 6.1 can assume modularity if the interconnections among the composing modules can be designed so that the value of $\mathcal{R}(\bar{X})$ as given in expression (7.8) is low.

7.3 Insulation Devices to Enforce Modularity

Of course, it is not always possible to design an interconnection such that the retroactivity is low. This is, for example, the case of an oscillator that has to time a downstream load: the load cannot be in general designed and the oscillator must perform well in the face of unknown and possibly variable load properties (Figure 7.2). Therefore, in analogy to what is performed in electrical circuits, one can design a device to be placed between the oscillator and the load so that the device output is not changed by the load and the device does not affect the behavior of the upstream oscillator. Specifically, consider a system S as the one shown in Figure 7.4 that takes u as input and gives y as output. We would like to design it in such a way that (a) the retroactivity r to the input is very small; (b) the effect of the retroactivity s to the output on the internal dynamics of the system is very small independently of s itself; (c) its input/output relationship is about linear. Such a system is said to enjoy the **insulation** property and will be called an insulation component or insulation device. Indeed, such a system will not affect an upstream system because $r \approx 0$ and it will keep the same output signal y *independently* of any connected downstream system. In electronics, amplifiers enjoy the insulation property by virtue of the features of the operational amplifier (op amp) that they employ [45] (Figure 7.7).

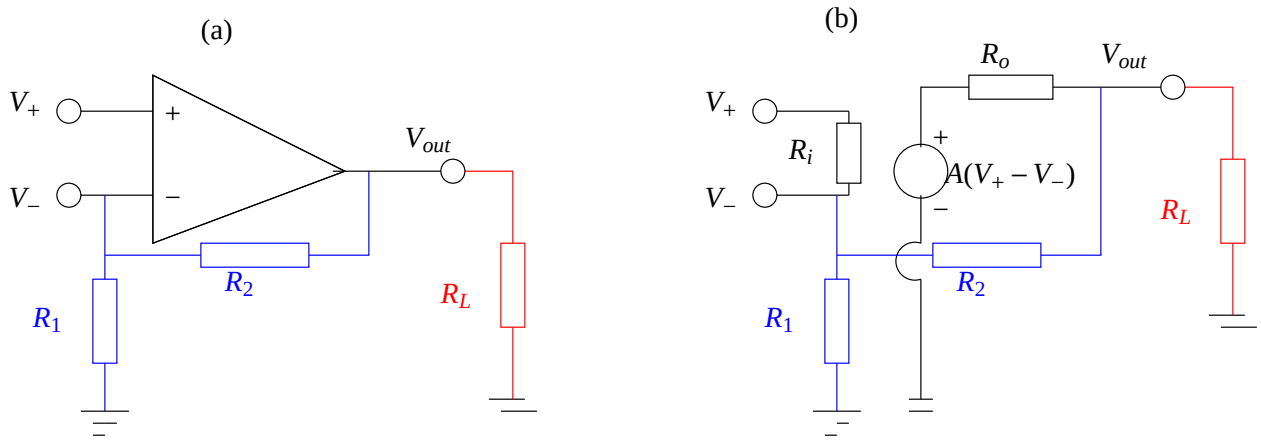


Figure 7.7: In diagram (a), we show the basic non-inverting amplifier circuit that is composed of the op amp plus a feedback circuit. The op amp is the triangular shape that takes as input the differential voltage $V_+ - V_-$ and gives as (open) output $V_{out} = A(V_+ - V_-)$, in which the gain A is infinity in the ideal op amp. The blue circuit components represent the feedback circuit, while the red component is the load. Letting $K = R_1/(R_1 + R_2)$, direct computation leads to $V_{out} \to V_+/K$ as $A \to \infty$. That is, the output voltage does not depend on the load: the retroactivity to the output is almost completely attenuated. In diagram (b), we zoom inside the op amp to show the abstraction of its internal structure. In an ideal op amp, $R_i = \infty$ so that it absorbs almost zero current and any upstream voltage generator will not experience a voltage drop at its output terminals upon interconnection with the amplifier. That is, the retroactivity to the input of the amplifier is almost zero.

The concept of amplifier in the context of a biochemical network has been considered before in relation to its robustness and insulation property from external disturbances ([42] and [41]). Here, we revisit the amplifier mechanism in the context of gene transcriptional networks with the objective of mathematically and computationally proving how suitable biochemical realizations of such a mechanism can attain properties (a), (b), and (c).

Retroactivity to the input

In electronic amplifiers, r is very small because the input stage of an op amp absorbs almost zero current (Figure **??**). This way, there is no voltage drop across the output impedance of an upstream voltage source. Equation (7.8) quantifies the effect of retroactivity on the dynamics of X as a function of biochemical parameters that characterize the interconnection mechanism with a downstream system. These parameters are the affinity of the binding site $1/k_d$, the total concentration of such binding site p_{TOT}, and the level of the signal $X(t)$. Therefore, to reduce the retroactivity, we can choose parameters such that (7.8) is small. A sufficient condition is

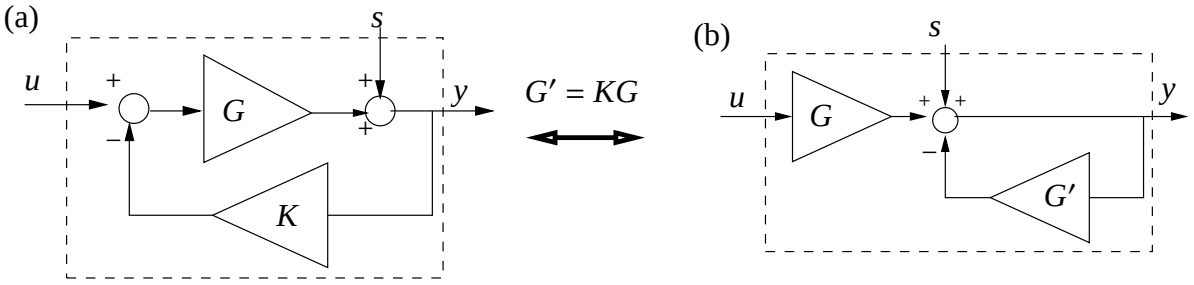


Figure 7.8: Diagram (a) shows the basic feedback/amplification mechanism by which amplifiers attenuate the effect of the retroactivity to the output s. Diagram (b) shows an alternative representation of the same mechanism of diagram (a), which will be employed to design biological insulation devices.

to choose k_d large (low affinity) and p_{TOT} small, for example. Having small value of p_{TOT} and/or low affinity implies that there is a small "flow" of protein X toward its target sites. Thus, we can say that a low retroactivity to the input is obtained when the "input flow" to the system is small. This interpretation establishes a nice analogy to the electrical case, in which low retroactivity to the input is obtained, as explained above, by a low input current. Such an interpretation can be further carried to the hydraulic example. In such an example, if the input flow to the downstream tank is small compared, for example, to the output flow of the downstream tank, the output pressure of the upstream tank will not be affected by the connection. Therefore, the retroactivity to the input of the downstream tank will be small.

Retroactivity to the output

In electronic amplifiers, the effect of the retroactivity to the output s on the amplifier behavior is reduced to almost zero by virtue of a large (theoretically infinite) amplification gain of the op amp and an equally large negative feedback mechanism that regulates the output voltage (Figure 7.7). Genetic realization of amplifiers have been previously proposed (see [38], for example). However, such realizations focus mainly on trying to reproduce the layout of the device instead of implementing the fundamental mechanism that allows it to properly work as an insulator. Such a mechanism can be illustrated in its simplest form by diagram (a) of Figure 7.8, which is very well known to control engineers. For simplicity, we have assumed in such a diagram that the retroactivity s is just an additive disturbance. The reason why for large gains G the effect of the retroactivity s to the output is negligible can be verified through the following simple computation. The output y is given by

$$y = G(u - Ky) + s,$$

which leads to

$$y = u\frac{G}{1+KG} + \frac{s}{1+KG}.$$

As G grows, y tends to u/K, which is independent of the retroactivity s.

Therefore, a central enabler to attenuate the retroactivity effect at the output of a component is to (1) amplify through a large gain the input of the component and (2) to apply a large negative output feedback. We next illustrate this general idea in the context of a simple hydraulic system.

