Chapter 6

Modeling Molecular Processes

Elementary cellular processes

In the last couple of chapters, we have used ODEs to model different biological problems. Each ODE has a function that describes the dynamics of some processes. You must be wondering how we develop those functions.

All biological phenomenon involves a large number of molecules participating in various biochemical processes. Often, we represent those processes as a molecular network. If you look carefully, you will find that we can group those processes in certain elementary biochemical processes. For example, ligand-receptor binding, transcription, translation, enzymatic reaction, etc. are elementary processes in molecular networks.

Biologists use certain functions or ODEs that capture the dynamics of these elementary processes. Once we learn those mathematical formulations, we can use them in different combinations to capture large network dynamics. At least that is what we hope.

In this chapter, we will learn how to develop ODE-based models for the elementary biochemical processes. Some of these ODEs are constructed following the Law of Mass action, even though the processes are not elementary chemical reactions.

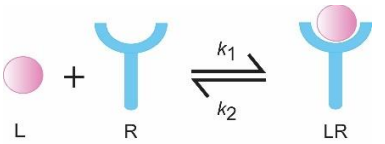
Often, we will not have detail mechanistic information about processes. In all such cases, we will use mathematical functions to capture the dynamics of the process. Those functions may not have one to one correspondence to the real molecular mechanism.

In our modeling approach, we will always use Occam's razor or parsimony principle– the simplest model that explains the phenomenon adequately is the best model. Keeping your model simple is an art, as there is always a threat of oversimplification.

Binding-unbinding

Binding of two molecules, like a ligand and its cognate receptor, is a frequent elementary process in cell biology. Binding and unbinding do not involve any covalent bond formation or breaking but depend upon non-covalent interactions.

Here in the following example, ligand L binds reversibly to the receptor R to form a complex LR.



Following system of ODEs represent the dynamics of ligand-receptor interactions,

$$\frac{d[L]}{dt} = -k_1[L][R] + k_2[LR] \tag{1}$$

$$\frac{d[R]}{dt} = -k_1[L][R] + k_2[LR] \tag{2}$$

$$\frac{d[LR]}{dt} = k_1[L][R] - k_2[LR] \tag{3}$$

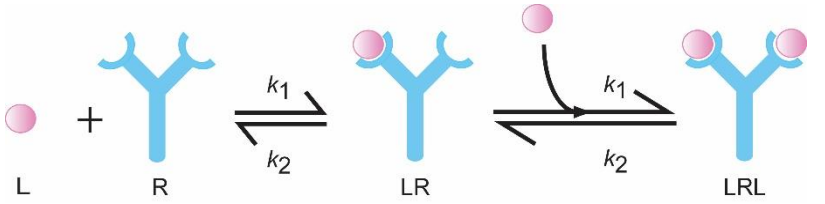
Here, k_1 and k_2 are rate constants. [L], [R], and [LR] are molar concentrations of free L, free R, and LR complex, respectively.

Conventionally, for molecular interactions in biology, k_1 and k_2 are not measured separately. Instead, one measures the affinity of the interaction in terms of the Dissociation Constant K_d. The dissociation constant is the inverse of the Equilibrium Constant (K_{eq}). By definition,

$$k_d = \frac{1}{K_{eq}} = \frac{[L][R]}{[LR]}$$

[L], [R], and [LR] are the molar concentrations of L, R, and LR at equilibrium.

Sometimes, more than two molecules bind to form a complex. For example, suppose R is a bivalent receptor, having two identical sites to bind ligand L. Therefore, R and L can form two types of complexes, RL and LRL. The following scheme represents the process.



We have considered the same rate constants, k_1, and k_2 for both the steps. However, this may not be true in some cases, and binding one ligand may help bind another molecule.

