Chapter 5

The Mathematics of Molecular Biology

5.1 A brief overview of molecular biology

The field of molecular biology aims to understand biological function in terms of the underlying molecular mechanisms that drive it. All cells are surrounded by a cell membrane — a double layer of lipid (fat) molecules that isolates the interior of the cell from the external environment. The fluid inside of the cell (the cytoplasm) contains a concentrated mixture of many different molecules, ranging in size from small inorganic molecules (such as water and dissolved salts and gases) to large biopolymers (proteins and nucleic acids), with organic molecules of a range of sizes laying in between these extremes; the external environment also typically presents at complex mixture of chemical species. It is the physical and chemical interactions between these molecules that underly the processes of life. The study of these systems ranges from developing an understanding of the fundamental chemical mechanisms by which these molecules react and interact (biochemistry) to more abstract models of how all the components work together to achieve biological function (molecular and cellular biology). While mathematical models for biochemical systems have played an important role in this discipline for decades, models for the complex interplay of molecules within the context of the cell have only recently become prominant. This emerging field of *systems biology* promises to play an increasingly important role in advancing our understanding of molecular biology in the years to come.

5.1.1 The central dogma

The fundamental concepts underlying our current understanding of molecular biology can be encapsulated in what is known as **The Central Dogma of Molecular Biology** (Figure 5.1):

- DNA (a polymer of deoxyribonucleic acids) is the primary information storage molecule in all organisms, and can be *replicated* to produce identical copies; this processes underlies reproduction; the sequence of all DNA in an organism is it's *genome*.

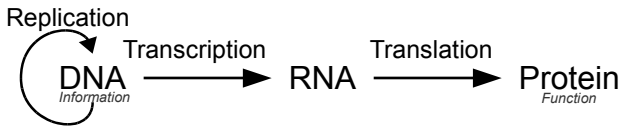

Figure 5.1: The Central Dogma of Molecular Biology

- The DNA of an organism contains a code for the sequence of many different proteins; each of these encoded sequences is a *gene.*
- Each gene (or small groups of related genes) can be converted into RNA (a polymer of ribonucleic acids) in a process called *transcription*. Each of these messenger RNA molecules (mRNA) is nearly identical to one strand of the DNA from which it was transcribed (over the region of transcription), and thus contains the code for the sequence of that particular protein.
- Proteins (polymers of amino acids) are made through a process of *translation*, where the encoded sequence in the mRNA is used as a template to create a new protein molecule of the appropriate sequence.
- It is the proteins that carry out most functions in an organism; among many roles, they act as enzymes to catalyze reactions, provide structural support to cells and tissues, and form receptors that sense external stimuli. The processes of replication and transcription are both mediated by proteins that both carry out the chemical process of making new DNA/RNA and that regulate when this is done.

It is important to note that the view of biology presented by the central dogma is a simplification that describes most, but not all, aspects of biological systems. For example, RNA can also play important functional roles. For example the processes of translation is driven by the ribosome, which is a large complex of many RNA molecules (rRNA) and proteins; the proteins play a primarily structural role, stabilizing the complex, while the rRNA actually catalyzes the synthesis of new protein. The decoding step of translation is additionally mediated by transfer RNA molecules (tRNA).

5.1.2 Molecular mechanisms of biological function

One of the the least satisfying statements of the central dogma is the phrasing of "proteins then carry out the functions of the cell." As these functions are truly what define an organisms features and behaviors, it many ways, the details of this step are of great interest. The functions of proteins are diverse, and may be classified in many different ways. At a very fundamental level, we may group functions as consisting of:

Catalysis: Proteins that catalyze (speed up) specific chemical reactions are known as *enzymes*. These reactions may involve the trasformation of some small organic molecule

into another (for example in the process of coverting sugar into useable energy), or my involve chemical modifications to biological macromolecules (other proteins, DNA, RNA, etc).

Macromolecular association: Many proteins can physically interact with other proteins, with DNA or with RNA. These interactions can have a direct effect on the function of the target (by reducing or enhancing enzymatic activity, for example) or may serve to physically restrict the target to a particular location.

Small molecule binding: Proteins can also interact with small molecules (organic or inorganic); while enzymes bind such molecules and aid their transfromation to another species, physical interactions that do not change the bound molecule can also occur. These interactions may have a role of modulating enzymatic activity or of modifying the strength of interactions with other macromolecules.

While some proteins may only have one of these general functions, many combine multiple functionalities.

Overall biological responses are caused by systems of many proteins, each following these fundmental mechanisms: cascades of chemical reactions (metabolic pathways) can be used to achieve large-scale chemical transformations; networks of physical and chemical interactions between proteins (signal transduction networks) can be used to pass information from one part of a cell or organism to another; physical interactions between various proteins and DNA (gene-regulatory networks) form the basis by which the expression patterns of particular genes are regulated.

5.2 Models of enzymatic activity

5.2.1 Introduction to biochemical kinetics

First let us discuss models of enzymatic activity, as one of the fundamental components of biochemical networks; how can we describe the dynamics of molecular species undergoing different chemical reactions? In chemical kinetics, we are interested in describing *how much* of various molecules are present in a given system as a function of time. Typically, we use units of concentration, which dictates the number of molecules per a given unit of volume. The standard unit is *molar*, or *moles per liter*, where 1 mole is equal to 6.02×10^{23} molecules. In biological system, it is most common to encounter concentrations is the range of nanomolar (1×10^{-9}M) to millimolar (1×10^{-3}M). Note that we typically denote the concentration of species X with the notation $[X]$, although you may also come across the notation c_X or even simply X.

Just as in our models of populations, biomass, and infectious disease, we are interested primarily in how concentrations of various *molecular species* change over time, and thus similar differential equation-based models are most appropriate. The Law of Mass Action provides a starting point, describing the rates of change for simple (irreversible) chemical reactions. For a reaction involving a single *reactant* molecule, A, which is converted to a

single *product* molecule, P, we have:

$$A \longrightarrow P \qquad \frac{d[P]}{dt} = \frac{-d[A]}{dt} = k_1[A] \tag{5.1}$$

which is referred to as a **first-order reaction**. For a reaction involving two reactants, A and B, (and again a single product):

$$A + B \longrightarrow P \qquad \frac{d[P]}{dt} = \frac{-d[A]}{dt} = \frac{-d[B]}{dt} = k_2[A][B] \tag{5.2}$$

which is referred to as a **second-order reaction**. Essentially, the Law of Mass Action states that the reaction rate will be proportional to the product of the reactant concentrations. If multiple molecules of a product are formed from one molecule of reactant, the basic rate law is the same, but the rates of change of each species must take into account the *stoichiometry* (the number of molecules of each type involved in the reaction). Thus:

$$A \longrightarrow 2P \qquad 2\frac{d[P]}{dt} = \frac{-d[A]}{dt} = k_1[A] \quad \text{or} \quad \frac{d[P]}{dt} = \frac{-1}{2}\frac{d[A]}{dt} = \frac{1}{2}k_1[A] \tag{5.3}$$

If there are two identical reactants involved, we need to consider the stoichiometry as well, but also note that the rate law follows second-order kinetics:

$$2A \longrightarrow P \qquad \frac{d[P]}{dt} = -2\frac{d[A]}{dt} = k_1[A]^2 \tag{5.4}$$

Now, in natural systems, many reactions are reversible — that is, note only may molecule A convert to P, but P can also convert to A. In this case, we can think of a system of two simultaneous reactions, one in each direction; the rate laws discussed above will describe the rate of change (or flux) along a given path. So:

$$A \underset{k_r[P]}{\overset{k_f[A]}{\rightleftharpoons}} P \quad \leftrightarrow \quad A \overset{k_f[A]}{\longrightarrow} P \quad \text{and} \quad P \overset{k_r[P]}{\longrightarrow} A \tag{5.5}$$

As in the models we have discussed previously, the overall rate of change of a given molecular species is the sum of the contributions from each reaction in which it is involved. Thus, for the reaction above:

$$\frac{d[A]}{dt} = -k_f[A] + k_r[P] \quad \text{and} \quad \frac{d[P]}{dt} = +k_f[A] - k_r[P] \tag{5.6}$$

For bimolecular reaction:

$$A + B \underset{k_r[P]}{\overset{k_f[A][B]}{\rightleftharpoons}} P \quad \leftrightarrow \quad A + B \overset{k_f[A][B]}{\longrightarrow} P \quad \text{and} \quad P \overset{k_r[P]}{\longrightarrow} A + B \tag{5.7}$$

and:

$$\frac{d[A]}{dt} = \frac{d[B]}{dt} = -k_f[A][B] + k_r[P] \quad \text{and} \quad \frac{d[P]}{dt} = +k_f[A][B] - k_r[P] \tag{5.8}$$

One of the key assumptions in mass action kinetics is that the reaction occurs in one step, involving a single "collision" of all the species involved. However, enzymatic reactions are *not* simple reactions like these. Instead, we may consider the steps involved the simplest of reactions:

1. The reactant (called a *substrate* in the context of an enzymatic reaction) physically associates with the enzyme. We call this a binding reaction.

2. While bound in the enzyme's active site, the substrate is converted to the product. This typically happens much faster than the reaction would with out the enzyme; the enzyme acts as a catalyst of the reaction.

3. After the reaction, the product "unbinds" or dissociates from the enzyme.

As each of these reactions may be reversible, we can write a general schematic for this mechanism:

$$E + S \underset{k_{-1}[ES]}{\overset{k_{+1}[E][S}{\rightleftharpoons}} ES \underset{k_{-1}[EP]}{\overset{k_{+2}[ES]}{\rightleftharpoons}} EP \underset{k_{-1}[E][P}{\overset{k_{+3}[EP]}{\rightleftharpoons}} E + P \tag{5.9}$$

which gives rise to a system of differential equations:

$$\begin{array}{rcccccc} \frac{d[E]}{dt} & = & -k_{+1}[E][S] & +k_{-1}[ES] & & & +k_{+3}[EP] & -k_{-3}[E][P] \\ \frac{d[S]}{dt} & = & -k_{+1}[E][S] & +k_{-1}[ES] & & & & \\ \frac{d[ES]}{dt} & = & +k_{+1}[E][S] & -k_{-1}[ES] & -k_{+2}[ES] & +k_{-2}[EP] & & \\ \frac{d[EP]}{dt} & = & & & +k_{+2}[ES] & -k_{-2}[EP] & -k_{+3}[EP] & +k_{-3}[E][P] \\ \frac{d[P]}{dt} & = & & & & & +k_{+3}[EP] & -k_{-3}[E][P] \end{array} \tag{5.10}$$

where corresponding terms have been aligned for clarity. Now, we see that this is a system of five equations in five variables, with six individual rate constants. This is certainly not overly complicated for numerical simulation, but this is just a single enzymatic reaction!