Hydraulic example. Consider the academic hydraulic example consisting of two connected tanks shown in Figure 7.9. The objective is to attenuate the effect of the pressure applied from the downstream tank to the upstream tank, so that the output pressure of the upstream system does not change when the downstream tank is connected. We let the input flow f_0 be amplified by a large factor G. Also, we

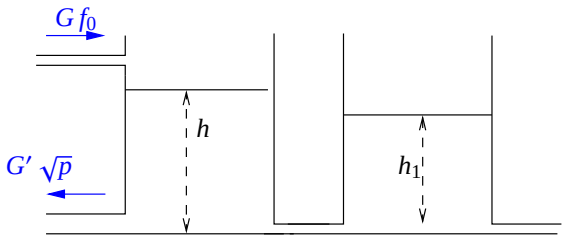


Figure 7.9: We amplify the input flow f_0 through a large gain G and we apply a large negative feedback by employing a large output pipe with output flow $G'\sqrt{p}$.

consider a large pipe in the upstream tank with output flow $G'\sqrt{p}$, with $G' \gg k$ and $G' \gg k_1$. Let p be the pressure at the output pipe of the upstream tank and p_1 the pressure at the bottom of the downstream tank. One can verify that the only equilibrium value for the pressure p at the output pipe of the upstream tank is obtained for $p > p_1$ and it is given by

$$p_{eq} = \left(\frac{Gf_0}{G' + (kk_1)/\sqrt{k_1^2+k^2}}\right)^2.$$

If we let G' be sufficiently larger than k_1 and k and we let $G' = KG$ for some positive $K = O(1)$, then for G sufficiently large $p_{eq} \approx (f_0/K)^2$, which does not depend on the presence of the downstream system. In fact, it is the same as the equilibrium value of the isolated upstream system $A\frac{dp}{dt} = \rho G f_0 - \rho G'\sqrt{p} - \rho k\sqrt{p}$ for G sufficiently large and for $G' = KG$ with $K = O(1)$.

Coming back to the transcriptional example, consider the approximated dynamics of equation (7.7) for X. Let us thus assume that we can apply a gain G to the

input $k(t)$ and a negative feedback gain G' to X with $G' = KG$. This leads to the new differential equation for the connected system (7.7) given by

$$\frac{dX}{dt} = (Gk(t) - (G' + \delta)X)(1 - d(t)), \tag{7.9}$$

in which we have defined $d(t) := \frac{d\gamma(y)}{dy}$, where $y(t)$ is given by the reduced system $\frac{dy}{dt} = Gk(t) - (G' + \delta)(y - \gamma(y))$. It can be shown (see [**?**] for details) that as G and thus as G' grow, the signal $X(t)$ generated by the connected system (7.9) becomes close to the solution $X(t)$ of the isolated system

$$\frac{dX}{dt} = Gk(t) - (G' + \delta)X, \tag{7.10}$$

that is, the presence of the disturbance term $d(t)$ will not significantly affect the time behavior of $X(t)$. Since $d(t)$ is a measure of the retroactivity effect on the dynamics of X, such an effect is thus attenuated by employing large gains G and G'. *How can we obtain a large amplification gain G and a large negative feedback G′ in a biological insulation component?* This question is addressed in the following chapter, in which we show two possible realizations of insulation devices.

7.4 Design of genetic circuits under the modularity assumption

Based on the modeling assumptions introduced in Chapter 2 and on the tools for studying the dynamics of a nonlinear system introduced in Chapter 3, a number of synthetic genetic circuits have been designed and fabricated by composing transcriptional modules through input/output connection (Figure 6.1). Through such a design procedure one seeks to predict the behavior of a circuit by the behavior of the composing units, once these have been well characterized in isolation. This approach is standard also in the design and fabrication of electronic circuitry.

The repressilator

Elowitz and Leibler [14] constructed the first operational oscillatory genetic circuit consisting of three repressors arranged in ring fashion, and coined it the "repressilator" (See diagram d) of Figure 6.1). The repressilator exhibits sinusoidal, limit cycle oscillations in periods of hours. The dynamical model of the repressilator can be thus obtained my composing three transcriptional modules in a loop fashion through input functions as in expression (**??**). Re-arranging the parameters, it can

thus be described by

$$\begin{aligned} \dot{r}_A &= -\delta r_A + f_1(C) \\ \dot{A} &= r_A - \delta A \\ \dot{r}_B &= -\delta r_B + f_2(A) \\ \dot{B} &= r_B - \delta B \\ \dot{r}_C &= -\delta r_C + f_3(B) \\ \dot{C} &= r_C - \delta C, \end{aligned} \tag{7.11}$$

where we consider two different cases for the shape of the input functions f_i: three identical repressions (the symmetric case) or two identical activations and one repression (the non-symmetric case). For the symmetric case, we thus assume that

$$f_1(p) = f_2(p) = f_3(p) = \frac{\alpha^2}{1+p^n}.$$

Since the regulation functions have all negative slope, and there is an odd number of them in the loop, there is only one equilibrium. One can then invoke Mallet-Paret's Theorem [29] or Hastings' Theorem [**?**] (see Chapter 3 for the details) to conclude that if the equilibrium point is unstable, the system admits a non-constant periodic orbit (see [**?**] for a detailed application of these theorems). Thus, one can search for parameter values to guarantee the instability of the equilibrium point. This procedure was followed by Elowitz in the design of the repressilator [14]. In particular, one can show that the repressilator in equations (7.11) has a periodic solution for the ratio α/δ satisfying the relation

$$\alpha^2/\delta^2 > \sqrt[n]{\frac{4/3}{n-4/3}}\left(1+\frac{4/3}{n-4/3}\right).$$

For the proof of this statement, the reader is referred to [**?**]. This relationship is plotted in the left plot of Figure 7.10. When n increases, the existence of an unstable equilibrium point is guaranteed for larger ranges of the other parameter values. Equivalently, for fixed values of α and δ, as n increases the robustness of the circuit oscillatory behavior to parametric variations in the values of α and δ increases. Of course, this "behavioral" robustness does not guarantee that other important features of the oscillator, such as the period value, are slightly changed when parameters vary. Numerical studies indicated that the period T approximatively follows $T \propto \frac{1}{\delta}$, and varies only little with α (right plot of Figure 7.10). From the figure, we can note that as the value of δ increases, the sensitivity of the period to the variation of δ itself decreases. However, increasing δ would necessitate the increase of the cooperativity n, therefore indicating a possible trade off that should be taken into account in the design process in order to balance the system complexity and robustness of the oscillations.

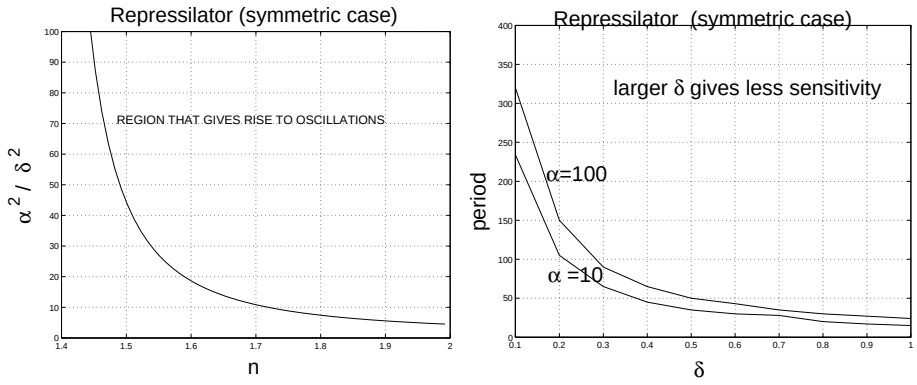


Figure 7.10: (Left) Space of parameters that give rise to oscillations for the repressilator in equations (**??**). (Right) Period as a function of δ and α.

A similar result for the existence of a periodic solution can be obtained for the non-symmetric case in which the input functions of the three transcriptional modules are modified to

$$\begin{aligned} f_1(p) &= \frac{\alpha_3^2}{1+p^n} \\ f_2(p) &= \frac{\alpha^2 p^n}{1+p^n} \\ f_3(p) &= \frac{\alpha^2 p^n}{1+p^n}, \end{aligned}$$

that is, two interactions are activations and one only is a repression. One can verify that there is one equilibrium point only and again invoke Mallet-Paret's Theorem [29] or Hastings' Theorem [**?**] to conclude that if the equilibrium point is unstable, the system admits a non-constant periodic solution. We can thus obtain the condition for oscillations again by establishing conditions on the parameters that guarantee an unstable equilibrium. These conditions are reported in Figure 7.11 (see [**?**] for the detailed derivations). One can conclude that it is possible to "over design" the circuit to be in the region of parameter space that gives rise to oscillations. It is also possible to show that increasing the number of elements in the oscillatory loop, the value of n sufficient for oscillatory behavior decreases. The design criteria for obtaining oscillatory behavior are thus summarized in Figures 7.10 and 7.11.

The activator-repressor clock

Consider the activator-repressor clock diagram shown in Figure 6.1 c). The transcriptional module for A has an input function that takes two inputs: an activator

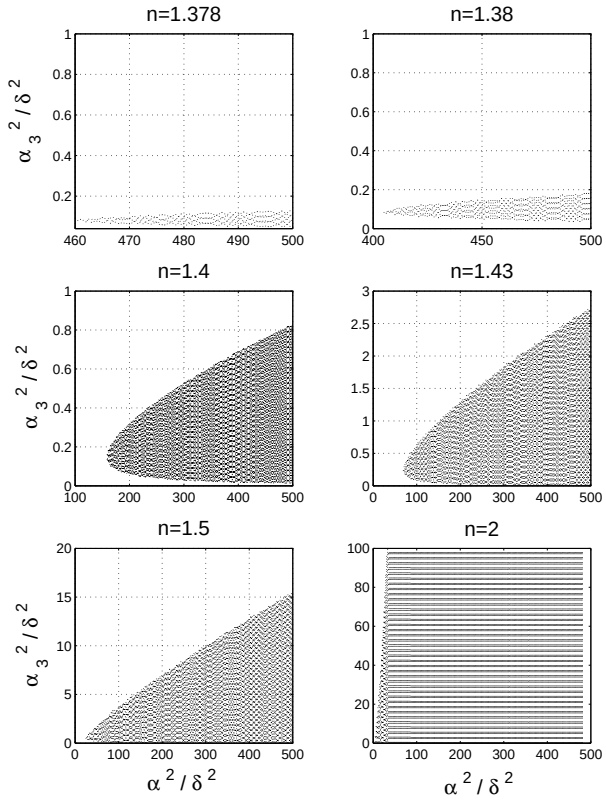


Figure 7.11: Space of parameters that give rise to oscillations for the repressilator (non-symmetric case).