For the bivalent binding, the ODEs for LR and LRL will take the following forms,

$$\frac{d[LR]}{dt} = 2k_1[L][R] - k_2[LR] - k_1[L][LR] + 2k_2[LRL] \tag{4}$$

$$\frac{d[LRL]}{dt} = k_1[L][LR] - 2k_2[LRL] \tag{5}$$

Note that LR can be formed either by binding L and R or by dissociating one L from LRL. R is bivalent. So, L can bind to R in two ways. That is why we have 2 in $2k_1[L][R]$. Similarly, one ligand dissociates from LRL to form LR. As there are two identical L in LRL, the dissociation can happen in two ways. Therefore, the rate of dissociation of LRL is also doubled.

If the ligand is bivalent or multivalent (as it happens in antigen-antibody binding), the stoichiometry of binding and unbinding get more complicated. You need to keep track of all the possible ways a complex can form or break.

Production and degradation

Molecules, like proteins, RNA, are produced and get degraded. When detail mechanistic information is not available, we can consider production as a zero-order process and degradation as a first-order process. So, for a molecule X, we can write,

$$\frac{dX}{dt} = k_s - k_d X \tag{6}$$

Here, k_S and k_d are rate constants for the production and degradation of X, respectively. In some cases, a signal may induce the production of X. We can represent such a system with the following ODE,

$$\frac{dX}{dt} = k_s S - k_d X \tag{7}$$

Here S is a measure of the signal strength. It can be constant or a time-dependent variable. Similarly, a signal can affect degradation too. We may use the following ODE for such a system,

$$\frac{dX}{dt} = k_s - k_d SX \tag{8}$$

Transcription and translation

Protein synthesis involves transcription, followed by translation. Transcription factors control the rate of transcription. Often the transcription of a gene is controlled by external signals through transcription factors. For example, signaling by growth factors, like EGF, activates pathways that eventually induce the expression of genes involved in the cell cycle. The extent of gene expression, therefore, the commitment to proliferation, is modulated by the strength of the signal. Biologists perform experiments where they modulate the transcription of a gene. For example, we can vary the level of expression of a gene under the control of a *Lac* operon, in *E. coli*, by varying the amount of IPTG in the culture medium. Similar inducible and controlled expression systems also exist for Eukaryotic systems.

In systems where transcriptional control is crucial, protein production is not modeled as a simple zero or first-order process. Instead, the process is broken in transcription and translation to accommodate transcriptional control of an input signal.

Interestingly, it has been observed that for many inducible systems, the extent of induction has a sigmoidal relation with the strength of the signal. At low signal strength, the level of transcription is very low. With increases in signal strength, the rate of transcription increases slowly. However, once the signal crosses a threshold, the

transcription rate increases sharply with little increase in signal. Eventually, the system saturates, and a further increase in signal does not alter the level of expression.

Transcription factors are often dimeric or multimeric, and they show cooperativity in binding to the promoter. One transcription factor may have multiple binding sites in the same promoter, and binding to one site facilitates binding to other sites. Such cooperativity gives rise to the sigmoidal behavior in the dose-response curve for transcription.

Cooperativity has been observed in other biological systems like the binding of oxygen to hemoglobin. Without going into the mechanistic details of cooperativity, one can model such sigmoidal behavior using the Hill function. Hill function is used frequently to capture the non-linear dose-response behavior of transcription. The general form of Hill function is,

$$y = \frac{x^n}{K^n + x^n} \tag{9}$$

Here, n and K are called Hill coefficient and Hill constant, respectively. The behavior of the Hill function is explained in Figure 1 with different values of n and K.