5.2.2 The Michaelis–Menten equation

However, knowing something about how many enzymes work allows us to simply these expressions significantly. First, we note that in many cases, the forward enzymatic reaction is highly preferred (*i.e.* $k_{+2} >> k_{-2}$) and the dissociation of product is much preferred over reassociation (*i.e.* $k_{+3} >> k_{-3}$). This simplies the schematic to:

$$E + S \underset{k_{-1}[ES]}{\overset{k_{+1}[E][S]}{\rightleftharpoons}} ES \overset{k_{+2}[ES]}{\longrightarrow} EP \overset{k_{+3}[EP]}{\longrightarrow} E + P \tag{5.11}$$

Additionally, it is often the case that dissociation and association reactions happen much faster than chemical (catalytic) reactions, and thus $k_{+3} >> k_{+2}$. This allows us to assume that every time the enzyme-substrate complex is converted to enzyme-product, it immediately dissociates irreversibly. This allows for one more simplification:

$$E + S \underset{k_{-1}[ES]}{\overset{k_{+1}[E][S]}{\rightleftharpoons}} ES \overset{k_{+2}[ES]}{\longrightarrow} E + P \tag{5.12}$$

which gives the system of differential equations:

$$\begin{array}{rcccc} \frac{d[E]}{dt} & = & -k_{+1}[E][S] & +k_{-1}[ES] & +k_{+2}[ES] \\ \frac{d[S]}{dt} & = & -k_{+1}[E][S] & +k_{-1}[ES] & \\ \frac{d[ES]}{dt} & = & +k_{+1}[E][S] & -k_{-1}[ES] & -k_{+2}[ES] \\ \frac{d[P]}{dt} & = & & & +k_{+2}[ES] \end{array} \tag{5.13}$$

Note that, $\frac{d[E]}{dt} = -\frac{d[ES]}{dt}$, which can be expressed alternatively as:

$$\frac{d[E]}{dt} + \frac{d[ES]}{dt} = 0 \qquad \leftrightarrow \qquad [E] + [ES] = constant = [E]_{\text{tot}} \tag{5.14}$$

This is simply an expression of the **law of conservation of mass** — the total amount of enzyme can not change, as there are no inputs or outputs in the system. Similarly, $\frac{d[S]}{dt} + \frac{d[P]}{dt} = -\frac{d[ES]}{dt}$, which gives:

$$\frac{d[S]}{dt} + \frac{d[ES]}{dt} + \frac{d[P]}{dt} = 0 \qquad \leftrightarrow \qquad [S] + [ES] + [P] = constant \tag{5.15}$$

Again, this is an expression of the conservation of mass of substrate and product; if there is no product present at the beginning of the reaction, the constant will be the initial substrate concentration. We have reduced the problem to four equations in four variables (although product does not play an explicit role in any of the expressions), and only three rate constants.

Now, consider what happens if a small amount of enzyme and a larger amount of substrate are combined. Initially, no enzyme-substrate complex is present ($[ES] = 0$), and thus, the only reaction with a non-zero rate is, $E + S \rightarrow ES$. As a result, the concentration of free enzyme ($[E]$) will begin to decrease, and the total amount of $[ES]$ will increase. The total amount of free substrate ($[S]$) will also decrease, but:

$$[S] >> [E]_{\text{tot}} \text{ and } [ES] < [E]_{\text{tot}} \therefore [S] = [S]_{\text{tot}} - [ES] \approx [S]_{tot} \tag{5.16}$$

That is, if there is a large excess of substrate, initially there will be no noticeable decrease in substrate concentration.

As $[ES]$ increases, both additional reactions ($ES \rightarrow E + S$ and $ES \rightarrow E + P$) will increase in rate, and as $[E]$ decreases, the rate of the initial binding reaction will similarly decrease. Eventually, a point will be reached when:

$$k_{+1}[E][S] = k_{-1}[ES] + k_{+2}[ES] \Rightarrow \frac{d[ES]}{dt} = 0 \tag{5.17}$$

This point is called **quasi-steady state**, since some of the components (levels of $[E]$ and $[ES]$) are not varying with time, but product is actively being formed. If the initial concentration of $[S]$ is large, or if there is an external supply of $[S]$ which replenishes used substrate, this quasi-steady state can persist for a significant length of time. Now, since the rate of product production is simply defined by $\frac{d[P]}{dt} = k_{+2}[ES]$, when $[ES]$ is constant, so will the rate of product production; a constant rate of product production means that product concentrations will increase linearly with time.

How can we characterize this quasi-steady state in more detail? Combining equations 5.14 and 5.17 gives:

$$k_{+1}([E]_{\text{tot}} - [ES])[S] = k_{-1}[ES] + k_{+2}[ES] \tag{5.18}$$

which rearranges to: [1]

$$[ES] = \frac{k_{+1}[E]_{\text{tot}}[S]}{k_{+1}[S] + k_{-1} + k_{+2}} = \frac{[E]_{\text{tot}}[S]}{\frac{k_{-1}+k_{+2}}{k_{+1}} + [S]} \tag{5.19}$$

[1] $k_{+1}[E]_{\text{tot}}[S] = k_{+1}[ES][S] + k_{-1}[ES] + k_{+2}[ES] = (k_{+1}[S] + k_{-1} + k_{+2})[ES] \rightarrow [ES] = \frac{k_{+1}[E]_{\text{tot}}[S]}{k_{+1}[S]+k_{-1}+k_{+2}}$

As $\frac{d[P]}{dt} = k_{+2}[ES]$, we then can write:

$$\frac{d[P]}{dt} = \frac{k_{+2}[E]_{\text{tot}}[S]}{\frac{k_{-1}+k_{+2}}{k_{+1}} + [S]} \tag{5.20}$$

which is often written as:

$$\frac{d[P]}{dt} = \frac{k_{\text{cat}}[E]_{\text{tot}}[S]}{K_{\text{m}} + [S]} \tag{5.21}$$

using the substitutions of $k_{\text{cat}} = k_{+2}$ and $K_{\text{m}} = \frac{k_{-1}+k_{+2}}{k_{+1}}$; the is the **Michaelis–Menten equation**. The Michaelis–Menten equation directly gives $\frac{d[P]}{dt}$ in terms of $[S]$ and $[E]_{\text{tot}}$, and thus simplifies a system of three differential equations to one.

Now, we may briefly examine the behavior of the Michaelis–Menten equation. First, we note that the rate of product creation is linear in the total enzyme concentration; a doubling of the total amount of enzyme present will double the rate at which product is made. In terms of substrate concentration, we have a saturating behavior, much like that seen in population dynamics with logistic growth and Holling's disc equation for predator response. At low $[S]$, when $K_{\text{m}} >> [S]$, we have:

$$\frac{d[P]}{dt} \approx \frac{k_{\text{cat}}[E]_{\text{tot}}[S]}{K_{\text{m}}} \tag{5.22}$$

which is linear in substrate concentration, with a slope of $\frac{k_{\text{cat}}[E]_{\text{tot}}}{K_{\text{m}}}$. However, at large $[S]$, $K_{\text{m}} << [S]$, and:

$$\frac{d[P]}{dt} \approx \frac{k_{\text{cat}}[E]_{\text{tot}}[S]}{[S]} = k_{\text{cat}}[E]_{\text{tot}} \tag{5.23}$$

which is a constant. This maximal rate of product creation, even when large excesses of substrate are present, is often referred to as V_{max}, and thus sometime the Michaelis–Menten equation is written as $\frac{d[P]}{dt} = \frac{V_{\text{max}}[S]}{K_{\text{m}}+[S]}$. However it is important to note that, in this representation, V_{max}, is implictly dependent on the total enzyme concentration, and thus is not a true constant.

Many enzymes tend to operate in the linear regime (in a biological context), and thus the ratio $\frac{k_{\text{cat}}}{K_{\text{m}}}$ is the key parameter defining the rate of product creation. As a result, this ratio is often referred to as the *catalytic efficiency* of the enzyme. Also, note that:

$$\frac{k_{\text{cat}}}{K_{\text{m}}} = \frac{k_{+1}k_{+2}}{k_{-1} + k_{+2}} \tag{5.24}$$

If the enzyme has a very high rate of chemical catalysis, then k_{+2} will be very large, and thus will dominate the denominator. In this regime:

$$\frac{k_{\text{cat}}}{K_{\text{m}}} \approx \frac{k_{+1}k_{+2}}{k_{+2}} = k_{+1} \tag{5.25}$$

and we see that the rate of catalysis is limited by the rate of binding. Since the rate of association can not exceed the rate at which molecules diffuse, this is often refered to at *diffusion-limited* catalysis. One result of this, is that there is little evolutionary pressure

to evolve an enyzme whose k_{cat} exceeds a certain value; once the diffusion limit is reached, further increases in k_{cat} will not affect the catalytic efficiency. Observationally, these so-call "perfect" enzymes have $\frac{k_{\text{cat}}}{K_{\text{m}}}$ values of roughly 10^8–10^9 $\text{M}^{-1}\text{s}^{-1}$.

If, on the otherhand, the enzyme only has a week rate of chemical catalysis, then $k_{+2} << k_{-1}$ and:

$$\frac{k_{\text{cat}}}{K_{\text{m}}} \approx \frac{k_{+1}k_{+2}}{k_{-1}} \tag{5.26}$$

The ratio $\frac{k_{+1}}{k_{-1}}$ is known as the equilibrium association constant, K_a, and its repicrocal ($\frac{k_{-1}}{k_{+1}}$) is the equilibrium dissocation constant, K_d:

$$K_a = \frac{k_{+1}}{k_{-1}} = \frac{[ES]_{eq}}{[E]_{eq}[S]_{eq}} \qquad K_d = \frac{k_{-1}}{k_{+1}} = \frac{[E]_{eq}[S]_{eq}}{[ES]_{eq}} \tag{5.27}$$

where $[X]_{eq}$ is the concentration of species X at equilibrium. Thus, when chemical catalysis is slow, K_{m} can be interpreted as a dissociation constant; it is important to note, however, that this interpretation is only valid when $k_{+2} << k_{-1}$.

How do we measure k_{cat} and K_M for a given enzyme? First, consider what happens if we mix an excess of substrate with a small amount of enzyme, and measure how much product is present over time. A plot of product concentration versus time will begin with a lag phase, corresponding to the pre-quasi-steady state regime, where product is binding to the enzyme, but little catalysis is occuring, followed by a linear increase in product concentration corresponding to quasi-steady state. Finally, the product concentration will level off at a constant value, when all substrate has been converted. The slope of the linear phase is precisely $\frac{d[P]}{dt}$ during quasi-steady state.

Now, imagine that the same experiment is repeated using different starting concentrations of substrate. We may then plot $\frac{d[P]}{dt}$ (at quasi-steady state) as a function of $[S]$, and should get a curve that matches the form of the Michaelis–Menten equation. As discussed earlier, this curve should two regimes: a linear regime at low $[S]$, with slope $\frac{k_{\text{cat}}[E]_{\text{tot}}}{K_{\text{M}}}$, and a saturated regime at high $[S]$ with a constant value of $k_{\text{cat}}[E]_{\text{tot}}$. Thus, an analysis of this plot directly gives us the values of interest.