A and a repressor B. The transcriptional module B has an input function that takes only an activator A as its input. Let r_A and r_B represent the concentration of m-RNA of the activator and of the repressor, respectively. Let A and B denote the protein concentration of the activator and of the repressor, respectively. Then, we consider the following four-dimensional model describing the rate of change of the species concentrations:

$$\begin{aligned} \dot{r}_A &= -\delta_1/\epsilon\, r_A + F_1(A,B) \\ \dot{A} &= \nu(-\delta_A A + k_1/\epsilon\, r_A) \\ \dot{r}_B &= -\delta_2/\epsilon\, r_B + F_2(A) \\ \dot{B} &= -\delta_B B + k_2/\epsilon\, r_B, \end{aligned} \tag{7.12}$$

in which the parameter ν regulates the difference of time-scales between the repressor and the activator dynamics, ϵ is a parameter that regulates the difference of time-scales between the m-RNA and the protein dynamics. The parameter ϵ deter-

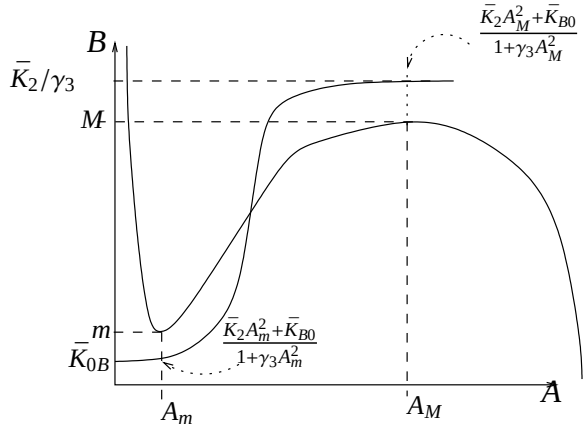


Figure 7.12: Shape of the curves in the A, B plane corresponding to $\dot{r}_B = 0, \dot{B} = 0$ and to $\dot{r}_A = 0, \dot{A} = 0$ as function of the parameters. Letting, $\bar{K}_1 = K_1(k_1/(\delta_1\delta_A))$, $\bar{K}_{A0} = K_{A0}(k_1/(\delta_1\delta_A))$, $\bar{K}_2 = K_2(k_2/(\delta_2\delta_B))$, $\bar{K}_{B0} = K_{B0}(k_2/(\delta_2\delta_B))$, we have $A_m = \frac{\bar{K}_1}{6\gamma_1}\left(1-(cos(\phi/3)-\sqrt{3}sin(\phi/3))\right)$, $A_M = \frac{\bar{K}_1}{6\gamma_1} + \frac{\bar{K}_1}{3\gamma_1}cos(\phi/3)$, $\phi = \text{atan}\left(\frac{\sqrt{\frac{27\bar{K}_{A0}}{4\gamma_1^2}(\frac{\bar{K}_1^3}{\gamma_1^2}-27\bar{K}_{A0})}}{\frac{\bar{K}_1^3}{4\gamma_1^3}-27\frac{\bar{K}_{A0}}{2\gamma_1}}\right)$, $m = \sqrt{\frac{\bar{K}_1 A_m^2+\bar{K}_{A0}-A_m(1+\gamma_1 A_m^2)}{\gamma_2 A_m}}$, $M = \sqrt{\frac{\bar{K}_1 A_M^2+\bar{K}_{A0}-A_M(1+\gamma_1 A_M^2)}{\gamma_2 A_M}}$.

mines how close model (7.12) is to a two-dimensional model in which the m-RNA dynamics are considered at the equilibrium. Thus, ϵ is a singular perturbation parameter (equations (7.12) can be taken to standard singular perturbation form by considering the change of variables $\bar{r}_A = r_A/\epsilon$ and $\bar{r}_B = r_B/\epsilon$). The details on singular perturbation can be found in Chapter 3. The functions F_1 and F_2 are the input functions and are given by

$$\begin{aligned} F_1(A,B) &= \frac{K_1 A^n + K_{A0}}{1+\gamma_1 A^n + \gamma_2 B^n} \\ F_2(A) &= \frac{K_2 A^n + K_{B0}}{1+\gamma_3 A^n}, \end{aligned}$$

in which K_1 and K_2 are the maximal transcription rates, while K_{A0} and K_{B0} are the basal transcription rates when no activator is present. The Hill coefficient n is chosen here to be $n = 2$. The values of ϵ and of ν do not affect the number of equilibria of the system, while the values of the other parameters are the ones that control the number of equilibria. The set of values of $K_i, k_i, \delta_i, \gamma_i, \delta_A, \delta_B$ that allow the existence of a unique equilibrium can be determined by employing graphical techniques. In particular, we can plot the curves corresponding to the sets of A, B values for which $\dot{r}_B = 0$ and $\dot{B} = 0$ and the set of A, B values for which $\dot{r}_A = 0$ and $\dot{A} = 0$ as in Figure 7.12. The intersection of these two curves provides the equilibria of the system and

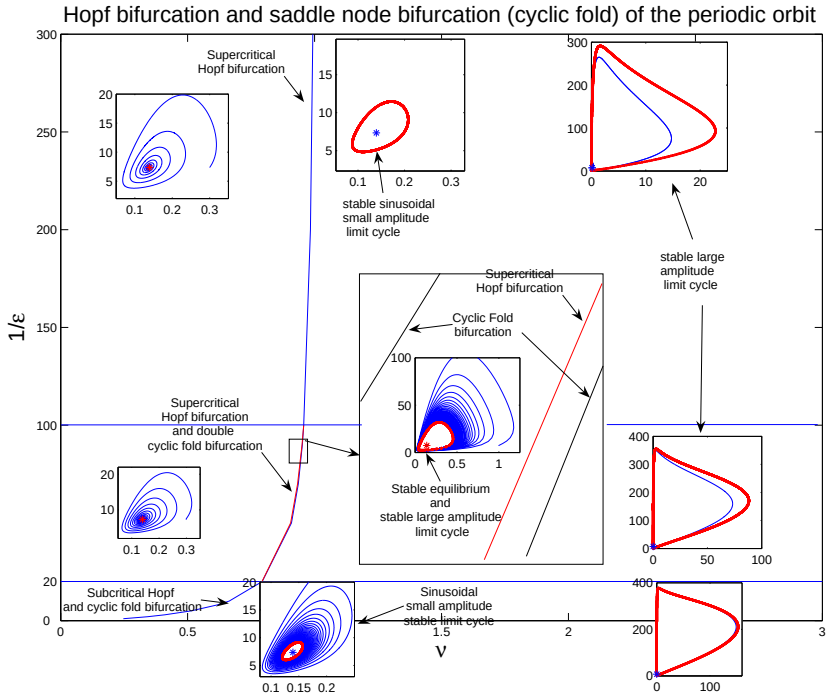


Figure 7.13: Design chart for the relaxation oscillator. One obtains sustained oscillations passed the Hopf bifurcation, for values of ν sufficiently large independently of the difference of time scales between the protein and the mRNA dynamics. We also notice that there are values of ν for which a stable equilibrium point and a stable orbit coexist and values of ν for which two stable orbits coexist. The interval of ν values for which two stable orbits coexist is too small to be able to numerically set ν in such an interval. Thus, this interval is not practically relevant. The values of ν for which a stable equilibrium and a stable periodic orbit coexist is instead relevant. This situation corresponds to the *hard excitation* condition [**?**] and occurs for realistic values of the separation of time-scales between protein and m-RNA dynamics. Therefore, this simple oscillator motif described by a four-dimensional model can capture the features that lead to the long term suppression of the rhythm by external inputs. *Birhythmicity* [**?**] is also possible even if practically not relevant due to the numerical difficulty of moving the system to one of the two periodic orbits. For more details, the reader is referred to [**?**, **?**].

conditions on the parameters can be determined that guarantee the existence of one equilibrium only. In particular, we require that the basal activator transcription rate when B is not present, which is proportional to $\bar{K}_{A0}$, is sufficiently smaller than the maximal transcription rate of the activator, which is proportional to $\bar{K}_1$. Also, $\bar{K}_{A0}$ must be non-zero. Also, in case $\bar{K}_1 >> \bar{K}_{A0}$, one can verify that $A_M \approx \bar{K}_1/2\gamma_1$ and thus $M \approx \bar{K}_1/2\sqrt{\gamma_1\gamma_2}$. As a consequence, if $\bar{K}_1/\gamma_1$ increases then so must do

