BIO723: Scientific Computing for Biologists

Paul M. Magwene
Fall 2012

November 20, 2012

In this lecture I focus on developing a few simple analytical models. The exposition and models described here are drawn from:

1. Uri Alon. Network motifs: theory and experimental approaches. *Nat Rev Genet*, 8(6):450–61, Jun 2007. doi:10.1038/nrg2102
2. Shai S Shen-Orr, Ron Milo, Shmoolik Mangan, and Uri Alon. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet*, 31(1):64–68, May 2002. doi:10.1038/ng881
3. U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman and Hall, 2007

1 Reaction Equations

The chemical kinetics "Law of Mass Action" states that the rate of a reaction is proportional to the concentrations of the reactants. For example, if the reaction is given by:

$$A \xrightarrow{k} B$$

then

$$\frac{d[A]}{dt} = -k[A] \qquad \text{and} \qquad \frac{d[B]}{dt} = k[A]$$

where t is time and k is a constant reaction rate, and where the brackets indicate the concentrations of the reactants.

Similarly for bimolecular reactions:

$$A + B \xrightarrow{k} C$$

the reaction rates are given by:

$$\begin{aligned} \frac{d[A]}{dt} &= -k[A][B] \\ \frac{d[B]}{dt} &= -k[A][B] \\ \frac{d[C]}{dt} &= k[A][B] \end{aligned}$$

This can be extended to an arbitrary number of reactants.

1.1 Reversible Reactions

Most chemical and biological reactions are reversible:

$$A \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} B$$

At equilibrium none of the concentrations are changing ($k_1[A] = k_{-1}[B]$) and therefore

$$[A][B] = k_{-1}/k_1 = K_{\text{eq}} \qquad \text{(the equilibrium constant)}$$

For a reversible bimolecular reaction:

$$A + B \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C$$

and at equilibrium $k_1[A][B] = k_{-1}[C]$ so that $[A][B]/[C] = k_{-1}/k_1 = K_{eq}$. For bimolecular reactions k_1 is often refered to as the association constant (k_{on}) while k_{-1} is the dissociation constant (k_{off}).

2 Enzyme-catalyzed reactions

Enzymes accelerate the rates of reactions, but do not affect their equilibrium. In other words, enzymes increase the values of the k's but do not affect K_{eq}.

The general scheme for an enzymatic reaction is:

$$\text{S} + \text{E} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \text{ES} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \text{E} + \text{P}$$

This describes an enzyme, E binding it's substrate S to form an enzyme-substrate complex ES. At the end of the reaction the substrate S is converted into product P and the enzyme is unaltered. Most if not all enzymatic reactions are reversible.

Enzymatic reactions do not follow mass action kinetics because there is usually a limited amount of enzyme present. If we assume that the rate at which $E + P$ forms is very fast compared to the rate at which ES dissociates back to $E + S$, then after a very brief period, a steady-state will be stablished in which $[ES]$ is essentially constant over time. We further assume the P does not accumulate (e.g. it is very rapidly removed by another reaction), so the overall scheme becomes:

$$\text{S} + \text{E} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \text{ES} \xrightarrow{k_2} \text{E} + \text{P}$$

Under those assumptions, simple enzymatic reactions are well approximated by the Michaelis-Menten equation, usually written as:

$$v = \frac{V_{\max}[S]}{K_m + [S]}$$

where $v = k_2[ES]$, $V_{\max} = k_2[E_t]$ where $[E_t] = [E] + [ES]$ and $K_m = (k_2 + k_{-1})/k_1$.