However, accurately fitting the linear and saturated regimes to separate lines requires that many data points are present in both. As an alternative, consider the reciprocal of the Michaelis–Menten equation:

$$\frac{1}{d[P]/dt} = \frac{K_{\text{m}} + [S]}{k_{\text{cat}}[E]_{\text{tot}}[S]} = \frac{K_{\text{m}}}{k_{\text{cat}}[E]_{\text{tot}}}\frac{1}{[S]} + \frac{1}{k_{\text{cat}}[E]_{\text{tot}}} \tag{5.28}$$

This is a *linear* equation in the reciprocal of $[S]$, and thus we can consider plotting $\frac{1}{d[P]/dt}$ versus $\frac{1}{[S]}$; this is known as a *double-reciprocal* or **Lineweaver–Burk** plot. The result should be a straight line, with slope, $\frac{K_{\text{m}}}{k_{\text{cat}}[E]_{\text{tot}}}$ and intercept $\frac{1}{k_{\text{cat}}[E]_{\text{tot}}}$. Thus:

$$K_{\text{m}} = \frac{m}{b} \qquad k_{\text{cat}} = \frac{1}{b[E]_{\text{tot}}} \tag{5.29}$$

where m is the slope and b is the y-intercept. As this simply involves a single fit of a straight line, the data points from all regimes can be used at once. It should also be noted that a computer-based non-linear least-squares fit directly to the Michaelis–Menten equation can also be performed, and this is often done.

5.3 Beyond Michaelis–Menten

The Michaelis–Menten equation was derived using a very simple model of an enzymatic mechanism, and many enzymes do not follow this. These deviations may include, as a partial list: enzymes with more than one substrate, enzymes with intermediates in the catalytic pathway, and enzymes that are regulated by other molecules.

5.3.1 Multiple substrates

An enzyme such as hexokinase (in the glycolytic pathway) involves two substrates, converting glucose and ATP into glucose-6-phosphate and ADP:

$$\text{hexokinase} + \text{glucose} + \text{ATP} \longrightarrow \text{hexokinase} + \text{G6P} + \text{ADP} \tag{5.30}$$

The net reaction is of the general form:

$$E + S_1 + S_2 \rightleftharpoons E + P_1 + P_2 \tag{5.31}$$

but what is a reasonable mechanism for how this occurs; there are multiple possibilities. One option is an *compulsory order ternary complex* mechanism:

$$E + S_1 + S_2 \rightleftharpoons ES_1 + S_2 \rightleftharpoons ES_1S_2 \rightleftharpoons EP_1P_2 \rightleftharpoons EP_1 + P_2 \rightleftharpoons E + P_1 + P_2 \tag{5.32}$$

in which first S_1 binds to E, followed by S_2, which catalysis directly forming P_1 and P_2, which are then release in order, P_2 then P_1. However, what if the order of binding doesn't matter? This gives rise to a *random order ternary complex* mechanism:

$$\begin{array}{ccccc} & ES_1 + S_2 & & EP_1 + P_2 & \\ \nearrow\!\!\!\nearrow & & \searrow & & \searrow\!\!\!\searrow \\ E + S_1 + S_2 & & ES_1S_2 \rightleftharpoons EP_1P_2 & & E + P_1 + P_2 \\ \searrow\!\!\!\searrow & & \nearrow & & \nearrow\!\!\!\nearrow \\ & ES_2 + S_1 & & EP_2 + P_1 & \end{array} \tag{5.33}$$

We also may have a case where the first substrate binds to the enzyme alone, modifying the enzyme as it reacts to form the first product; the modified enzyme may then bind to the second substrate, returning to its original form as the second product is formed. This is a *ping-pong* or *substituted enzyme* mechanism:

$$\begin{array}{ccccccc} E + S_1 & \rightleftharpoons & ES_1 & \rightleftharpoons & E*P_1 & \rightleftharpoons & E*{+}P_1 \\ +S_2 & & +S_2 & & +S_2 & & +S_2 \\ E*{+}S_2 & \rightleftharpoons & E*S_2 & \rightleftharpoons & EP_2 & \rightleftharpoons & E + P_2 \\ +P_1 & & +P_1 & & +P_1 & & +P_1 \end{array} \tag{5.34}$$

While we will not go through the details, for each of these mechanisms, a quasi-steady state rate law can be derived, giving a single rate of product formation that depends only on $[S_1]$, $[S_2]$ and $[E]_{\text{tot}}$.

5.3.2 Regulated enzymatic reactions

Many enzymatic reactions are additionally controled by other molecules. For example, the glycolytic enzyme phosphofructokinase-1, which catalyzes the reaction of fructose-6-phosphate and ATP to form fructose-1,6-biphosphate and ADP, is *activated* by AMP, as well as by fructose-2,6-biphosphate. High levels of AMP are indicative of low energy levels in the cell, and this regulation "speeds up" energy production in this regime. High levels of fructose-2,6-biphosphate (which is made from the substrate fructose-6-phosphate by an other enzyme, phosphofructokinase-2) indicate that glycolysis is "backed up", and again the regulation acts to speed up the overall process.

Hexokinase is also regulated, but in a negative manner; its product (glucose-6-phophate) inhibits the enzyme. High levels of the product are an indication that downstream enzymes that use the product are saturated, and thus the regulation acts to slow down production.

Inhibition

An inhibitor is a molecule which binds to an enzyme and reduces its ezymatic activity; this may be the product of the reaction itself (product inhibition) or some other molecule. Schematically, we can write a set of reactions:

$$\begin{array}{ccccc} E+S+I & \rightleftharpoons & ES+I & \longrightarrow & E+P+I \\ \Updownarrow & & \Updownarrow & & \Updownarrow \\ EI+S & \rightleftharpoons & ESI & \longrightarrow & EI+P \end{array} \tag{5.35}$$

where the vertical arrows represent inhibitor binding, and both the enzyme alone and enzyme bound to inhibitor are considered to follow a Michaelis–Menten like mechanism. For I to act as an inhibitor, catalysis by the inhibitor-bound form must be weaker than that of the free enzyme; this may be a result of a lower $k_{\rm cat}$, a higher $K_{\rm m}$, or both. In *competitive* inhibition, we have the extreme case, where $K_{\rm m}$ for the inhibitor bound form is infinite; at the molecular level, this means that when the inhibitor is bound, it is impossible to also bind substrate; the EIS complex can never form. This often occurs when the inhibitor and the substrate interact with the protein at the same physical location. In this case, we can rewrite our schematic as:

$$EI + S \underset{k_{-I}[EI]}{\overset{k_{+I}[E][I]}{\rightleftharpoons}} E + S + I \underset{k_{-1}[ES]}{\overset{k_{+1}[E][S]}{\rightleftharpoons}} ES + I \overset{k_{+2}[ES]}{\longrightarrow} E + P + I \tag{5.36}$$

Where we have added in the rate laws for the elementary component reactions. At quasi-steady-state, we require:

$$\frac{d[ES]}{dt} = 0 \qquad \text{and} \qquad \frac{d[EI]}{dt} = 0 \tag{5.37}$$

The second equation leads to:

$$k_{-I}[EI] = k_{+I}[E][I] \qquad \rightarrow \qquad \frac{[E][I]}{[EI]} = \frac{k_{-I}}{k_{+I}} = K_I \tag{5.38}$$

where K_I is termed the *inhibition constant* and may be interpretted as the equilibrium dissociation constant of the inhibitor. Since the reactions to the right of the inhibitor binding step are identical to the original Michaelis–Menten mechanism, we also have:

$$k_{+1}[E][S] = k_{-1}[ES] + k_{+2}[ES] \qquad \rightarrow \qquad \frac{[E][S]}{[ES]} = \frac{k_{-1} + k_{+2}}{k_{+1}} = K_\mathrm{m} \tag{5.39}$$

and conservation of enzyme mass requires:

$$[E] + [ES] + [EI] = [E]_\mathrm{tot} \tag{5.40}$$

Combining these gives the quasi-steady-state concentration of ES:[2]

$$[ES] = \frac{[E]_\mathrm{tot}}{\left(1 + \frac{K_\mathrm{m}}{[S]} + \frac{[I]}{K_I}\frac{K_\mathrm{m}}{[S]}\right)} = \frac{[E]_\mathrm{tot}[S]}{[S] + K_\mathrm{m} + \frac{[I]}{K_I}K_\mathrm{m}} = \frac{[E]_\mathrm{tot}[S]}{K_\mathrm{m}\left(1 + \frac{[I]}{K_I}\right) + [S]} \tag{5.41}$$

Since $\frac{d[P]}{dt} = k_{+2}[ES]$ and taking $k_\mathrm{cat} = k_{+2}$:

$$\frac{d[P]}{dt} = k_\mathrm{cat}[ES] = \frac{k_\mathrm{cat}[E]_\mathrm{tot}[S]}{K_\mathrm{m}\left(1 + \frac{[I]}{K_I}\right) + [S]} \tag{5.42}$$

This is very nearly the same as the Michaelis–Menten equation, but with K_m scaled by a factor of $1 + \frac{[I]}{K_I}$. Thus, the Michaelis constant (which is inversely proportional to the slope of the linear phase of the rate curve) will increase linearly with inhibitor concentration.