$\bar{K}_2/\gamma_3$. Finally, $A_m \approx 0$, and $m \approx \sqrt{K_{A0}/\gamma_2 A_m}$. As a consequence, the smaller $\bar{K}_{A0}$ becomes, the smaller $\bar{K}_{B0}$ must be (see [**?**] for more details). Assume that the values of $K_i, k_i, \delta_i, \gamma_i, \delta_A, \delta_B$ have been chosen so that there is a unique equilibrium and we numerically study the occurrence of periodic solutions as the difference in time-scales between protein and m-RNA, ϵ, and the difference in time-scales between activator and repressor, ν, are changed. In particular, we perform bifurcation analysis with ϵ and ν the two bifurcation parameters.
These bifurcation results are summarized by Figure 7.13. The reader is referred to [**?**] for the details of the numerical analysis. In terms of the ϵ and ν parameters, it is thus possible to "over design" the system: if the activator dynamics is sufficiently sped up with respect to the repressor dynamics, the system parameters move across a Hopf bifurcation (Hopf bifurcation was introduced in Chapter 3) and stable oscillations will arise. From a fabrication point of view, the activator dynamics can be sped up by adding suitable degradation tags to the activator protein. The region of the parameter space in which the system exhibits almost sinusoidal damped oscillations is on the left-hand side of the curve corresponding to the Hopf bifurcation. Since the data of [6] exhibits almost sinusoidal damped oscillations, it is possible that the clock is operating in a region of parameter space on the "left" of the curve corresponding to the Hopf bifurcation. If this were the case, increasing the separation of time-scales between the activator and the repressor, ν, may lead to a stable limit cycle.

7.5 Biological realizations of an insulation component

In the previous section, we have proposed a general mechanism in order to create an insulation component. In particular, we have specified how one can alter the biological features of the interconnection mechanism in order to have low retroactivity to the input r and we have shown a general method to attenuate the retroactivity to the output s. Such a method consists of a large amplification of the input and a large negative output feedback. The insulation component will be inserted in place of the transcriptional component of Figure 7.5. This will guarantee that the system generating Z, an oscillator, for example, will maintain the same behavior as in isolation and also that the downstream system that accepts X as its input will not alter the behavior of X. The net result of this is that the oscillator generating signal Z will be able to time downstream systems with the desired phase and amplitude independently of the number and the features of downstream systems. In this section, we determine two possible biological mechanisms that can be exploited to obtain a large amplification gain to the input Z of the insulation component and a large negative feedback on the output X of the insulation component. Both mechanisms realize the negative feedback through enhanced degradation. The first design realizes amplification through transcriptional activation, while the second design through phosphorylation of a protein that is in abundance in the system.

Design 1: Amplification through transcriptional activation

In this design, we obtain a large amplification of the input signal $Z(t)$ by having promoter p_0 (to which Z binds) be a strong, non leaky, promoter. The negative feedback mechanism on X relies on enhanced degradation of X. Since this must be large, one possible way to obtain an enhanced degradation for X is to have a protease, called Y, be expressed by a strong constitutive promoter. The protease Y will cause a degradation rate for X, which is larger if Y is more abundant in the system. This design is schematically shown in Figure 7.14.

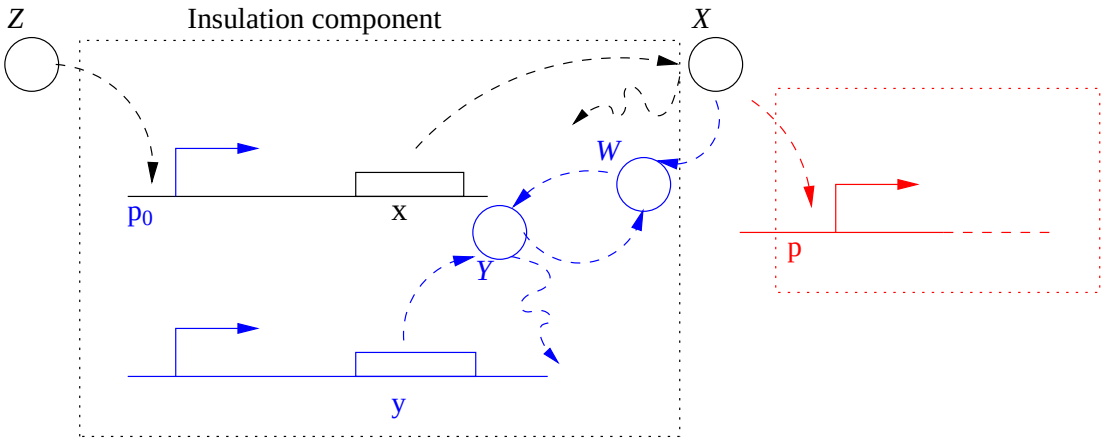


Figure 7.14: We highlight in blue the parts that Design 1 affects. In particular, a negative feedback occurring through post-translational regulation and a promoter that produces a large signal amplification are the central parts of this design. The red part indicates the downstream component that takes as input the concentration of protein X.

In order to investigate whether such a design realizes a large amplification and a large negative feedback on X as needed, we analyze the full input/output model for the block in the dashed box of Figure 7.14. In particular, the expression of gene x is assumed to be a two-step process, which incorporates also the mRNA dynamics. Incorporating these dynamics in the model is relevant for the current study because they may contribute to an undesired delay between the Z and X signals. The reaction of the protease Y with protein X is modeled as the two-step reaction

$$\mathrm{X} + \mathrm{Y} \underset{\eta_2}{\overset{\eta_1}{\rightleftharpoons}} \mathrm{W} \overset{\beta}{\rightarrow} \mathrm{Y},$$

which can be found in standard references (see [**?**], for example). The input/output system model of the insulation component that takes Z as an input and gives X as

an output is given by the following equations

$$\frac{dZ}{dt} = k(t) - \delta Z + \boxed{k_{-}Z_p - k_{+}Z(p_{0,TOT} - Z_p)} \tag{7.13}$$

$$\frac{dZ_p}{dt} = k_{+}Z(p_{0,TOT} - Z_p) - k_{-}Z_p \tag{7.14}$$

$$\frac{dm_X}{dt} = GZ_p - \delta_1 m_X \tag{7.15}$$

$$\frac{dX}{dt} = \nu m_X - \eta_1 YX + \eta_2 W - \delta_2 X + \boxed{k_{off}C - k_{on}X(p_{TOT} - C)} \tag{7.16}$$

$$\frac{dW}{dt} = \eta_1 XY - \eta_2 W - \beta W \tag{7.17}$$

$$\frac{dY}{dt} = -\eta_1 YX + \beta W + \alpha G - \gamma Y + \eta_2 W \tag{7.18}$$

$$\frac{dC}{dt} = -k_{off}C + k_{on}X(p_{TOT} - C), \tag{7.19}$$

in which we have assumed that the expression of gene z is controlled by a promoter with activity $k(t)$. These equations will be studied numerically and analyzed mathematically in a simplified form. The variable Z_p is the concentration of protein Z bound to the promoter controlling gene x, $p_{0,TOT}$ is the total concentration of the promoter p_0 controlling gene x, m_X is the concentration of messenger RNA of X, C is the concentration of X bound to the downstream binding sites with total concentration p_{TOT}, γ is the decay rate of the protease Y. The value of G is the production rate of X mRNA per unit concentration of Z bound to the promoter controlling x; the promoter controlling gene y has strength αG, for some constant α, and it has the same order of magnitude strength as the promoter controlling x. The terms in the box in equation (7.13) represent the retroactivity r to the input of the insulation component in Figure **??**. The terms in the box in equation (7.16) represent the retroactivity s to the output of the insulation component of Figure **??**. The dynamics of equations (7.13)–(7.19) without s (the elements in the box in equation (7.16)) describe the dynamics of X with no downstream system.

We mathematically explain why system (7.13)–(7.19) allows to attenuate the effect of s on the X dynamics. Equations (7.13) and (7.14) simply determine the signal $Z_p(t)$ that is the input to equations (7.15)–(7.19). For the discussion regarding the attenuation of the effect of s, it is not relevant what the specific form of signal $Z_p(t)$ is. Let then $Z_p(t)$ be any bounded signal $\nu(t)$. Since equation (7.15) takes $\nu(t)$ as an input, we will have that $m_X = G\bar{\nu}(t)$, for a suitable signal $\bar{\nu}(t)$. Let us assume for the sake of simplifying the analysis that the protease reaction is a one step reaction, that is, $X + Y \xrightarrow{\beta} Y$. Therefore, equation (7.18) simplifies to $\frac{dY}{dt} = \alpha G - \gamma Y$ and equation (7.16) simplifies to $\frac{dX}{dt} = \nu m_X - \beta YX - \delta_2 X + k_{off}C - k_{on}X(p_{TOT} - C)$. If we consider the protease to be at its equilibrium, we have that $Y(t) = \alpha G/\gamma$. As a

consequence, the X dynamics becomes

$$\frac{dX}{dt} = \nu G\bar{v}(t) - (\beta\alpha G/\gamma + \delta_2)X + \boxed{k_{off}C - k_{on}X(p_{TOT} - C)},$$

with C determined by equation (7.19). By using the same singular perturbation argument employed in the previous section, we obtain that the dynamics of X will be after a fast transient approximatively given by

$$\frac{dX}{dt} = (\nu G\bar{v}(t) - (\beta\alpha G/\gamma + \delta_2)X)(1 - d(t)), \tag{7.20}$$

in which $0 < d(t) < 1$ is the effect of the retroactivity s. Then, as G increases, $X(t)$ becomes closer to the solution of the isolated system

$$\frac{dX}{dt} = \nu G\bar{v}(t) - (\beta\alpha G/\gamma + \delta_2)X,$$

as explained in Section **??**[1].