A graph showing the rate of reaction, v, as a function of substrate concentration $[S]$ has a hyperbolic shape as shown in Fig. 1

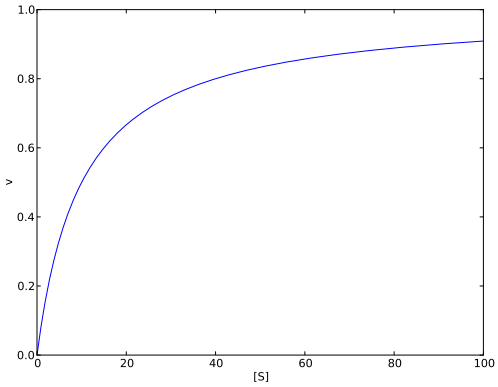

Figure 1: A curve described by the Michaelis-Menten equation with $V_{\max} = 1$ and $K_m = 10$.

3 Hill Equation

Biochemical reactions that are more complex then the simple enzymatic reactions that the Michaelis-Menten equation describes are often described by the Hill Equation. The Hill equation was originally derived in the context of cooperative enzymatic reactions where an enzyme binds several substrate molecules at the same time and the binding of each subsequent substrate molecules enhances the affinity of the enzyme for the substrate. This sort of cooperative reaction is described by the formula:

$$v = \frac{V_{\max}[S]^n}{K_d + [S]^n}$$

where $V_{\max}$ is the maximum velocity of the reaction, $[S]$ is the concentration of the substrate, and K_d is the dissociation constant. The Michaelis-Menten equation can be viewed as a special case of the Hill equation with $n = 1$.

The Hill equation, written in this form, describes a sigmoid relationship between the substrate concentration and the velocity of the reaction. The larger the value of n the more 'step-like' the relationship as illustrated in the Fig. 2. The Hill equation is also used to describe non-enzymatic reactions. For example, consider a transcription factor, X, that positively regulates the transcription of gene Y. We describe the rate of production of Y as $f(X)$ where:

$$f(X) = \frac{\beta X^n}{K^n + X^n}$$

where K is the 'activation coefficient' and is related to the affinity between X and it's binding sites. β is the maximum expression level of the promoter, and again n governs the steepness of the function.

If rather than activating Y's transcription, X is a transcriptional repressor we can write:

$$f(X) = \frac{\beta}{1 + (X/K)^n}$$

which yields curves like those shown in Fig. 3. Remember that both of these equation describe the production of Y as a function of the levels of X, not the temporal dynamics of Y which we'll look at after developing a few more ideas.

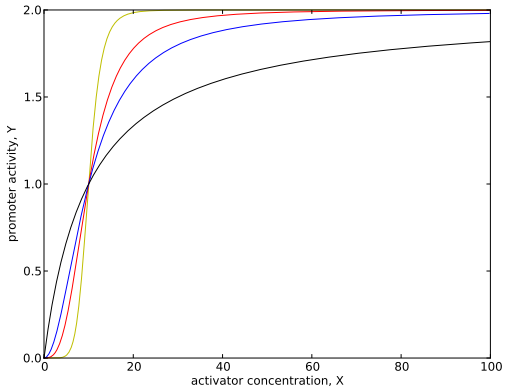

Figure 2: Hill functions with varying degrees of Hill coefficients describing transcriptional activation – $n = 1$ (black), $n = 2$ (blue), $n = 3$ (red), $n = 8$ (yellow).

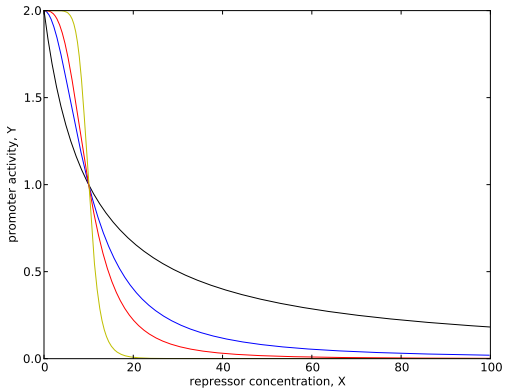

Figure 3: Hill functions with varying degrees of Hill coefficients describing transcriptional repression – $n = 1$ (black), $n = 2$ (blue), $n = 3$ (red), $n = 8$ (yellow).