Activation

Now consider a molecule that binds to an enzyme and *enhances* the activity; we can right the same general schematic as for inhibition:

$$\begin{array}{ccccc} E + S + X & \rightleftharpoons & ES + X & \longrightarrow & E + P + X \\ \Updownarrow & & \Updownarrow & & \Updownarrow \\ EX + S & \rightleftharpoons & ESX & \longrightarrow & EX + P \end{array} \tag{5.43}$$

but with the activator-bound rate larger (by an increase in k_cat, a decrease in K_m, or both). Again we may consider an extreme case, where the enzyme is unable to bind to substrate in the absence of the inhibitor (K_m for the *unbound* enzyme is infinite). This gives rise to the schematic:

$$E + S + X \underset{k_{-X}[EX]}{\overset{k_{+X}[E][X]}{\rightleftharpoons}} EX + S \underset{k_{-1}[EXS]}{\overset{k_{+1}[EX][S]}{\rightleftharpoons}} EXS + I \overset{k_{+2}[EXS]}{\longrightarrow} EX + P + I \tag{5.44}$$

This is very similar to the schematic for inhibition, and thus to achieve quasi-steady-state, we have similar results:

$$\frac{d[EXS]}{dt} = 0 \qquad \text{and} \qquad \frac{d[EX]}{dt} = 0 \tag{5.45}$$

[2] $[E] + [ES] + [EI] = [E]_\mathrm{tot} \rightarrow [E] + [ES] + \frac{[E][I]}{K_I} = [E]_\mathrm{tot} \rightarrow [ES] + \left(1 + \frac{[I]}{K_I}\right)[E] = [E]_\mathrm{tot} \rightarrow [ES] + \left(1 + \frac{[I]}{K_I}\right)\frac{K_\mathrm{m}[ES]}{[S]} = [E]_\mathrm{tot} \rightarrow \left(1 + \frac{K_\mathrm{m}}{[S]} + \frac{[I]}{K_I}\frac{K_\mathrm{m}}{[S]}\right)[ES] = [E]_\mathrm{tot} \rightarrow [ES] = \frac{[E]_\mathrm{tot}}{\left(1 + \frac{K_\mathrm{m}}{[S]} + \frac{[I]}{K_I}\frac{K_\mathrm{m}}{[S]}\right)}$

giving:

$$k_{-X}[EX] = k_{+X}[E][X] \quad \rightarrow \quad \frac{[E][X]}{[EX]} = \frac{k_{-X}}{k_{+X}} = K_X \tag{5.46}$$

and:

$$k_{+1}[EX][S] = k_{-1}[EXS] + k_{+2}[EXS] \quad \rightarrow \quad \frac{[EX][S]}{[EXS]} = \frac{k_{-1} + k_{+2}}{k_{+1}} = K_{\mathrm{m}} \tag{5.47}$$

with conservation of enzyme mass requiring:

$$[E] + [EX] + [EXS] = [E]_{\mathrm{tot}} \tag{5.48}$$

Using the same process as for the inhibited enzyme, these combine to give (at quasi-steady-state):

$$[ES] = \frac{[E]_{\mathrm{tot}}[S]}{K_{\mathrm{m}}\left(1 + \frac{K_X}{[X]}\right) + [S]} \tag{5.49}$$

and:

$$\frac{d[P]}{dt} = k_{\mathrm{cat}}[EXS] = \frac{k_{\mathrm{cat}}[E]_{\mathrm{tot}}[S]}{K_{\mathrm{m}}\left(1 + \frac{K_X}{[X]}\right) + [S]} \tag{5.50}$$

Note that the scaling factor of K_{m} here is $1 + \frac{K_X}{[X]}$, which will approach 1 as $[X]$ becomes larges and will approach infinity as $[X]$ approaches zero. Again, the effect of the activator is on the slope of the linear regime of the rate curve.

Complex regulatory mechanisms

The extreme models of inhibition and activation that were discussed above both impacted only the K_{m} of the reaction, with V_{max} (which depends solely on k_{cat}) unaffected. In a more complicated system, the binding of a molecule to an enzyme may modulate both k_{cat} and K_{m}, perhaps in different ways. We may generally classify the possible responses into four types, using k'_{cat} and K'_{m} to denote the constants with the molecule bound, and k_{cat} and K_{m} to denote free-enzyme constants:

- $k'_{\mathrm{cat}} > k_{\mathrm{cat}}$ and $\frac{k'_{\mathrm{cat}}}{K'_{\mathrm{m}}} > \frac{k_{\mathrm{cat}}}{K_{\mathrm{m}}}$. In this situation both the initial slope and the maximal rate of the reaction are elevated with the molecule bound; the molecule is thus an **activator** across all substrate concentrations.

- $k'_{\mathrm{cat}} < k_{\mathrm{cat}}$ and $\frac{k'_{\mathrm{cat}}}{K'_{\mathrm{m}}} < \frac{k_{\mathrm{cat}}}{K_{\mathrm{m}}}$. In this situation both the initial slope and the maximal rate of the reaction are *lowered* with the molecule bound; the molecule is thus an **inhibitor** across all substrate concentrations.

- $k'_{\mathrm{cat}} > k_{\mathrm{cat}}$ but $\frac{k'_{\mathrm{cat}}}{K'_{\mathrm{m}}} < \frac{k_{\mathrm{cat}}}{K_{\mathrm{m}}}$. In this situation the initial slope is lowered with the molecule bound, but the maximal rate of the reaction is elevated; the molecule will thus act as an *inhibitor* at low substrate concentrations, but as an *activator* at high levels of substrate.

- $k'_{\text{cat}} < k_{\text{cat}}$ but $\frac{k'_{\text{cat}}}{K'_{\text{m}}} > \frac{k_{\text{cat}}}{K_{\text{m}}}$. In this situation the initial slope is increased with the molecule bound, while the maximal rate of the reaction is lowered; the molecule will thus act as an *activator* at low substrate concentrations, but as an *inhibitor* at high levels of substrate.

These general concepts still rely on the fundamental assumption that the Michaelis–Menten model is an appropriate way to describe the kinetics of the enzyme in both states. Deviations from the Michaelis–Menten model can lead to even more complex behaviors.

5.3.3 Cooperative enzymes — The Hill Equation

The kinetics of a Michaelis–Menten-like enzyme (regulated or otherwise) have a linear dependence on substrate at low substrate concentrations and a saturated (constant) rate at high levels of substrate. This has important ramifactions for the "sensitivity" of how the rate changes with increasing substrate. Consider, for example, the question of how much substrate concentration must increase to change the rate from 10% of the maximal value to 90% (a 9-fold increase in rate). We simply solve the Michaelis–Menten equation for an rate of an arbitrary fraction (x) of the maximal rate:

$$xV_{\max} = \frac{V_{\max}[S]}{K_{\text{m}} + [S]}$$

which rearranges to:[3]

$$[S] = \frac{x}{(1-x)} K_{\text{m}} \tag{5.51}$$

This gives $[S] = \frac{1}{9}K_{\text{m}}$ for $x = 10\%$ and $[S] = 9K_{\text{m}}$ for $x = 90\%$, which means that substrate concentrations must increase by 81-fold in order to achieve this (9-fold) increase in rate ($\frac{[S]_{90\%}}{[S]_{10\%}} = \frac{9K_{\text{m}}}{9^{-1}K_{\text{m}}} = 81$). In general, for any shift from y% to 100-y%, we have $\frac{[S]_{100-y\%}}{[S]_{y\%}} = \frac{(1-y)y^{-1}K_{\text{m}}}{y(1-y)^{-1}K_{\text{m}}} = \frac{(y-1)^2}{y^2}$, thus, to change from 1% to 99% of the maximal rate requires a concentration change of $\frac{99^2}{1^2} = 9801$, an almost 10,000-fold change!

Note that this equation also allows us to specify the substrate concentration at which we reach half-maximal activity: $[S]_{1/2} = \frac{0.5}{1-0.5}K_{\text{m}} = K_{\text{m}}$; that is, K_{m} can be thought of as the substrate concentration that gives have maximal activity.

How might we gain a sharper response (a larger change in rate for a smaller change in concentration)? Consider the following equation, known as the **Hill equation**:

$$\frac{d[P]}{dt} = \frac{V_{\max}[S]^h}{K_{1/2}^h + [S]^h} \tag{5.52}$$

Note that this reduces to the Michaelis–Menten equation if $h = 1$, and that if $[S] = K_{1/2}$:

$$\frac{d[P]}{dt} = \frac{V_{\max}K_{1/2}^h}{K_{1/2}^h + K_{1/2}^h} = \frac{V_{\max}K_{1/2}^h}{1K_{1/2}^h} = \frac{1}{2}V_{\max}$$

[3] $xV_{\max} = \frac{V_{\max}[S]}{K_{\text{m}}+[S]} \rightarrow xV_{\max}K_{\text{m}} + xV_{\max}[S] = V_{\max}[S] \rightarrow xV_{\max}K_{\text{m}} = (1-x)V_{\max}[S] \rightarrow \frac{xK_{\text{m}}}{(1-x)} = [S]$

Thus, $K_{1/2}$, like K_{m}, is the substrate concentration that gives have maximal activity.

At high substrate concentration ($[S] >> K_{1/2}$):

$$\frac{d[P]}{dt} = \frac{V_{\max}[S]^h}{K_{1/2}^h + [S]^h} \approx \frac{V_{\max}[S]^h}{[S]^h} = V_{\max} \tag{5.53}$$

which is a saturated response just like the Michaelis–Menten result. However at low substrate concentration ($[S] << K_{1/2}$):

$$\frac{d[P]}{dt} = \frac{V_{\max}[S]^h}{K_{1/2}^h + [S]^h} \approx \frac{V_{\max}[S]^h}{K_{1/2}^h} \tag{5.54}$$

In this regime, the rate changes with the h-th power of $[S]$ (or $\frac{[S]}{K_{1/2}}$). If $h > 1$, then this curve will fall *below* $y = x$ for all $\frac{[S]}{K_{1/2}} < 1$; the rate increases *more slowly* with increasing concentration than does the Michaelis–Menten equation. For $\frac{[S]}{K_{1/2}} > 1$, however, the curve lies *above* the $y = x$ line, and thus the Hill equation gives a *sigmoidal* transition from low to high rate with increasing concentration. The sharpness of this curve increases with increasing h; we often say that increasing h makes the transition more "step-like". Note that the Hill equation shares many qualitative characteristics with the logistic equation used in population models; the primary difference is that at low values the logistic equation follows an exponential curve, rather than a power-law.

Now let's revisit the question of how much substrate concentration must be changed to achieve a given change in rate:

$$xV_{\max} = \frac{V_{\max}[S]^h}{K_{1/2}^h + [S]^h}$$

rearranges to:[4]

$$[S] = \sqrt[h]{\frac{x}{(1-x)}} K_{1/2} \tag{5.55}$$

In general, for any shift from y% to 100-y%, we thus have:

$$\frac{[S]_{100-y\%}}{[S]_{y\%}} = \frac{\sqrt[h]{(1-y)y^{-1}}K_{1/2}}{\sqrt[h]{y(1-y)^{-1}}K_{1/2}} = \sqrt[h]{\frac{(y-1)^2}{y^2}} \tag{5.56}$$

Thus, for the two cases we considered previously, we have $\frac{[S]_{90\%}}{[S]_{10\%}} = \sqrt[h]{\frac{90^2}{10^2}} = \sqrt[h]{81}$ and $\frac{[S]_{99\%}}{[S]_{1\%}} = \sqrt[h]{\frac{99^2}{1^2}} = \sqrt[h]{9801}$. If $h = 2$, we thus require a 9-fold change in concentration to achieve a 9-fold change in rate, and a 99-fold change in $[S]$ to increase the rate 99-fold. With $h = 4$, a three-fold increase in concentration would increase the rate by a factor of 9, and an increase in $[S]$ of slightly below 10 ($\sqrt{99}$) would increase the rate by 99-fold.