We now turn to the question of minimizing the retroactivity to the input r because its effect can alter the input signal $Z(t)$. In order to decrease r, we guarantee that the retroactivity measure given in equation (**??**) is small. This is seen to be true if $(\bar{k}_d + Z)^2/(p_{0,TOT}\bar{k}_d)$ is very large, in which $1/\bar{k}_d = k_+/k_-$ is the affinity of the binding site p_0 to Z. Since after a short transient, $Z_p = (p_{0,TOT}Z)/(\bar{k}_d + Z)$, for Z_p not to be a distorted version of Z, it is enough to ask that $k_d \gg Z$. This, combined with the requirement that $(\bar{k}_d + Z)^2/(p_{0,TOT}\bar{k}_d)$ is very large, leads to the requirement $p_{0,TOT}/\bar{k}_d \ll 1$. Summarizing, for not having distortion effects between Z and Z_p and small retroactivity r, we need that

$$\bar{k}_d \gg Z \text{ and } p_{0,TOT}/\bar{k}_d \ll 1. \tag{7.21}$$

Simulation results. Simulation results are presented for the insulation system of equations (7.13)–(7.19) as the mathematical analysis of such a system is only valid under the approximation that the protease reaction is a one step reaction. In all simulations, we consider protein decay rates to be $0.01min^{-1}$ to obtain a protein half life of about one hour. We consider always a periodic forcing $k(t) = 0.01(1 + \sin(\omega t))$, in which we assume that such a periodic signal has been generated by a synthetic biological oscillator. Therefore, the oscillating signals are chosen to have a period that is about 12 times the protein half life in accordance to what is experimentally observed in the synthetic clock of [6]. All simulation results were obtained by using MATLAB (Simulink), with variable step ODE solver ODE23s. For large gains ($G = 1000$, $G = 100$), the performance considerably improves compared to the case in which X was generated by a plain transcriptional component accepting Z as an input (Figure 7.6). For lower gains ($G = 10$, $G = 1$), the performance starts to degrade for $G = 10$ and becomes not acceptable for $G = 1$ (Figure

[1]See the supplementary material for the mathematical details.

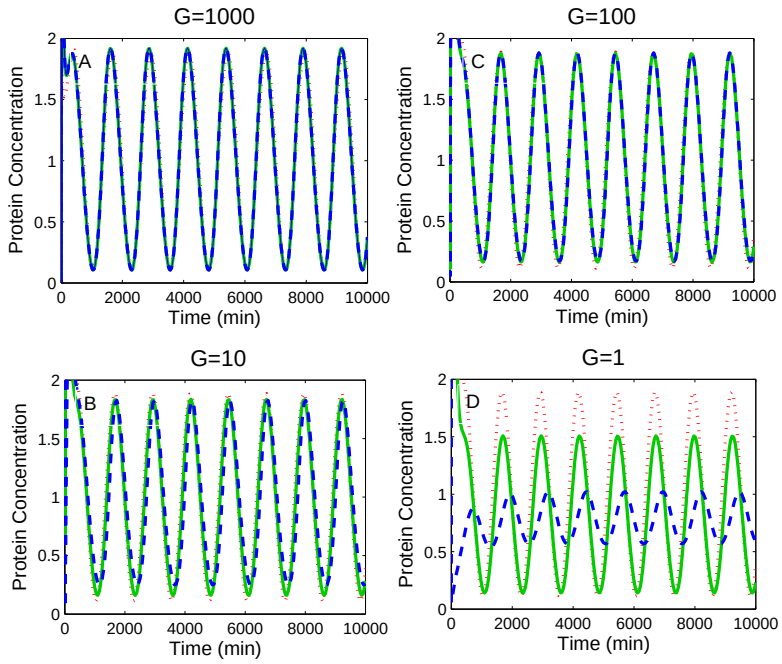


Figure 7.15: Design 1: results for different gains G. In all plots, red (dotted line) is the input Z to the insulation device, green (solid line) is the output X of the insulation device in isolation (without the downstream binding sites p), blue (dashed line) is the output X of the insulation device when downstream sites p are present. In all plots, $k(t) = 0.01(1 + sin(\omega t))$, $p_{TOT} = 100$, $k_{off} = k_{on} = 10$, $\delta = 0.01$, and $\omega = 0.005$. The parameter values are $\delta_1 = 0.01$, $p_{0,TOT} = 1$, $\eta_1 = \eta_2 = \beta = \gamma = 0.01$, $k_- = 200$, $k_+ = 10$, $\alpha = 0.1$, $\delta_2 = 0.1$, $\nu = 0.1$, and $G = 1000, 100, 10, 1$. The retroactivity to the output is not well attenuated for values of the gain $G = 1$ and the attenuation capability begins to worsen for $G = 10$.

7.15). Since we can view G as the number of transcripts produced per unit time (one minute) per complex of protein Z bound to promoter p_0, values $G = 100, 1000$ may be difficult to realize *in vivo*, while the values $G = 10, 1$ could be more easily realized. The values of the parameters chosen in Figure 7.15 are such that $\bar{k_d} \gg Z$ and $p_{0,TOT} \ll \bar{k_d}$. This is enough to guarantee that there is small retroactivity r to the input of the insulation device independently of the value of the gain G, according to relations (7.21). The poorer performance of the device for $G = 1$ is therefore entirely due to poor attenuation of the retroactivity s to the output.

Design 2: Amplification through phosphorylation

In this design, the amplification of Z is obtained by having Z activate the phosphorylation of a protein X, which is available in the system in abundance. That is,

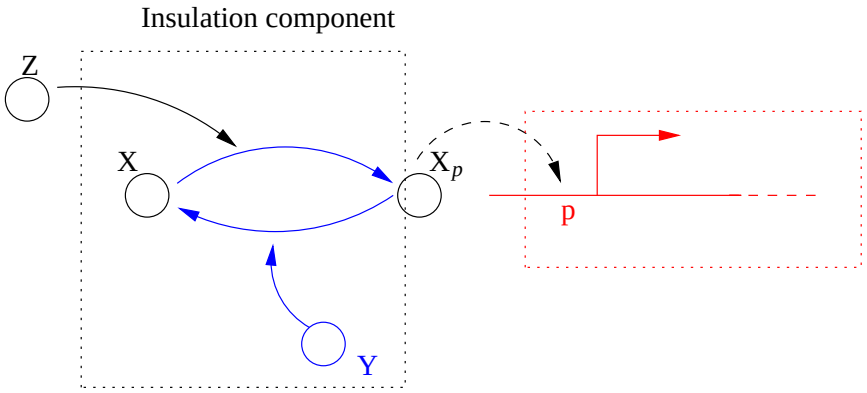


Figure 7.16: The dashed box contains the insulation device. The blue parts highlight the mechanism that provides negative feedback and amplification. Negative feedback occurs through a phosphatase Y that converts the active form X_p back to its inactive form X. Amplification occurs through Z activating the phosphorylation of X.

Z is a kinase for a protein X. The phosphorylated form of X, called X_p, binds to the downstream sites, while X does not. A negative feedback on X_p is obtained by having a phosphatase Y activate the dephosphorylation of protein X_p. Protein Y is also available in abundance in the system. This mechanism is depicted in Figure 7.16. A similar design has been proposed by [42, 41], in which a MAPK cascade plus a negative feedback loop that spans the length of the MAPK cascade is considered as a feedback amplifier. Our design is much simpler as it involves only one phosphorylation cycle and does not require the additional feedback loop. In fact, we realize a strong negative feedback by the action of the phosphatase that converts the active protein form X_p to its inactive form X. This negative feedback, whose strength can be tuned by varying the amount of phosphatase in the system, is enough to mathematically and computationally show that the desired insulation properties are satisfied.