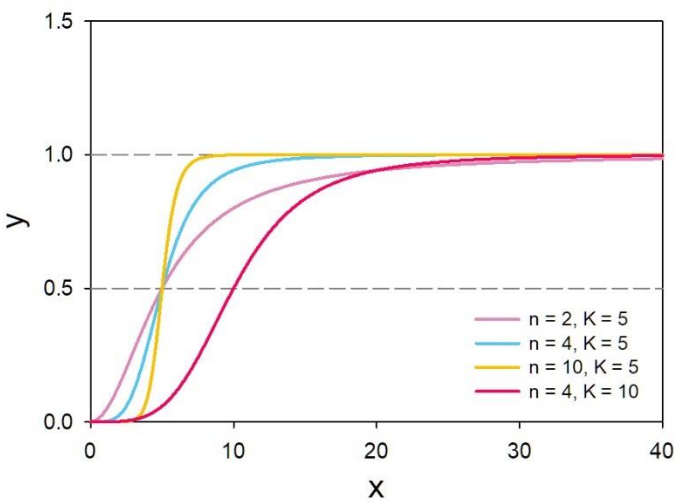


Figure 1: Behaviours of Hill function for different values of Hill coefficient (n) and Hill constant (K).

In the Hill function, for $x \geq 0$, y varies between zero and one ($0 \leq y \leq 1$). The Hill coefficient, n, determines how fast y reaches to one. To understand this, compare the curve for $n = 2$ and $n = 4$ in Figure 1. For both the curves K = 5. Hill coefficient is a measure of cooperativity. When $n = 1$, we get a rectangular hyperbola. The y vs x curve is sigmoidal when $n > 1$. When n is very high, the sigmoidal curve changes very sharp, just like a switch (check the curve for n =10, K =5). In most transcriptional systems, n varies from 2 to 4.

The Hill constant, K, is the value of x, for which $y = 0.5$. In other words, K determines the value of x for which y is half-saturated. To understand this, compare the curve for $n = 4$, $K = 5$, with $n = 4$, $K = 10$, in Figure 1.

Some transcription factor inhibits gene expression. Often, such inhibition also involves cooperativity among transcription factors and shows a sigmoidal behavior. For transcriptional inhibition, we consider that the cooperativity is negative and the Hill function is accordingly modified (Equation 10).

$$y = \frac{x^{-n}}{K^{-n} + x^{-n}} = \frac{K^n}{K^n + x^n} \tag{10}$$

The behavior of inverse Hill function for different values of n and K is shown in Figure 2.

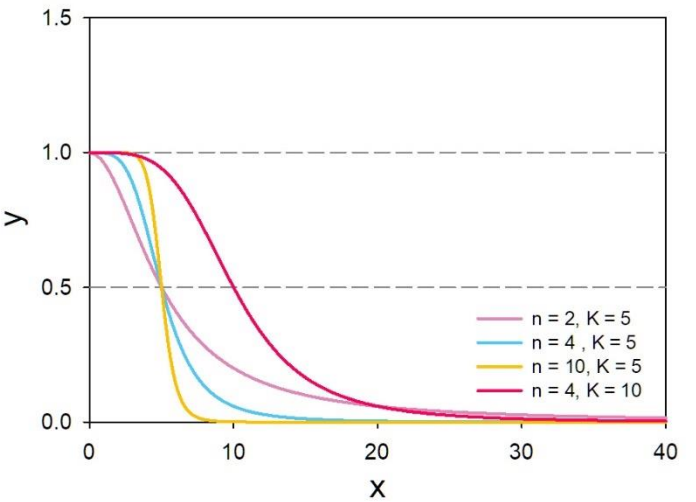


Figure 2: Behaviours of inverse Hill function for different values of Hill coefficient (n) and Hill constant (K).

Now, we will build an ODE-based model for gene expression using Hill function. We will consider transcription and translation as two separate processes. Suppose, *S* is a signal or transcription factor that induces transcription of the gene to produce the mRNA, *m*. The mRNA is translated into protein *P*. Both mRNA and the protein have finite life-time and get degraded. The following system of ODEs can be used to model the system,

$$\frac{dm}{dt} = k_b + k_r \frac{S^n}{K^n + S^n} - k_m m \tag{11}$$

$$\frac{dP}{dt} = k_l m - k_p P \tag{12}$$

Here, k_b is the rate of basal level of transcription of the gene that happens even in absence of any inducing signal. k_r is the maximum rate of transcription upon induction, and $k_b << k_r$. k_m and k_P are first-order rate constants for degradation of the mRNA and the protein, respectively.

k_l is the rate constant for translation. Note that the rate of translation depends upon the concentration of the mRNA, but concentration of mRNA does not change by translation.