[4] $xV_{\max} = \frac{V_{\max}[S]^h}{K_{1/2}^h+[S]^h} \rightarrow xV_{\max}K_{1/2}^h + xV_{\max}[S]^h = V_{\max}[S]^h \rightarrow xV_{\max}K_{1/2}^h = (1-x)V_{\max}[S]^h \rightarrow \frac{xK_{1/2}^h}{(1-x)} = [S]^h \rightarrow \frac{x^{1/h}K_{1/2}}{(1-x)^{1/h}} = [S]$

Note that *hemoglobin*, the protein in red blood cells that carries oxygen and carbon dioxide between the lungs and tissues, binds oxygen with a Hill-like behavior in terms of oxygen levels and a Hill coefficient (h) of just under 3. Oxygen pressure in the lungs is roughly 0.2 atm, while oxygen pressure in the tissues is about 0.025 atm or less, approximately an 8-fold difference. With a Michaelis–Menten-like binding profile ($h = 1$), this would provide (at best) less than a three-fold difference in the amount of oxygen in hemoglobin leaving and returning from the lungs; with $h = 3$ hemoglobin leaving the lungs can contain more than 20 times the oxygen of that returning from the tissues. This allows for a near perfect absorption of oxygen (and release of CO_2) in the high-oxygen environment of the lungs, and near perfect release of oxygen (and absorption of CO_2) in the low-oxygen environment of the tissues.

A mechanistic derivation of the Hill equation

The Hill equation was not initially derived, but rather was postulated (by Archibald Hill, in 1910) as a reasonable model for precisely the problem of oxygen transport by hemoglobin mentioned above. However, it can be shown that the Hill equation can be mapped directly onto chemical-kinetic model, as with Michaelis–Menten.

Consider an enzyme with two identical binding (active) sites which each follow a Michaelis–Menten mechanism. With one substrate molecule binding and reacting, we have:

$$E + S \underset{k_{-1}[ES]}{\overset{k_{+1}[E][S]}{\rightleftharpoons}} ES \overset{k_{+2}[ES]}{\longrightarrow} E + P \tag{5.57}$$

However, there is also a possibility that a substrate molecule will binding to an enzyme already containing susbrate in the other site, and thus we also have:

$$ES + S \underset{k'_{-1}[ES_2]}{\overset{k'_{+1}[ES][S]}{\rightleftharpoons}} ES_2 \overset{k'_{+2}[ES_2]}{\longrightarrow} ES + P \tag{5.58}$$

where the rates in this case may or may not be different than the singly-bound case. If the rates of the two reactions are the same (the two sites are *independent*) then it is just as if there was twice as much enzyme present, and thus:

$$\frac{d[P]}{dt} = \frac{2k_{\mathrm{cat}}[E]_{tot}[S]}{K_{\mathrm{m}} + [S]} \tag{5.59}$$

with $K_{\mathrm{m}} = \frac{k_{-1}+k_{+2}}{k_{+1}}$ as before.

If the rates of the second (doubly-bound) reaction is faster than the first, then we say the sites are **cooperative** (if the doubly-bound reaction is slower, the term anti-cooperative is used); cooperativity requires:

$$k'_{\mathrm{cat}} > k_{\mathrm{cat}} \qquad \text{or} \qquad \frac{k'_{\mathrm{cat}}}{K'_{\mathrm{m}}} > \frac{k_{\mathrm{cat}}}{K_{\mathrm{m}}} \tag{5.60}$$

In the extreme case, we may have both $k_{\mathrm{cat}} = 0$ (the first reaction does not occur at all) and $K'_{\mathrm{m}} << K_{\mathrm{m}}$. Note that if $k_{\mathrm{cat}} = k_{+2} = 0$, then $K_{\mathrm{m}} = \frac{k_{-1}}{k_{+1}}$, which is just an equilibrium

dissociatation constant, $K_{\rm m} = \frac{[E][S]}{[ES]}$. Since there is only one reaction forming product:

$$\frac{d[P]}{dt} = k'_{\rm cat}[ES_2] \tag{5.61}$$

and again, we wish to find the quasi-steady-state value of $[ES_2]$. From the quasi-state assumption ($\frac{d[ES_2]}{dt} = 0$), we have:

$$\frac{d[ES_2]}{dt} = k'_{+1}[ES][S] - k'_{-1}[ES_2] - k'_{\rm cat}[ES_2] = 0 \tag{5.62}$$

which gives:

$$k'_{+1}[ES][S] = k'_{-1}[ES_2] + k'_{\rm cat}[ES_2] \rightarrow [ES][S] = \frac{k'_{-1} + k'_{\rm cat}}{k'_{+1}}[ES_2] = K'_{\rm m}[ES_2] \tag{5.63}$$

The conservation of mass of enzyme also requires:

$$[E]_{\rm tot} = [E] + [ES] + [ES_2] \tag{5.64}$$

Since we also know that $K_{\rm m} = \frac{[E][S]}{[ES]}$ and thus $[ES] = \frac{[E][S]}{K_{\rm m}}$, these results become:

$$[ES][S] = K'_{\rm m}[ES_2] \rightarrow \frac{[E][S]^2}{K_{\rm m}} = K'_{\rm m}[ES_2] \rightarrow [E] = \frac{K_{\rm m}K'_{\rm m}[ES_2]}{[S]^2} \tag{5.65}$$

and

$$[E]_{\rm tot} = [E] + [ES] + [ES_2] \rightarrow [E]_{\rm tot} = \left(1 + \frac{[S]}{K_{\rm m}}\right)[E] + [ES_2] \tag{5.66}$$

which combine to give:

$$\begin{aligned}
[E]_{\rm tot} &= \left(1 + \frac{[S]}{K_{\rm m}}\right)\frac{K_{\rm m}K'_{\rm m}[ES_2]}{[S]^2} + [ES_2] && (5.67)\\
&= \left(\frac{K_{\rm m}K'_{\rm m}}{[S]^2} + \frac{[S]}{K_{\rm m}}\frac{K_{\rm m}K'_{\rm m}}{[S]^2} + 1\right)[ES_2] && (5.68)\\
&= \left(\frac{K_{\rm m}K'_{\rm m}}{[S]^2} + \frac{K'_{\rm m}}{[S]} + 1\right)[ES_2] && (5.69)\\
&= \left(\frac{K_{\rm m}K'_{\rm m}}{[S]^2} + \frac{K'_{\rm m}][S]}{[S]^2} + \frac{[S]^2}{[S]^2}\right)[ES_2] && (5.70)\\
&= \left(\frac{K_{\rm m}K'_{\rm m} + K'_{\rm m}[S] + [S]^2}{[S]^2}\right)[ES_2] && (5.71)\\
& && (5.72)
\end{aligned}$$

$$[ES_2] = \left(\frac{[E]_{\rm tot}[S]^2}{K_{\rm m}K'_{\rm m} + K'_{\rm m}[S] + [S]^2}\right) \tag{5.73}$$

Thus, we have the result that:

$$\frac{d[P]}{dt} = k'_{\rm cat}[ES_2] = \frac{k'_{\rm cat}[E]_{\rm tot}[S]^2}{K_{\rm m}K'_{\rm m} + K'_{\rm m}[S] + [S]^2} \tag{5.74}$$

If $K_{\rm m}$ is very large with respect to substrate concentration ($K_{\rm m} >> [S]$), then the first term in the denominator will dominate the second, and we have:

$$\frac{d[P]}{dt} = k'_{\rm cat}[ES_2] = \frac{k'_{\rm cat}[E]_{\rm tot}[S]^2}{K_{\rm m}K'_{\rm m} + [S]^2} \tag{5.75}$$

which is precisely the Hill equation ($h = 2$), with $V_{\rm max} = k'_{\rm cat}[E]_{\rm tot}$ and $K_{1/2} = \sqrt{K_{\rm m}K'_{\rm m}}$.

The same result can be obtained by considering that, when $K'_{\rm m} << K_{\rm m}$, the levels of $[ES]$ will generally be small compared to $[ES_2]$, since substrate will preferentially form the doubly-bound complex over the singly-bound state. Thus:

$$[E]_{\rm tot} = [E] + [ES] + [ES_2] \approx [E] + [ES_2] \tag{5.76}$$

which gives:

$$\begin{aligned}
[E]_{\rm tot} &= \frac{K_{\rm m}K'_{\rm m}[ES_2]}{[S]^2} + [ES_2] && (5.77)\\
&= \left(\frac{K_{\rm m}K'_{\rm m}}{[S]^2} + 1\right)[ES_2] && (5.78)\\
&= \left(\frac{K_{\rm m}K'_{\rm m} + [S]^2}{[S]^2}\right)[ES_2] && (5.79)\\
& && (5.80)
\end{aligned}$$

$$[ES_2] = \left(\frac{[E]_{\rm tot}[S]^2}{K_{\rm m}K'_{\rm m} + [S]^2}\right) \tag{5.81}$$

and:

$$\frac{d[P]}{dt} = k'_{\rm cat}[ES_2] = \frac{k'_{\rm cat}[E]_{\rm tot}[S]^2}{K_{\rm m}K'_{\rm m} + [S]^2} \tag{5.82}$$

Using similar approximations for an enzyme with n active-sites, and a strong preference for the ES_n state, will lead to the Hill equation with $h = n$.

5.4 Pathways and networks of reactions

In the previous sections we have worked through the mathematics of simple models for enzymatic reactions under quasi-steady-state. While all the results were derived using a mechanistic model and certain assumptions about relative rates, it should be noted that in many cases these models are applied without a fundamental understanding of the mechanistics details of the reaction. Rather, these equations are used as general functional forms, with parameters fit to observed data. If, for example, the Hill equation with $h = 3$ is a good fit to observation, then it may be a reasonable mathematical model to use in predicting rates; however, it should not be presumed that this implies that the assumptions used in the

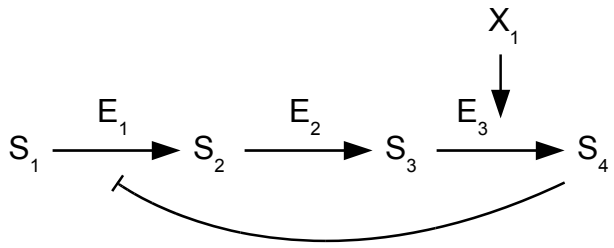

Figure 5.2: An example metabolic pathway schematic.

derivation are true. Also, it is important to note that when a model is a good fit to some set of data, predictions that *interpolate* between observations are generally much more robust than those that require as significant degree of *extrapolation* outside the region where the model was fit.

Now, given a set of basic functional forms for enzymatic reactions, as well as the more fundamental mass-action kinetics laws, we can consider how to use these in describing pathways and networks of biochemical reactions. We may classify these in different ways, and some of the classes of particular interest are metabolic pathways, signal transduction networks, and gene-regulatory networks. Pathways and networks are often used interchangeably, particular in the context of metabolism and signal transduction; if any distinction can be made, pathways tend to be more linear in their topology, while networks are more interconnected.

5.4.1 Metabolic pathways

A metabolic network is a set of enzyme-catalyzed reactions which work together to achieve a single large-scale chemical conversion (metabolism). The glycolytic pathway and Kreb's cycle, which convert glucose into energy (in the form of ATP) is a perfect example. The components of a metabolic pathway are:

- A set of enyzymes that carry out catalysis ($\{E_i\}$).
- A set of molecules that are substrates and products for the various reactions of the pathway ($\{S_i\}$).
- A set of regulatory molecules that act an inhibitors and/or activators of the enzymes ($\{X_i\}$); in some cases these regulators are, in fact, also substrates and/or products of reactions in the network.