4 Simplifying Models using Logic Approximations

To simplify analysis it's often convenient to approximate step-like sigmoidal functions like those produced by the Hill equation with functions using logic approximations. We'll assume that a gene, Y, is either on or off – i.e. $f(X) = 0$ or $f(X) = \beta$. To do this we can rewrite the formula for Y as:

$$f(X) = \beta\ \Theta(X > K)$$

where the function Θ is zero if the statement inside the parentheses is false or one if the statement is true.

When X is a repressor we can write:

$$f(X) = \beta\ \Theta(X < K)$$

4.1 Multi-dimensional Input Functions

What if a gene needs two or more activator proteins to be transcribed? We can describe the amount of Z transcribed as a function of active forms of X and Y with a function like:

$$f(X,Y) = \beta\ \Theta(X > K_x)\Theta(Y > K_y)$$

The above equation describes 'AND' logic (i.e. *both* X and Y have to be above their threshold levels, K_x and K_y, for Z to be transcribed). In a similar manner we can define 'OR' logic:

$$f(X,Y) = \beta\ \Theta(X > K_x \text{ or } Y > K_y)$$

A SUM function would be defined like this:

$$f(X,Y) = \beta_x X + \beta_y Y$$

4.2 Dynamics of the Logic Approximation

Again, let's assume X is a transcriptional activator of Y. For the moment let's assume there is a constant concentration of X (above the threshold K_y) so that the rate of production of Y is given by $f(X) = \beta$. Y is lost due to degradation and dilution at a rate, α, proportional to the amount of Y. A differential equation to describe the change of Y over time is:

$$\frac{dY}{dt} = \beta - \alpha Y$$

At steady state $dY/dt = 0$, which means that the rate of production and the rate of degradation are perfectly balanced ($\beta = \alpha Y$) and we can calculate the steady state value, Y_{st} as:

$$Y_{st} = \frac{\beta}{\alpha}$$

Assume Y has come to it's steady state. What happens if we take away the activating signal, X? (i.e. β suddenly becomes zero). In that case:

$$\frac{dY}{dt} = -\alpha Y$$

and solving for Y as a function of time, t, we get:

$$Y(t) = Y_{st} e^{-\alpha t}$$

In a similar manner we can consider the case where Y is initially zero. If we solve the differential equation $dY/dt = \beta - \alpha Y$ we find:

$$Y(t) = Y_{st}(1 - e^{-\alpha t})$$

4.3 Response Time

The response time of a system is defined as the time to reach halfway between the initial and final levels in a dynamic process. We'll designate this $T_{1/2}$. If we solve one of the above equations (decay

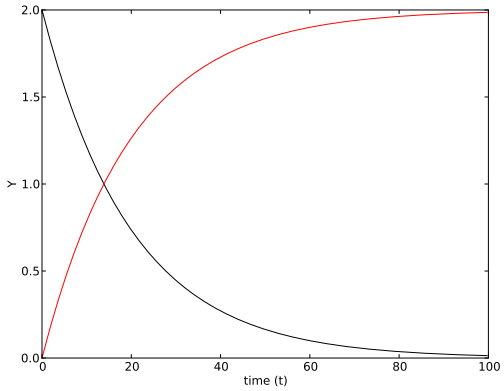

Figure 4: Decay from steady state (black) or accumulation (red) of a protein Y under the logic approximation described above.

from steady state or build up from zero) by setting $Y(t) = Y_{st}/2$ we fine that following formula for response time:

$$T_{1/2} = \frac{\log(2)}{\alpha}$$

Interestingly *the response time is only a function of the degradation/dilution rate*, not the rate of production! This leads us to an important rule of thumb for signaling and regulatory network – proteins with key roles in signal transduction (as well as modifications like phosphorylation) must have high turnover rates if their concentration need to be adjusted quickly.