When the signal inhibits gene expression we will replace the Hill function in Equation 11, with an inverse Hill function.

In some cases, particularly in Eukaryotic systems, we can assume that transcription is much faster than translation and concentration of mRNA reaches the steady state before the start of translation. With this assumption, we can combine Equations 11 and 12 into a single ODE.

At the steady state, $\frac{dm}{dt} = 0$. Therefore, following Equation 11, the steady state value of *m* is,

$$m_{ss} = \frac{k_b}{k_m} + \frac{k_r}{k_m} \frac{S^n}{K^n + S^n} \tag{13}$$

Using m_{ss} in Equation 12, we can write

$$\frac{dP}{dt} = k_l m_{ss} - k_P P$$
$$= \frac{k_l k_b}{k_m} + \frac{k_l k_r}{k_m} \frac{S^n}{K^n + S^n} - k_P P \quad (14)$$

Clubbing multiple parameters in Equation 14 we get,

$$\frac{dP}{dt} = q_b + q_s \frac{S^n}{K^n + S^n} - k_P P \quad (15)$$

Here, $q_b = k_1 k_b / k_m$, and $q_s = k_l k_r / k_m$.

It is recommended that transcription and translation must be considered explicitly for modeming an inducible system. However, that requires two ODEs for each gene in the model. That may add computational difficulties in modelling large gene expression network. That's why Equation 15 is frequently used in modelling transcriptional networks, even in prokaryotic systems and without any explicit assumption of steady state for mRNA.

Enzymatic reactions

Most of the cellular processes involve enzymes. Though most of the enzymes are proteins, substrates are of diverse types, from small molecules like Glucose to large proteins. In an enzymatic reaction, the enzyme forms an enzyme-substrate complex and eventually facilitates substrate to product conversion. Like any other catalyst total concentration of an enzyme does not change in this reaction.

Usually, we are ignorant of the mechanism of catalysis by an enzyme. Then the simplest way to think of an enzymatic reaction as a second order reaction and use the following ODE,

$$\frac{d[P]}{dt} = k_1 [E][S]$$

here, $[P]$, $[E]$, and $[S]$ are the molar concentration of product, enzyme and substrate, respectively.

Sometime the substrate to product conversion is reversible and two different enzymes catalyse those. For example, a kinase phosphorylates a target protein and corresponding phosphatases remove the phosphorylation from that protein. For such reversible process, involving two enzymes, E_1 and E_2, we can use the following ODE,

$$\frac{d[P]}{dt} = k_1[E_1][S] - k_2[E_2][S]$$

One can go into the details of an enzymatic reaction and consider an elaborate reaction scheme as shown below. Here, the enzyme forms a complex with the substrate that converts into product-enzyme complex. Eventually, the product is released. All the steps are reversible.

$$S + E \rightleftharpoons SE \rightleftharpoons PE \rightleftharpoons P + E$$

We can simplify this scheme, by making some assumptions. Usually, the interconversion between substrate-enzyme complex (SE) and product-enzyme complex (PE) is very fast. Therefore, we can consider them as a single entity, C as shown in the following scheme.

$$S + E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C \xrightarrow{k_3} P + E$$

The second assumption is that the product does not bind to the enzyme. Therefore, conversion from C to product and Enzyme is unidirectional. With these two assumptions, we can write the following ODEs,

$$\frac{d[S]}{dt} = -k_1[S][E] + k_2[C]$$

$$\frac{d[E]}{dt} = -k_1[S][E] + k_2[C] + k_3[C]$$

$$\frac{d[C]}{dt} = k_1[S][E] - k_2[C] - k_3[C]$$

$$\frac{d[P]}{dt} = k_3[C]$$

This scheme is useful, if we model a system with a handful of enzymatic reactions. However, it is difficult to analyze or simulate a large network like signal transduction network using this approach. The difficulties are of two-fold. Every enzymatic step need four ODEs. That increases complexity of the model. Secondly, we need to know (or guess) three rate constants for each enzymatic reaction and usually these rate constants are not measured in experiments.