We may express such a network by a set of arrows that connect the substrate and product of each reaction; each of these may be labeled by the appropriate enzyme. Regulation is indicated by a second set of arrows that connect the regulatory molecule to the reaction that it controls; an arrowhead is customarily used to indicate activation, and a bar is used to indicate inhibition; an example is given in Figure 5.2.

Each reaction can then be associated with a differential equation that gives the rate of that particular reaction — this may be the Michaelis–Menten equation, the Hill equation, or any of the other general forms. For the example given, we may have:

1. Catalysis of S_1 to S_2 by E_1, inhibited by S_4 (competetively):

$$\nu_1 = \frac{k_{\mathrm{cat},1}[E_1]_{\mathrm{tot}}[S_1]}{K_{\mathrm{m},1}\left(1 + \frac{[S_4]}{K_{I4,1}}\right) + [S_1]}$$

2. Catalysis of S_2 to S_3 by E_2, following simple Michaelis–Menten mechanism:

$$\nu_2 = \frac{k_{\mathrm{cat},2}[E_2]_{\mathrm{tot}}[S_2]}{K_{\mathrm{m},2} + [S_2]}$$

3. Catalysis of S_3 to S_4 by E_3, requiring activator X_1:

$$\nu_3 = \frac{k_{\mathrm{cat},3}[E_3]_{\mathrm{tot}}[S_3]}{K_{\mathrm{m},3}\left(1 + \frac{K_{X1,3}}{[X_1]}\right) + [S_3]}$$

Given the rate of each underlying reaction, we can write a differential equation for the rate of change of each chemical species by adding together the rates of all reactions in which the species is either a substrate or product; the contribution is negative when the species is a substrate (as it is consumed during the reaction) and positibe when it is a product. If a single reaction creates more than one molecule of product (or uses more than one molecule of reactant), the contribution is multiplied by this number (called the *stoichiometric coefficient*).

Since enzymes are neither created nor consumed in any of these reactions, we generally take the total enzyme concentrations to all be constant. Furthermore, since our general forms are based on quasi-steady-state conditions and are given in terms of total enzyme concentration, we do not need to be concerned with any time-dependence of enzyme levels. In other words, we only need to define the equations for the rates of change of the substrates and products of the pathway. If there are activators or inhibitors that are *not* substrates or products these only need to be considered explicitly if their concentrations vary in some time-dependent manner.

This results in a system of differential equations, one for each substrate/product; for the example:

$$\begin{aligned}
\frac{d[S_1]}{dt} &= -\nu_1 \\
&= -\frac{k_{\text{cat},1}[E_1]_{\text{tot}}[S_1]}{K_{\text{m},1}\left(1+\frac{[S_4]}{K_{I4,1}}\right)+[S_1]} && (5.83)\\
\frac{d[S_2]}{dt} &= +\nu_1 - \nu_2 \\
&= \frac{k_{\text{cat},1}[E_1]_{\text{tot}}[S_1]}{K_{\text{m},1}\left(1+\frac{[S_4]}{K_{I4,1}}\right)+[S_1]} - \frac{k_{\text{cat},2}[E_2]_{\text{tot}}[S_2]}{K_{\text{m},2}+[S_2]} && (5.84)\\
\frac{d[S_3]}{dt} &= +\nu_2 - \nu_3 \\
&= \frac{k_{\text{cat},2}[E_2]_{\text{tot}}[S_2]}{K_{\text{m},2}+[S_2]} - \frac{k_{\text{cat},3}[E_3]_{\text{tot}}[S_3]}{K_{\text{m},3}\left(1+\frac{K_{X1,3}}{[X_1]}\right)+[S_3]} && (5.85)\\
\frac{d[S_4]}{dt} &= +\nu_3 \\
&= \frac{k_{\text{cat},3}[E_3]_{\text{tot}}[S_3]}{K_{\text{m},3}\left(1+\frac{K_{X1,3}}{[X_1]}\right)+[S_3]} && (5.86)
\end{aligned}$$

This is fundamentally the same type of model as we have seen both for population dynamics and epidemiological modeling, and the same numerical methods can be used. For example, given some set of initial concentrations, we can simulate the variation in the concentration of all chemical species in the system as a function of time, or we can attempt to solve the system (numerically) for stationary points.

Note that for the above example, $[S_1]$ is strictly decreasing, and $[S_4]$ is strictly increasing. As a result, the long term behavior of the system can only result in complete conversion of $[S_1]$ to $[S_4]$. However, imagine that there is some external supply of $[S_1]$, delivered at some rate, and that there is a sink for $[S_4]$ to leave the system (again at some rate); biologically these might represent transporters which control influx and and eflux from the cell, or externally controlled variables in a controlled experiment. To model such a system, we simply add these terms to the system of differential equations. If the influx/eflux rates are constant (given by RR_{in} and R_{out}), then the system becomes:

$$\frac{d[S_1]}{dt} = R_{in} - \frac{k_{\mathrm{cat},1}[E_1]_{\mathrm{tot}}[S_1]}{K_{\mathrm{m},1}\left(1+\frac{[S_4]}{K_{I4,1}}\right)+[S_1]} \tag{5.87}$$

$$\frac{d[S_2]}{dt} = \frac{k_{\mathrm{cat},1}[E_1]_{\mathrm{tot}}[S_1]}{K_{\mathrm{m},1}\left(1+\frac{[S_4]}{K_{I4,1}}\right)+[S_1]} - \frac{k_{\mathrm{cat},2}[E_2]_{\mathrm{tot}}[S_2]}{K_{\mathrm{m},2}+[S_2]} \tag{5.88}$$

$$\frac{d[S_3]}{dt} = \frac{k_{\mathrm{cat},2}[E_2]_{\mathrm{tot}}[S_2]}{K_{\mathrm{m},2}+[S_2]} - \frac{k_{\mathrm{cat},3}[E_3]_{\mathrm{tot}}[S_3]}{K_{\mathrm{m},3}\left(1+\frac{K_{X1,3}}{[X_1]}\right)+[S_3]} \tag{5.89}$$

$$\frac{d[S_4]}{dt} = \frac{k_{\mathrm{cat},3}[E_3]_{\mathrm{tot}}[S_3]}{K_{\mathrm{m},3}\left(1+\frac{K_{X1,3}}{[X_1]}\right)+[S_3]} - R_{out} \tag{5.90}$$

Here, we have both positive and negative terms for *all* components, and thus we may reach some non-trival stationary point. This type of model can be used to determine optimal conditions for industrial-level biological reactions, such as the production of ethanol from corn.

5.4.2 Signal transduction networks

Signal transduction refers to the process by which external stimuli — changes in enviroment, the presence of chemicals released by nearby cells, and so on — are converted into changes in cellular behavior. In many cases, the external stimuli are sensed by protein receptors on the surface of the cell. In this case, the signal transduction process must convert this extracellular event to biochemical changes *inside* the cell. In many cases, these signals are propogated through direct regulation of the activity of enzymes and the physical interaction of non-enzymatic proteins with each other and with various enzymes. This regulation may be achieved in one of several ways:

Physical association: The properties of a protein complex (two or more physically associated proteins) may be different than the properties of the isolated proteins. For example, the equilibrium binding constant for another protein may be changed or the constants of an enzymatic reaction (k_{cat} and K_{m}) may be perturbed.

Chemical modification: Proteins may be chemically modified in several ways. Among other modifications, phosphate groups can be added in a site-specific manner to serines, threonines and tyrosines by enzymes called **kinases** and removed by **phosphatases**; lysines may have between one and three methyl groups added, or a single actyl group, again in a dynamic manner; and the sugar *N*-acetyl glucosamine can be added to serines and threonines, sometimes at positions that can also be phosphorylated. These modifications can them alter both the association and catalytic constants of the modified proteins.

Binding of second messengers: Second messengers are small molecules whose concentrations in the cell are tightly coupled to signal transduction events; cyclic AMP (cAMP), calcium ions (Ca^{2+}), inositol triphosphate (IP_3) and diacylglycerol (DAG) are among

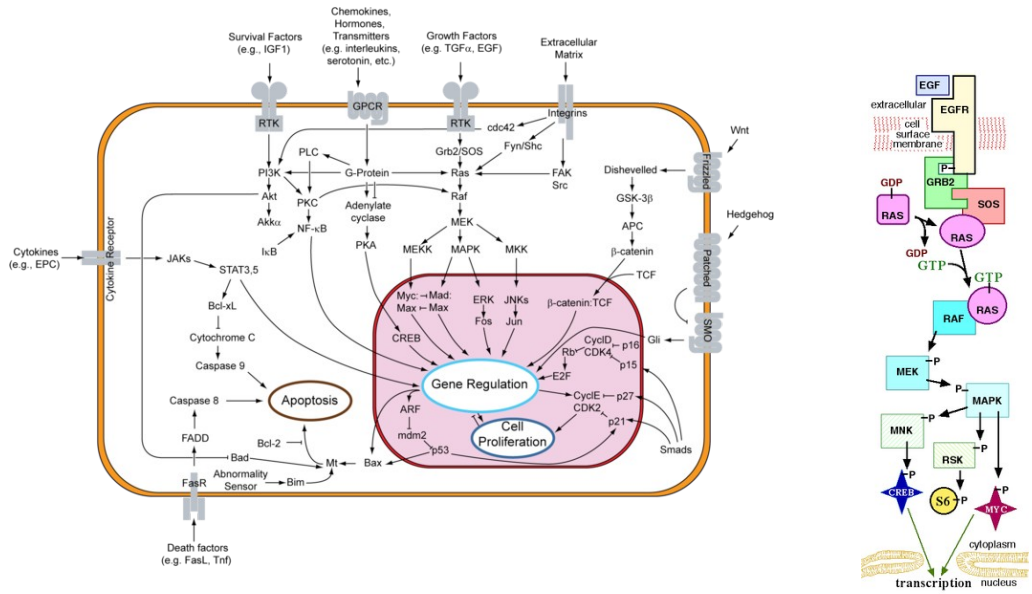

Figure 5.3: A schematic of (a) the main signal transduction pathways controlling cellular growth and differentiation; and (b) the Epidermal Growth Factor (EGF) signaling pathway.

the most common of these. Second messengers can bind to various proteins, again regulating their ability to interact with other proteins or to catalyze particular chemical reactions.

As it is the relative levels of various protein (enzyme) components that are are direct interest, these are the key variables in a signal transduction network. Again, the model is initially laid out as a schematic, relating how each component of the system can change, and how this impacts overall function. Then, each reaction is described a single rate law, and these are combined into a system of differential equations describing the rate of change of each species. Note that in this type of system, each form of a protein must be considered separately: free (unmodified) protein, phosphorylated protein forms, and all complexes containing the protein must be considered as different biochemical species.