We consider two different models for the phosphorylation and dephosphorylation processes. A one step reaction model is initially considered to illustrate what biochemical parameters realize the input gain G and the negative feedback G'. Then, we turn to a more realistic two step model to perform a parametric analysis and numerical simulation. The one step model that we consider is the one of [22]:

$$Z + X \xrightarrow{k_1} Z + X_p,$$

and

$$Y + X_p \xrightarrow{k_2} Y + X.$$

We assume that there is plenty of protein X and of phosphatase Y in the system and that these quantities are conserved. The conservation of X gives $X + X_p + C = X_{TOT}$, in which X is the inactive protein, X_p is the phosphorylated protein that binds to

the downstream sites p, and C is the complex of the phosphorylated protein X_p bound to the promoter p. The X_p dynamics can be described by the first equation in the following model

$$\frac{dX_p}{dt} = k_1 X_{TOT} Z(t)\left(1 - \frac{X_p}{X_{TOT}} - \boxed{\frac{C}{X_{TOT}}}\right) - k_2 Y X_p + \boxed{k_{off} C - k_{on} X_p (p_{TOT} - C)} \quad (7.22)$$

$$\frac{dC}{dt} = -k_{off} C + k_{on} X_p (p_{TOT} - C). \quad (7.23)$$

The boxed terms represent the retroactivity s to the output of the insulation system of Figure 7.16. For a weakly activated pathway ([22]), $X_p \ll X_{TOT}$. Also, if we assume that the concentration of total X is large compared to the concentration of the downstream binding sites, that is, $X_{TOT} \gg p_{TOT}$, equation (7.22) is approximatively equal to

$$\frac{dX_p}{dt} = k_1 X_{TOT} Z(t) - k_2 Y X_p + k_{off} C - k_{on} X_p (p_{TOT} - C).$$

Denote $G = k_1 X_{TOT}$ and $G' = k_2 Y$. Exploiting again the difference of time scales between the X_p dynamics and the C dynamics, after a fast initial transient, the dynamics of X_p can be well approximated by

$$\frac{dX_p}{dt} = (GZ(t) - G' X_p)(1 - d(t)), \quad (7.24)$$

in which $0 < d(t) < 1$ is the effect of the retroactivity s to the output after a short transient. Therefore, for G and G' large enough, $X_p(t)$ tends to the solution $X_p(t)$ of the isolated system $\frac{dX_p}{dt} = GZ(t) - G' X_p$, as explained in Section **??**[2]. As a consequence, the effect of the retroactivity to the output s is attenuated by increasing $k_1 X_{TOT}$ and $k_2 Y$ enough. That is, to obtain large input and feedback gains, one should have large phosphorylation/dephosphorylation rates and/or a large amount of protein X and phosphatase Y in the system. This reveals that the values of the phosphorylation/dephosphorylation rates cover an important role toward the realization of the insulation property of the module of Figure **??**.

We next consider a more complex model for the phosphorylation and dephosphorylation reactions and perform a parametric analysis to highlight the roles of the various parameters for attaining the insulation properties. In particular, we consider a two-step reaction model such as those in [23]. According to this model, we have the following two reactions for phosphorylation and dephosphorylation, respectively:

$$\mathrm{X} + \mathrm{Z} \underset{\beta_2}{\overset{\beta_1}{\rightleftarrows}} \mathrm{C}_1 \xrightarrow{k_1} \mathrm{X}_p + \mathrm{Z}, \quad (7.25)$$

and

$$\mathrm{Y} + \mathrm{X}_p \underset{\alpha_2}{\overset{\alpha_1}{\rightleftarrows}} \mathrm{C}_2 \xrightarrow{k_2} \mathrm{X} + \mathrm{Y}, \quad (7.26)$$

[2] See the supplementary material for the mathematical details.

in which C_1 is the [protein X/kinase Z] complex and C_2 is the [phosphatase Y/protein X_p] complex. Additionally, we have the conservation equations $Y_{TOT} = Y + C_2$, $X_{TOT} = X + X_p + C_1 + C_2 + C$, because proteins X and Y are not degraded. Therefore, the differential equations modeling the insulation system of Figure 7.16 become

$$\frac{dZ}{dt} = k(t) - \delta Z \boxed{-\beta_1 Z X_{TOT}(1 - \frac{X_p}{X_{TOT}} - \frac{C_1}{X_{TOT}} - \frac{C_2}{X_{TOT}} - \boxed{\frac{C}{X_{TOT}}}) + (\beta_2 + k_1)C_1} \quad (7.27)$$

$$\frac{dC_1}{dt} = -(\beta_2 + k_1)C_1 + \beta_1 Z X_{TOT}(1 - \frac{X_p}{X_{TOT}} - \frac{C_1}{X_{TOT}} - \frac{C_2}{X_{TOT}} - \boxed{\frac{C}{X_{TOT}}}) \quad (7.28)$$

$$\frac{dC_2}{dt} = -(k_2 + \alpha_2)C_2 + \alpha_1 Y_{TOT} X_p(1 - \frac{C_2}{Y_{TOT}}) \quad (7.29)$$

$$\frac{dX_p}{dt} = k_1 C_1 + \alpha_2 C_2 - \alpha_1 Y_{TOT} X_p(1 - \frac{C_2}{Y_{TOT}}) + \boxed{k_{off} C - k_{on} X_p(p_{TOT} - C)} \quad (7.30)$$

$$\frac{dC}{dt} = -k_{off} C + k_{on} X_p(p_{TOT} - C), \quad (7.31)$$

in which the expression of gene z is controlled by a promoter with activity $k(t)$. The terms in the large box in equation (7.27) represent the retroactivity r to the input, while the terms in the small box in equation (7.27) and in the boxes of equations (7.28) and (7.30) represent the retroactivity s to the output. We assume that $X_{TOT} \gg p_{TOT}$ so that in equations (7.27) and (7.28) we can neglect the term C/X_{TOT} because $C < p_{TOT}$. Also, phosphorylation and dephosphorylation reactions in equations (7.25) and (7.26) can occur at a much faster rate (on the time scale of a second [**?**]) than protein production and decay processes (on the time scale of minutes [3]). Choosing X_{TOT} and Y_{TOT} sufficiently large, the separation of time-scales between equation (7.27) and equations (7.28–7.31) can be explicitly modeled by letting $\epsilon = \delta/k_{off}$, $k_{on} = k_{off}/k_d$, and by defining the new rate constants $b_1 = \beta_1 X_{TOT}\epsilon/\delta$, $a_1 = \alpha_1 Y_{TOT}\epsilon/\delta$, $b_2 = \beta_2\epsilon/\delta$, $a_2 = \alpha_2\epsilon/\delta$, $c_i = \epsilon k_i/\delta$. Letting $z = Z + C_1$ (the total amount of kinase) be the slow variable, we obtain the system in the standard singular perturbation form

$$\begin{aligned}
\frac{dz}{dt} &= k(t) - \delta(z - C_1) \\
\epsilon\frac{dC_1}{dt} &= -\delta(b_2 + c_1)C_1 + \delta b_1(z - C_1)(1 - \frac{X_p}{X_{TOT}} - \frac{C_1}{X_{TOT}} - \frac{C_2}{X_{TOT}}) \\
\epsilon\frac{dC_2}{dt} &= -\delta(c_2 + a_2)C_2 + \delta a_1 X_p(1 - \frac{C_2}{Y_{TOT}}) \\
\epsilon\frac{dX_p}{dt} &= \delta c_1 C_1 + \delta a_2 C_2 - \delta a_1 X_p(1 - \frac{C_2}{Y_{TOT}}) + \boxed{\delta C - \delta/k_d(p_{TOT} - C)X_p} \\
\epsilon\frac{dC}{dt} &= -\delta C + \delta/k_d(p_{TOT} - C)X_p, \quad (7.32)
\end{aligned}$$

in which the boxed terms represent the retroactivity to the output s. We then compute the dynamics on the slow manifold by letting $\epsilon = 0$. When we set $\epsilon = 0$, the

terms due to the retroactivity s vanish. This means that if the internal dynamics of the insulation device evolve on a time scale that is much faster than the dynamics of the input signal Z, then (provided we also have $X_{TOT} \gg p_{TOT}$) the retroactivity s to the output has no effect on the dynamics of X_p at the quasi steady state. This is a crucial feature of this design. Letting $\gamma = (\beta_2 + k_1)/\beta_1$ and $\bar{\gamma} = (\alpha_2 + k_2)/\alpha_1$, setting $\epsilon = 0$ in the third and fourth equations of (7.32) the following relationships can be obtained:

$$C_1 = F_1(X_p) = \frac{\frac{X_p Y_{TOT} k_2}{\bar{\gamma} k_1}}{1 + X_p/\bar{\gamma}}, \quad C_2 = F_2(X_p) = \frac{\frac{X_p Y_{TOT}}{\bar{\gamma}}}{1 + X_p/\bar{\gamma}}. \tag{7.33}$$

Using expressions (7.33) in the second of equations (7.32) with $\epsilon = 0$ leads to

$$F_1(X_p)(b_2 + c_1 + \frac{b_1 Z}{X_{TOT}}) = b_1 Z(1 - \frac{X_p}{X_{TOT}} - \frac{F_2(X_p)}{X_{TOT}}). \tag{7.34}$$

Assuming for simplicity that $X_p \ll \bar{\gamma}$, we obtain that $F_1(X_p) \approx \frac{X_p Y_{TOT} k_2}{\bar{\gamma} k_1}$ and that $F_2(X_p) \approx \frac{X_p}{\bar{\gamma}} Y_{TOT}$. As a consequence of these simplifications, equation (7.34) leads to

$$X_p = \frac{b_1 Z}{\frac{b_1 Z}{X_{TOT}}(1 + Y_{TOT}/\bar{\gamma} + (Y_{TOT} k_2)/(\bar{\gamma} k_1)) + \frac{Y_{TOT} k_2}{\bar{\gamma} k_1}(b_2 + c_1)} := m(Z).$$