We can reduce the number ODEs and parameters, with the assumption of quasi-steady state for C. This assumption is valid when the total concentration of the enzyme is much smaller than total concentration of the substrate. In stricter mathematical terms, quasi-steady state for C is achieved only when $[E]_T \ll [S]_T + \frac{k_2+k_3}{k_1}$. $[E]_T$ and $[S]_T$ are total concentrations of enzyme and substrate, respectively.

With the assumption of quasi-steady state, you can rearrange those four ODEs and derive the famous Michaelis–Menten equation for an enzymatic reaction,

$$\frac{d[P]}{dt} = \frac{k_3[E]_T[S]}{\frac{k_2+k_3}{k_1}+[S]} = \frac{v_{\max}[S]}{K_M+[S]} \tag{16}$$

$[S]$ and $[E]$ are concentrations of free substrate and free enzyme. $k_3.[E]_T = v_{\max}$ is the maximum rate of product formation and $(k_2+k_3)/k_1 = K_M$, is the Michaelis–Menten constant. K_M is the substrate concentration at which the rate of product formation is half of $v_{\max}$. The rate constant k_3 is called the turnover number.

In experimental enzymology, Michaelis–Menten equation is widely used to understand enzyme kinetics. Therefore, Michaelis–Menten constant and turnover numbers are known for many enzymes.

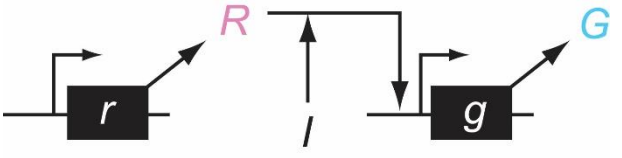
Michaelis–Menten equation is widely used in mathematical modeling of enzymatic networks, even when we are not sure of the validity of the quasi-steady state assumption in a particular system. However, I will suggest against liberal use of Michaelis–Menten equation in modeling molecular networks.

Exercises

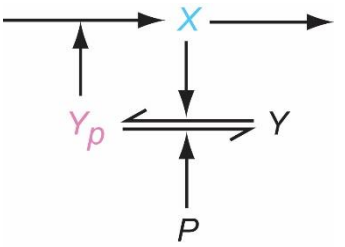
1. An inducer I, induces the expression of a reporter gene G in *E. coli*. The induction has a sigmoidal behavior. Derive the relation between the dose of I and the steady state level of the reporter protein. Also, explain this relation schematically.

2. A cell surface receptor R is internalized by endocytosis. The endocytosed receptor returns to the surface by exocytosis. k_{in} and k_{out} are rate constants for endocytosis and exocytosis, respectively. Calculate the fraction of the total receptor internalized in time *t*. Consider that the total number of R is conserved, and at t = 0, all R was on the cell surface.

3. R is a transcription regulator that binds to an inducer I and then induces the expression of the gene g. R is produced constitutively. Create an ODE-based model for this system. Use the Hill function appropriately.



4. Y is a protein that gets phosphorylated to Yp. This reaction is catalyzed by X following Michaelis-Menten kinetics. Yp gets de-phosphorylated by P. This reaction also follows Michaelis-Menten kinetics. The amount of P is constant. X is produced constitutively at a rate *r* and gets degraded by a first-order process. Yp can induce the production of X, and that is a first-order process. Create an ODE-based model for this system. Also, make a rough schematic diagram for X and Yp nullclines. Consider, the total amount of Y, phosphorylated, and non-phosphorylated, is constant.



5. A small network involving two molecules, X and Y, is shown below. The external signal S induces X and Y production. These two molecules are not produced in the absence of S. X gets degraded by a first-order reaction. X controls the degradation of Y. Y does not get degraded in the absence of X. Create a simplified model for this system and check whether the system will have bifurcation with respect to S.