As an example, consider signaling by epidermal growth factor, which plays a central role in regulating cellular growth. A cartoon-based schematic of this pathway, along how this pathway fits into the larger network that governs cellular growth and development are shown in Figure 5.3. While these abstract, cartoon-like models can help give an intuitive understanding of the overall mechanisms that drive biology, they lack mechanistic detail toand thus can not be directly mapped onto a mathematical model. As a more detailed schematic, we may write the system of chemical reactions that the cartoon model attempts to capture.

- EGF is able to bind to EGFR (on its extracellular domain), in a reversible reaction; this complex is then able to dimerize, again reversibly:

$$EGF + EGFR \rightleftharpoons EGF.EGFR \tag{5.91}$$

$$2EGF.EGFR \rightleftharpoons (EGF.EGFR)_2 \tag{5.92}$$

- The cytoplasmic domain of EGFR is a kinase; in the dimeric complex, each kinase is able to catalyze phosphorylation of the other member of the complex. This is essentially an irreversible reaction, but a phosphatase can catalyze the reverse reaction:

$$(EGF.EGFR)_2 \longrightarrow (EGF.EGFR^*)_2 \tag{5.93}$$

$$(EGF.EGFR)_2 \xleftarrow{P} (EGF.EGFR^*)_2 \tag{5.94}$$

- Phosphorylated receptor can bind reversibly to GRB2, and this complex can, in turn, bind reversibly to SOS:

$$(EGF.EGFR^*)_2 + GRB2 \rightleftharpoons (EGF.EGFR^*)_2.GRB2 \tag{5.95}$$

$$(EGF.EGFR^*)_2.GRB2 + SOS \rightleftharpoons (EGF.EGFR^*)_2.GRB2.SOS \tag{5.96}$$

- The receptor/GRB2/SOS complex is a catalyst that promotes the exchange of GDP for GTP in the protein Ras; Ras is itself an enzyme that coverts GTP back to GDP, but does not release the product on it's own:

$$Ras.GDP \xrightarrow{(EGF.EGFR^*)_2.GRB2.SOS} Ras.GTP \tag{5.97}$$

$$Ras.GDP \longleftarrow Ras.GTP \tag{5.98}$$

- Ras.GTP can bind reversibly to the kinase Raf (and thus activate it):

$$Ras.GTP + Raf \rightleftharpoons Ras.GTP.Raf \tag{5.99}$$

- Raf is a kinase active when bound to Ras.GTP; it phosphorylates another kinase, MEK, again activating it; phosphorylated Mek in turn phosphorylates another kinase, MAPK (also known as ERK). This is known as a MAP kinase cascade, and each step is reversible only by the action on another enzyme (a phosphatase):

$$MEK \xrightarrow{Ras.GTP.Raf} MEK^* \tag{5.100}$$

$$MEK \xleftarrow{P} MEK^* \tag{5.101}$$

$$MAPK \xrightarrow{MEK^*} MAPK^* \tag{5.102}$$

$$MAPK \xleftarrow{P} MAPK^* \tag{5.103}$$

- MAPK (MAP kinase) acts as a *master regulator* with many targets that it phosphorylates. These include other kinases (such as MNK and RSK, each of which target other proteins) and transcription factors, such as Myc. Again, phosphatases are required to reverse these reactions.

$$MNK \xrightarrow{MAPK^*} MNK^* \tag{5.104}$$

$$MNK \xleftarrow{P} MNK^* \tag{5.105}$$

$$RSK \xrightarrow{MAPK^*} RSK^* \tag{5.106}$$

$$RSK \xleftarrow{P} RSK^* \tag{5.107}$$

$$Myc \xrightarrow{MAPK^*} Myc^* \tag{5.108}$$

$$Myc \xleftarrow{P} Myc^* \tag{5.109}$$

- The targets of MNK include transcription factors such as CREB, and the targets of RSK include ribosomal proteins (S6):

$$CREB \xrightarrow{MNK^*} CREB^* \tag{5.110}$$

$$CREB \xleftarrow{P} CREB^* \tag{5.111}$$

$$S6 \xrightarrow{RSK^*} S6^* \tag{5.112}$$

$$S6 \xleftarrow{P} S6^* \tag{5.113}$$

Phosphoryation of transcription factors like CREB and Myc directly affect *transcription* of various genes; phosphorylation of the ribosome directly affects the *translation* of transcribed genes (mRNA) into protein. Thus, there is a change in the overall levels of many different proteins in the cell, which results in a change of cellular state (cell growth and replication). The next section will outline how this affects might be modeled.

From these chemical schematics, we can use our models of biochemical reactions (developed in the previous sections) to write rates laws for each of these reactions. For molecular association, we use second-order mass-action kinetics:

$$EGF + EGFR \longrightarrow EGF.EGFR: \qquad \nu_{+1} = k_{+1}[EGF][EGFR] \tag{5.114}$$

and for dissocation, first order:

$$EGF.EGFR \longrightarrow EGF + EGFR: \qquad \nu_{-1} = k_{-1}[EGF.EGFR] \tag{5.115}$$

with similar expresssions for all binding reactions. For enzyme-catalyzed reactions, we may use the Michaelis–Menten form (or another model, as appropriate):

$$S6 \xrightarrow{RSK^*} S6^*: \qquad \nu_{+14} = \frac{k_{cat,+14}[RSK^*][S6]}{K_{m,+14} + [S6]} \tag{5.116}$$

$$S6^* \xrightarrow{P} S6: \qquad \nu_{-14} = \frac{k_{cat,-14}[P][S6^*]}{K_{m,-14} + [S6^*]} \tag{5.117}$$

All phosphorylation and dephosphorylation steps, with the exception of the initial phosphorylation of the receptor, will follow a similar model. We also have to special reactions: the autophosphorylation of the EGFR dimer, and the conversion of bound GTP to GDP by

Ras. In both of these cases, the molecular association step of catalysis has already been considered, and thus we only have a first-order chemical step to describe:

$$(EGF.EGFR)_2 \longrightarrow (EGF.EGFR^*)_2 : \qquad \nu_{+3} = k_{+3}[(EGF.EGFR)_2] \tag{5.118}$$

$$Ras.GTP \longrightarrow Ras.GDP : \qquad \nu_{-6} = k_{-6}[Ras.GTP] \tag{5.119}$$

In both of these cases, the reaction in the other direction follows an enzyme catalyzed model, for which we may use the Michaelis–Menten expression:

$$(EGF.EGFR^*)_2 \xrightarrow{P} (EGF.EGFR)_2 : \qquad \nu_{-3} = \frac{k_{cat,-3}[P][(EGF.EGFR)_2]}{K_{m,-3} + [(EGF.EGFR)_2]} \tag{5.120}$$

$$Ras.GDP \xrightarrow{(EGF.EGFR^*)_2.GRB2.SOS} Ras.GTP : \qquad \nu_{+6} = \frac{k_{cat,+6}[E2GS][Ras.GDP]}{K_{m,+6} + [Ras.GDP]} \tag{5.121}$$

where, $[E2GS]$ has been used as shorthand for $[(EGF.EGFR^*)_2.GRB2.SOS]$.

5.4.3 Gene-regulatory networks

Gene-regulatory networks describe the system that control the transcription of specific genes in a given cell under various conditions. Fundamentally, genes are transcribed (that is, the mRNA corresponding to the gene is synthesized) by the binding of RNA polymerase to the genes initiation site, followed by a processive process of synthesis with RNA polymerase moving along the gene until a termination sequence is reached. However, not all genes are equally accessible by RNA polymerase, and many proteins, know as **transcription factors** regulate this accessiblity. *Activators* or *enhancers* are proteins that increase the expression of a gene by physically associating with the DNA in the genes regulatory region; *repressors* may also bind in the regulatory region, but have the effect of decreasing gene expression.

The level of gene expression may be thought of as the rate of production of mRNA; when this is high, more mRNA will accumulate over time. Since mRNA is the template for protein synthesis, the concentration of mRNA will directly influence the rate of production of the protein of the gene; again, higher levels of mRNA will lead to accumulation of higher concentrations of protein. Thus, for a given gene, we can describe two processes: mRNA synthesis (dependent on different transcription factors) and degradation, and protein synthesis (dependent on mRNA levels) and degradation. If we use the notation X for degradation rates and Ω for synthesis, we will have the general form for each mRNA/protein pair (A):

$$\frac{d[A_{prot}]}{dt} = \Omega_{A,prot} - X_{A,prot} \tag{5.122}$$

$$\frac{d[A_{rna}]}{dt} = \Omega_{A,rna} - X_{A,rna} \tag{5.123}$$

The simplest form of the degradation terms is a unimolecular elementary reaction, corresponding to a constant rate of degradation per unit concentration:

$$X_{A,prot} = \chi_{A,prot}[A_{prot}] \qquad X_{A,rna} = \chi_{A,rna}[A_{rna}] \tag{5.124}$$

where χ is the degradation constant. This model assumes that degradation is not regulated by any other species; in certain cases protein degradation can, in fact, be highly regulated. The simplest form for protein synthesis will be a rate that is linearly related to the concentration of mRNA of that gene:

$$\Omega_{A,prot} = \omega_A[A_{rna}] \tag{5.125}$$

with ω_A the rate per unit concentration. This essentially makes the assumption that there is an unlimited supply of translational material (ribosomes, tRNA, amino acids). The rate of mRNA synthesis will depend on the concentations of the transcription factors which regulate the gene, $\Omega_{A,rna} = f([T_i], [T_j], ...)$, but the precise functional form will depend on the mechanism of regulation. Thus, we obtain the general form:

$$\frac{d[A_{prot}]}{dt} = \omega_A[A_{rna}] - \chi_{A,prot}[A_{prot}] \tag{5.126}$$

$$\frac{d[A_{rna}]}{dt} = \Omega_{A,rna}([T_i], [T_j], \cdots) - \chi_{A,rna}[A_{rna}] \tag{5.127}$$

The proteins corresponding to different genes have different functions, and some of these are to act as transcription factors — it is this that leads to the concept of a *network* of regulation. The expression of one transcription factor can have a direct impact on the expression of others (and of itself). Thus, our basic two component (mRNA and protein) model for a single gene can be extended, with the mRNA expression step dependent on the concentration of one or more proteins in the network.

Now let us discuss some of the possible models for activation of a gene by a transcription factor. All our models will be based on several key assumptions:

- In the absence of any (explictly modeled) transcription factor, each gene is transcribed at a constant rate, μ_A^0.

- When transcription factors are appropriately bound to the promoter region of a gene, the gene is transcribed at a constant rate of μ_A^i, where i denotes the particular DNA–transcription factor complex of interest.

- Binding of transcription factors to DNA is assumed to be much faster than the initiation of transcription.