In order not to have distortion from Z to X_p, we require that

$$Z \ll \frac{Y_{TOT} \frac{k_2}{k_1} \frac{\gamma}{\bar{\gamma}}}{1 + \frac{Y_{TOT}}{\bar{\gamma}} + \frac{Y_{TOT}}{\bar{\gamma}} \frac{k_2}{k_1}}, \tag{7.35}$$

so that $m(Z) \approx Z \frac{X_{TOT} \bar{\gamma} k_1}{Y_{TOT} \gamma k_2}$ and therefore we have a linear relationship between X_p and Z with gain from Z to X_p given by $\frac{X_{TOT} \bar{\gamma} k_1}{Y_{TOT} \gamma k_2}$. In order not to have attenuation from Z to X_p we require that the gain is greater than or equal to one, that is,

$$\text{input/output gain} \approx \frac{X_{TOT} \bar{\gamma} k_1}{Y_{TOT} \gamma k_2} \geq 1. \tag{7.36}$$

Requirements (7.35), (7.36), and $X_p \ll \bar{\gamma}$ are enough to guarantee that we do not have nonlinear distortion between Z and X_p and that X_p is not attenuated with respect to Z. In order to guarantee that the retroactivity r to the input is sufficiently small, we need to quantify the retroactivity effect on the Z dynamics due to the binding of Z with X. To achieve this, we proceed as in Section **??** by computing the Z dynamics on the slow manifold, which gives a good approximation of the dynamics of Z if $\epsilon \approx 0$. Such a dynamics is given by

$$\frac{dZ}{dt} = (k(t) - \delta Z)\left(1 - \frac{dF_1}{dX_p}\frac{dX_p}{dz}\right),$$

in which $\frac{dF_1}{dX_p}\frac{dX_p}{dz}$ measures the effect of the retroactivity r to the input on the Z dynamics. Direct computation of $\frac{dF_1}{dX_p}$ and of $\frac{dX_p}{dz}$ along with $X_p \ll \overline{\gamma}$ and with (7.35) leads to $\frac{dF_1}{dX_p}\frac{dX_p}{dz} \approx X_{TOT}/\gamma$, so that in order to have small retroactivity to the input, we require that

$$\frac{X_{TOT}}{\gamma} \ll 1. \tag{7.37}$$

Concluding, for having attenuation of the effect of the retroactivity to the output s, we require that the time scale of the phosphorylation/dephosphorylation reactions is much faster than the production and decay processes of Z (the input to the insulation device) and that $X_{TOT} \gg p_{TOT}$, that is, the total amount of protein X is in abundance compared to the downstream binding sites p. To obtain also a small effect of the retroactivity to the input, we require that $\gamma \gg X_{TOT}$ as established by relation (7.37). This is satisfied if, for example, kinase Z has low affinity to binding with X. To keep the input/output gain between Z and X_p close to one (from equation (7.36)), one can choose $X_{TOT} = Y_{TOT}$, and equal coefficients for the phosphorylation and dephosphorylation reactions, that is, $\gamma = \overline{\gamma}$ and $k_1 = k_2$.

Simulation results. System in equations (7.27–7.31) was simulated with and without the downstream binding sites p, that is, with and without, respectively, the terms in the small box of equation (7.27) and in the boxes in equations (7.30) and (7.28). This is performed to highlight the effect of the retroactivity to the output s on the dynamics of X_p. The simulations validate our theoretical study that indicates that when $X_{TOT} \gg p_{TOT}$ and the time scales of phosphorylation/dephosphorylation are much faster than the time scale of decay and production of the protein Z, the retroactivity to the output s is very well attenuated (Figure 7.17, plot A). Similarly, the time behavior of Z was simulated with and without the terms in the large box in equation (7.27), that is, with and without X to which Z binds, to verify whether the insulation component exhibits retroactivity to the input r. In particular, the accordance of the behaviors of $Z(t)$ with and without its downstream binding sites on X (Figure 7.17, plot B), indicates that there is no substantial retroactivity to the input r generated by the insulation device. This is obtained because $X_{TOT} \ll \gamma$ as indicated in equation (7.37), in which $1/\gamma$ can be interpreted as the affinity of the binding of X to Z. Our simulation study also indicates that a faster time scale of the phosphorylation/dephosphorylation reactions is necessary, even for high values of X_{TOT} and Y_{TOT}, to maintain perfect attenuation of the retroactivity to the output s and small retroactivity to the output r. In fact, slowing down the time scale of phosphorylation and dephosphorylation, the system looses its insulation property (Figure 7.18). In particular, the attenuation of the effect of the retroactivity to the output s is lost because there is not enough separation of time scales between the Z dynamics and the internal device dynamics. The device also displays a non negligible amount of retroactivity to the input because the condition $\gamma \ll X_{TOT}$ is not satisfied anymore.

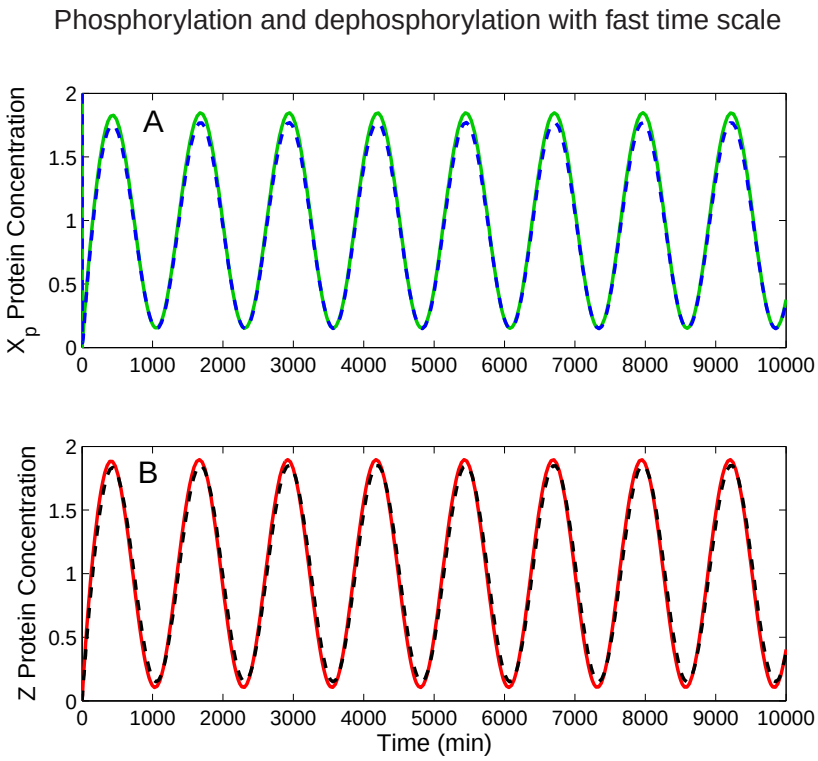


Figure 7.17: Simulation results for system in equations (7.27–7.31). In all plots, p_{TOT} = 100, $k_{off} = k_{on} = 10$, $\delta = 0.01$, $k(t) = 0.01(1 + sin(\omega t))$, and $\omega = 0.005$. In subplots A and B, $k_1 = k_2 = 50$, $\alpha_1 = \beta_1 = 0.01$, $\beta_2 = \alpha_2 = 10$, and $Y_{TOT} = X_{TOT} = 1500$. In subplot A, the signal $X_p(t)$ without the downstream binding sites p is in green (solid line), while the same signal with the downstream binding sites p is in blue (dashed line). The small error shows that the effect of the retroactivity to the output s is attenuated very well. In subplot B, the signal $Z(t)$ without X to which Z binds is in red (solid), while the same signal $Z(t)$ with X present in the system ($X_{TOT} = 1500$) is in black (dashed line). The small error confirms a small retroactivity to the input. The values of the complexes concentrations C_1 and C_2 oscillate about 0.4, so they are comparable to the values of X_p.