The last point is particularly important, as it addresses one of the fundamental challenges in modeling of gene expression — DNA is present as only a small number of discrete molecules (1 or 2 in most cases). If a gene has only a single transcription factor binding site, and exists in only one copy in a cell, then only one of two states (transcription factor bound or unbound) can exist at any given time; with two copies of a gene (and a single binding site) this becomes three states (one, two, or no copies bound). These discrete changes are difficult to deal with using continuous concentration variables, but if transcription factors associate and dissociate from the gene much faster than transcriptional activation, then we may assume that an equilibrium state of binding will be reached. This equilibrium will

describe the fraction of the gene spent in each state, ξ_A^i, in a time-averaged sense, which leads to a general form of the rate of transcription:

$$\Omega_{A,rna} = \xi_A^0 \mu_A^0 + \sum_{i=1}^{N} \xi_A^i \mu_A^i \tag{5.128}$$

where we have allowed for multiple transcription factor complexes with different effects on transcription (N is the total number of such complexes). Note that, as we are dealing with fractions of the DNA in each state:

$$\xi_A^0 + \sum_{i=1}^{N} \xi_A^i = 1 \qquad \rightarrow \qquad \xi_A^0 = 1 - \sum_{i=1}^{N} \xi_A^i \tag{5.129}$$

This allows us to rewrite the general expression:[5]

$$\Omega_{A,rna} = \mu_A^0 + \sum_{i=1}^{N} \xi_A^i \left(\mu_A^i - \mu_A^0\right) \tag{5.130}$$

where the term in parantheses $(\mu_A^i - \mu_A^0)$ is the difference in the transcription factor bound and unbound rates.

If we have a single possible transcription factor binding complex, this becomes:

$$\Omega_{A,rna} = \mu_A^0 + \xi_A \left(\mu_A - \mu_A^0\right) \tag{5.131}$$

with a single variable ξ_A that must be determined. We wish to obtain ξ_A as a function of the concentrations of our transcription factors, and this will depend on the nature of the complex that is formed. First consider the simplest possible mechanism — a single transcription factor binds to the DNA:

$$DNA + TF \rightleftharpoons DNA.TF \tag{5.132}$$

Equilbrium binding in this model is given by the dissociation constant:

$$K_d = \frac{[DNA][TF]}{[DNA.TF]} \tag{5.133}$$

Recall that the dissociation constant is the reciprocal of the binding, or association, constant. Now, the total concentration of DNA is, as discussed above, a constant, discrete value directly related to the number of copies of the gene present. For simplicity, we choose units of DNA concentration such that the total concentration is 1 ($[DNA] + [DNA.TF] = 1$). This leads to an equivalence that $[DNA.TF] = \xi_A$ (and that $[DNA] = \xi_A^0 = 1 - \xi_A$) and thus:[6]

$$K_d = \frac{(1-\xi_A)[TF]}{\xi_A} \qquad \rightarrow \qquad \xi_A = \frac{[TF]}{(K_d + [TF])} \tag{5.134}$$

[5] $Omega_{A,rna} = \xi_A^0\mu_A^0 + \sum_{i=1}^{N} \xi_A^i\mu_A^i = \left(1 - \sum_{i=1}^{N} \xi_A^i\right)\mu_A^0 + \sum_{i=1}^{N} \xi_A^i\mu_A^i = \mu_A^0 - \sum_{i=1}^{N} \xi_A^i\mu_A^0 + \sum_{i=1}^{N} \xi_A^i\mu_A^i = \mu_A^0 + \sum_{i=1}^{N} \left(\xi_A^i\mu_A^i - \xi_A^i\mu_A^0\right) = \mu_A^0 + \sum_{i=1}^{N} \xi_A^i \left(\mu_A^i - \mu_A^0\right)$

[6] $K_d = \frac{(1-\xi_A)[TF]}{\xi_A} \rightarrow K_d\xi_A = (1-\xi_A)[TF] \rightarrow K_d\xi_A = [TF] - \xi_A[TF] \rightarrow (K_d + [TF])\xi_A = [TF] \rightarrow \xi_A = \frac{[TF]}{(K_d+[TF])}$

Note that this has a form similar to the Michaelis–Menten equation, with a linear regime at low $[TF]$ and a saturated regime at high $[TF]$. The overall rate of transcription is then:

$$\Omega_{A,rna} = \mu_A^0 + \left(\mu_A - \mu_A^0\right) \frac{[TF]}{(K_d + [TF])} \tag{5.135}$$

Another common mechanism of transcription activation involves dimerization of the transcription factor as a prerequisite for binding; this may involve either *homodimerization* (two identical models binding) or *heterodimerization* (two different molecules binding). In the case of homodimerization, the result from above becomes:

$$\xi_A = \frac{[(TF)_2]}{(K_d + [(TF)_2])} \tag{5.136}$$

If we also assume that the transcription dimerization is fast, then we can also write the equilibrium expression:

$$2TF \rightleftharpoons (TF)_2 \qquad K_{d_o} = \frac{[TF]^2}{[(TF)_2]} \tag{5.137}$$

These combine (using $[(TF)_2]) = \frac{[TF]^2}{K_{d_o}}$ to give:

$$\xi_A = \frac{\frac{[TF]^2}{K_{d_o}}}{K_d + \frac{[TF]^2}{K_{d_o}}} = \frac{[TF]^2}{K_d K_{d_o} + [TF]^2} \tag{5.138}$$

which has the form of the Hill equation, which a cooperativity coefficient of 2 and $K_{1/2} = K_d K_{d_o}$ (an effective half-maximal concentration. The overall rate of transcription is:

$$\Omega_{A,rna} = \mu_A^0 + \left(\mu_A - \mu_A^0\right) \frac{[TF]^2}{K_d K_{d_o} + [TF]^2} \tag{5.139}$$

If we have heterodimerization, we have:

$$\xi_A = \frac{[TF_a.TF_b]}{(K_d + [TF_a.TF_b])} \tag{5.140}$$

and:

$$TF_a + TF_b \rightleftharpoons TF_a.TF_b \qquad K_{d_o} = \frac{[TF_a][TF_b]}{[TF_a.TF_b]} \tag{5.141}$$

which combine give:

$$\xi_A = \frac{[TF_a][TF_b]}{K_d K_{d_o} + [TF_a][TF_b]} \tag{5.142}$$

and:

$$\Omega_{A,rna} = \mu_A^0 + \left(\mu_A - \mu_A^0\right) \frac{[TF_a][TF_b]}{K_d K_{d_o} + [TF_a][TF_b]} \tag{5.143}$$

Now, there are two extremes of transcription factor action:

1. An ideal **activator** may be essential for transcription, with a basal rate (μ_A^0) of zero.

2. An ideal **repressor** may completely shut down transcription, with the bound-state rate (μ_A) zero.

Using the homodimerization-based model as an example, we have activation ($\mu_A^0 = 0$) giving:

$$\Omega_{A,rna} = \frac{\mu_A[TF]^2}{K_d K_{d_o} + [TF]^2} \tag{5.144}$$

and repression ($\mu_A = 0$) giving

$$\Omega_{A,rna} = \mu_A^0 \left(1 - \frac{[TF]^2}{K_d K_{d_o} + [TF]^2}\right) \tag{5.145}$$

Coupling of signal transduction and gene-regulatory networks

Note that in many cases, gene-regulatory networks are directly coupled to signal transduction networks; a transcription factor many only function effectively when phosphorylated by a specific kinase, which is in turn activated by a signaling cascade. This is how long-term cellular responses can be achieved (such as cellular differentiation) by short term signals (such as those during particular stages of development). How can we integrate this sort of regulation into our models?

Consider a system with one kinase (E), one phosphatase (P) and one transcription factor (TF), with the kinase acting on the transciption factor. Also assume that the transcription factor regulates both itself and the kinase (as an activator), and that it must be phosphorylated and homodimerize in order to function. Assume the phosphatase is expressed in an unregulated manner. We thus have three RNA synthesis reactions:

$$\Omega_{E,rna} = \frac{\mu_E[TF^*]^2}{K_{1/2,E} + [TF^*]^2}, \quad \Omega_{TF,rna} = \frac{\mu_{TF}[TF^*]^2}{K_{1/2,TF} + [TF^*]^2}, \quad \Omega_{P,rna} = \mu_P^0 \tag{5.146}$$

where TF^* indicates the phosphorylated form. These are incorporated into the differential equations for the rates of RNA production:

$$\frac{d[E_{rna}]}{dt} = \frac{\mu_E[TF^*]^2}{K_{1/2,E} + [TF^*]^2} - \chi_{E,rna}[E_{rna}] \tag{5.147}$$

$$\frac{d[TF_{rna}]}{dt} = \frac{\mu_{TF}[TF^*]^2}{K_{1/2,TF} + [TF^*]^2} - \chi_{TF,rna}[TF_{rna}] \tag{5.148}$$

$$\frac{d[TF_{rna}]}{dt} = \mu_P^0 - \chi_{TF,rna}[TF_{rna}] \tag{5.149}$$

and we have the standard forms for protein production of enzyme and phosphatase:

$$\frac{d[E_{prot}]}{dt} = \omega_E[E_{rna}] - \chi_{E,prot}[E_{prot}] \tag{5.150}$$

$$\frac{d[P_{prot}]}{dt} = \omega_P[P_{rna}] - \chi_{P,prot}[P_{prot}] \tag{5.151}$$

For the transcription factor, we need to be careful, because we additionally we have the enzymatic phosphorylation and dephosphorylation, whose rates may be given:

$$\nu_k = \frac{k_{cat,k}[E_{prot}][TF_{prot}]}{K_{m,k} + [TF_{prot}]}, \qquad \nu_p = \frac{k_{cat,p}[P_{prot}][TF^*]}{K_{m,p} + [TF^*]} \tag{5.152}$$

The kinase activity decreases unphosphorylated $[TF]$, and the phosphatase reaction increases it. Along with the synthesis and degradation terms we have:

$$\frac{d[TF_{prot}]}{dt} = \omega_{TF}[TF_{rna}] - \chi_{TF,prot}[TF_{prot}] - \frac{k_{cat,k}[E_{prot}][TF_{prot}]}{K_{m,k} + [TF_{prot}]} + \frac{k_{cat,p}[P_{prot}][TF^*]}{K_{m,p} + [TF^*]} \tag{5.153}$$

Finally, we also must consider the phosphorylated form of the enzyme; protein synthesis will only make the unphosphorylated form, but we might assume that both forms are degraded at equal rates. Adding in the kinase and phosphatase rates we have:

$$\frac{d[TF^*]}{dt} = -\chi_{TF,prot}[TF^*] + \frac{k_{cat,k}[E_{prot}][TF_{prot}]}{K_{m,k} + [TF_{prot}]} - \frac{k_{cat,p}[P_{prot}][TF^*]}{K_{m,p} + [TF^*]} \tag{5.154}$$

These seven differential equations then fully describe the rules for the dynamics of the system; numerical methods can be used to understand the behavior of the system in more detail.