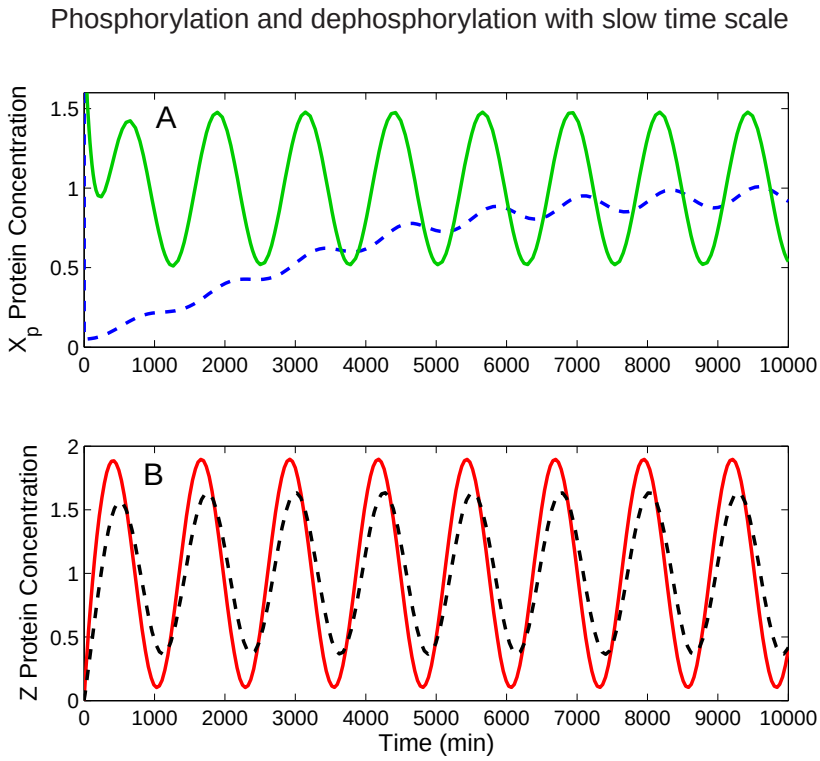


Figure 7.18: In all plots, $p_{TOT} = 100$ and $k_{off} = k_{on} = 10$, $\delta = 0.01$, $k(t) = 0.01(1 + sin(\omega t))$, and $\omega = 0.005$. Phosphorylation and dephosphorylation rates are slower than the ones in Figure 7.17, that is, $k_1 = k_2 = 0.01$, while the other parameters are left the same, that is, $\alpha_2 = \beta_2 = 10$, $\alpha_1 = \beta_1 = 0.01$, and $Y_{TOT} = X_{TOT} = 1500$. In subplot A, the signal $X_p(t)$ without the downstream binding sites p is in green (solid line), while the same signal with the downstream binding sites p is in blue (dashed line). The effect of the retroactivity to the output s is dramatic. In subplot B, the signal $Z(t)$ without X in the system is in red (solid line), while the same signal $Z(t)$ with X in the system is in black (dashed line). The device thus also displays a large retroactivity to the input r.

Chapter 8
Design Tradeoffs

Chapter 9
Design Examples

Bibliography

[1] K. J. Åström and R. M. Murray. *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton University Press, 2008. Available at http://www.cds.caltech.edu/~murray/amwiki.

[2] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson. *The Molecular Biology of the Cell*. Garland Science, fifth edition edition, 2008.

[3] U. Alon. *An introduction to systems biology. Design principles of biological circuits*. Chapman-Hall, 2007.

[4] W. Arber and S. Linn. DNA modification and restriction. *Annual Review of Biochemistry*, 38:467–500, 1969.

[5] D. P. Atherton. *Nonlinear Control Engineering*. Van Nostrand, New York, 1975.

[6] M. R. Atkinson, M. A. Savageau, J. T. Meyers, and A. J. Ninfa. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell*, pages 597–607, 2003.

[7] D. Baker, G. Church, J. Collins, D. Endy, J. Jacobson, J. Keasling, P. Modrich, C. Smolke, and R. Weiss. ENGINEERING LIFE: Building a FAB for biology. *Scientific American*, June 2006.

[8] N Barkai and S Leibler. Robustness in simple biochemical networks. *Nature*, 387(6636):913–7, 1997.

[9] A. Becskei and L. Serrano. Engineering stability in gene networks by autoregulation. *Nature*, 405:590–593, 2000.

[10] F. D. Bushman and M. Ptashne. Activation of transcription by the bacteriophage 434 repressor. *Proceedings of the National Academy of Sciences*, pages 9353–9357, 1986.

[11] A. J. Courey. *Mechanisms in Transcriptional Regulation*. Wiley-Blackwell, 2008.

[12] H. de Jong. Modeling and simulation of genetic regulatory systems: A literature review. *Journal of Computational Biology*, 9:67–103, 2002.

[13] D. Del Vecchio, A. J. Ninfa, and E. D. Sontag. Modular cell biology: Retroactivity and insulation. *Nature/EMBO Molecular Systems Biology*, 4:161, 2008.

[14] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, 2000.

[15] D. Endy. Foundations for engineering biology. *Nature*, 438:449–452, 2005.

[16] T.S. Gardner, C.R. Cantor, and J.J. Collins. Construction of the genetic toggle switch in *Escherichia Coli*. *Nature*, page 339342, 2000.

[17] D. T. Gillespie. *Markov Processes: An Introduction For Physical Scientists*. Academic Press, 1976.

[18] D. T. Gillespie. A rigorous derivation of the chemical master equation. *Physica A*, (188):404–425, 1992.

[19] D. Graham and D. McRuer. *Analysis of Nonlinear Control Systems*. Wiley, New York, 1961.

[20] J. Greenblatt, J. R. Nodwell, and S. W. Mason. Transcriptional antitermination. *Nature*, 364(6436):401–406, 1993.

[21] L.H. Hartwell, J.J. Hopfield, S. Leibler, and A.W. Murray. From molecular to modular cell biology. *Nature*, 402:47–52, 1999.

[22] R. Heinrich, B. G. Neel, and T. A. Rapoport. Mathematical models of protein kinase signal transduction. *Molecular Cell*, 9:957–970, 2002.

[23] C. F. Huang and J. E. Ferrell. Ultrasensitivity in the mitogen-activated proteinkinase cascade. *Proc. Natl. Acad. Sci.*, 93(19):10078–10083, 1996.

[24] B. Ingalls. A frequency domain approach to sensitivity analysis of biochemical networks. *Journal of Physical Chemistry B-Condensed Phase*, 108(3):143–152, 2004.

[25] F. Jacob and J. Monod. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.*, 3:318–56, 1961.

[26] N. G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, 1992.

[27] P. Kokotovic, H. K. Khalil, and J. O'Reilly. *Singular Perturbation Methods in Control*. SIAM, 1999.

[28] H. Madhani. *From a to alpha: Yeast as a Model for Cellular Differentiation*. CSHL Press, 2007.

[29] J. Mallet-Paret and H.L. Smith. The Poincaré-Bendixson theorem for monotone cyclic feedback systems. *J. of Dynamics and Differential Equations.*, 2:367–421, 1990.

[30] C. J. Morton-Firth, T. S. Shimizu, and D. Bray. A free-energy-based stochastic simulation of the tar receptor complex. *Journal of Molecular Biology*, 286(4):1059–74, 1999.

[31] J. D. Murray. *Mathematical Biology*, Vols. I and II. Springer-Verlag, New York, 3rd edition, 2004.

[32] R. M. Murray. *Optimization-Based Control*. http://www.cds.caltech.edu/~murray/amwiki/OBC, Retrieved 20 December 2009.

[33] National Center for Biotechnology Information. A science primer. Retrieved 20 December 2009, 2004. http://www.ncbi.nlm.nih.gov/About/primer/genetics.html.

[34] National Human Genome Research Institute. Talking glossary of genetic terms. Retrieved 20 December 2009. http://www.genome.gov/glossary.

[35] R. Phillips, J. Kondev, and J. Theriot. *Physical Biology of the Cell*. Garland Science, 2008.

[36] M. Ptashne. *A genetic switch*. Blackwell Science, Inc., 1992.

[37] C. V. Rao, J. R. Kirby, and A. P. Arkin. Design and diversity in bacterial chemotaxis: A comparative study in escherichia coli and bacillus subtilis. *PLoS Biology*, 2(2):239–252, 2004.

[38] G. De Rubertis and S. W. Davies. A genetic circuit amplifier: Design and simulation. *IEEE Trans. on Nanobioscience*, 2(4):239–246, 2003.

[39] J. Saez-Rodriguez, A. Kremling, H. Conzelmann, K. Bettenbrock, and E. D. Gilles. Modular analysis of signal transduction networks. *IEEE Control Systems Magazine*, pages 35–52, 2004.

[40] J. Saez-Rodriguez, A. Kremling, and E.D. Gilles. Dissecting the puzzle of life: modularization of signal transduction networks. *Computers and Chemical Engineering*, 29:619–629, 2005.

[41] H. M. Sauro and B. Ingalls. MAPK cascades as feedback amplifiers. Technical report, http://arxiv.org/abs/0710.5195, Oct 2007.

[42] H. M. Sauro and B. N. Kholodenko. Quantitative analysis of signaling networks. *Progress in Biophysics & Molecular Biology*, 86:5–43, 2004.

[43] H.M. Sauro. The computational versatility of proteomic signaling networks. *Current Proteomics*, 1(1):67–81, 2004.

[44] M. A. Savageau. Biochemical systems analysis. i. some mathematical properties of the rate law for the component enzymatic reactions. *J. Theoretical Biology*, 25:365–369, 1969.

[45] D. L. Schilling and C. Belove. *Electronic Circuits: Discrete and Integrated*. McGraw Hill, 1968.

[46] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.*, 31(1):64–68, 2002.

[47] L. Villa-Komaroff, A. Efstratiadis, S. Broome, P. Lomedico, R. Tizard, S. P. Naber, W. L. Chick, and W. Gilbert. A bacterial clone synthesizing proinsulin. *Proc. Natl. Acad. Sci. U.S.A.*, 75(8):372731, 1978.

[48] T.-M. Yi, Y. Huang, M. I. Simon, and J. Doyle. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proceedings of the National Academy of Sciences*, 97(9):4649–53, 2000.

Index