5 Feed Forward Loops

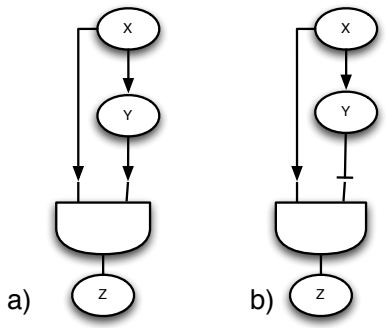

Figure 5: a) A coherent feed forward loop; b) an incoherent feed forward loop.

We're now going to use some of these tools to look at a class of network motifs, called Feed Forward Loops (FFLs), found in signaling and regulatory networks. FFLs involve interactions between three components, with the basic topology illustrated in Fig. 5. Depending on the signs of the edges (whether activating or repressing) we can classify FFLs as 'coherent' or 'incoherent.' We'll take a look at an example of each class.

5.1 A Coherent FFL

The most common type of coherent FFL is illustrated in Fig. 5a. In this system X is an activator of Y and both X and Y regulate the production of Z with AND logic (i.e. both X and Y must be above particular thresholds in order to trigger the production of Z). Using our logic approximation framework we will model this network as follows.

- For Y:

$$Y = f(X) = \beta_y \, \Theta(X > K_{xy})$$

$$\frac{dY}{dt} = \beta_y \, \Theta(X > K_{xy}) - \alpha_y Y$$

- For Z:

$$Z = g(X, Y) = \beta_z \, \Theta(X > K_{xz}) \Theta(Y > K_{yz})$$

$$\frac{dZ}{dt} = \beta_z \, \Theta(X > K_{xz}) \Theta(Y > K_{yz}) - \alpha_z Z$$

As before we can solve for Y as a function of time and calculate what it's steady state value will be:

$$Y(t) = Y_{st}(1 - e^{-\alpha_y t})$$

and

$$Y_{st} = \beta_y / \alpha_y$$

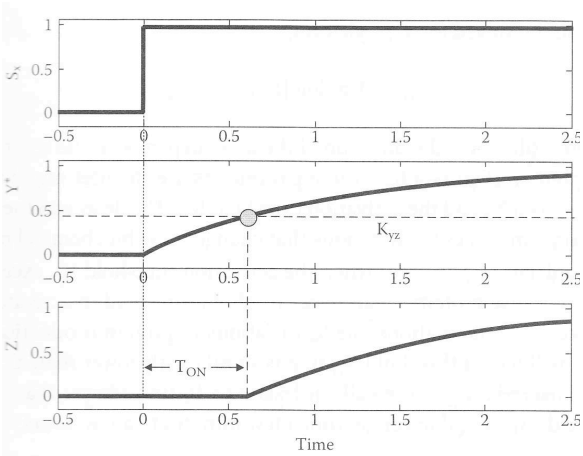

Figure 6: Delay time, T_{on}, associated with a coherent FFL, figure from [1].

How about Z? Since Z is governed by an AND function it needs both X and Y to be above their respective thresholds, K_{xz} and K_{yz}. For the sake of simplicity let's assume that both Y and Z have the same threshold with respect to X, i.e. $K_{xy} = K_{xz}$. This allows us just to consider how long it

takes for Y to reach the threshold value K_{yz}. Given this we can calculate the delay before Z turns on, T_{on} as follows.

$$Y(T_{\text{on}}) = Y_{st}(1 - e^{-\alpha_y T_{\text{on}}}) = K_{yz}$$

and solving for T_{on} we find:

$$T_{\text{on}} = 1/\alpha_y \log[1/(1 - K_{yz}/Y_{st})]$$

Thus we see that the delay before Z turns on is a function of the degradation rate of Y and the ratio between Y_{st} and K_{yz}. This delay in the production on Z is illustrated in Fig. 6.

As discussed in the article by Shen-Orr et al. [7] a feed forward loop of the type we've just discussed can act as a type of filter – a sign-sensitive delay that keeps Z from firing in response to transient noisy signals from X, but shuts down Z immediately once the signal from X is removed. In Fig. 7 we illustrate the dynamics of Y and Z to a short and long input signal X.

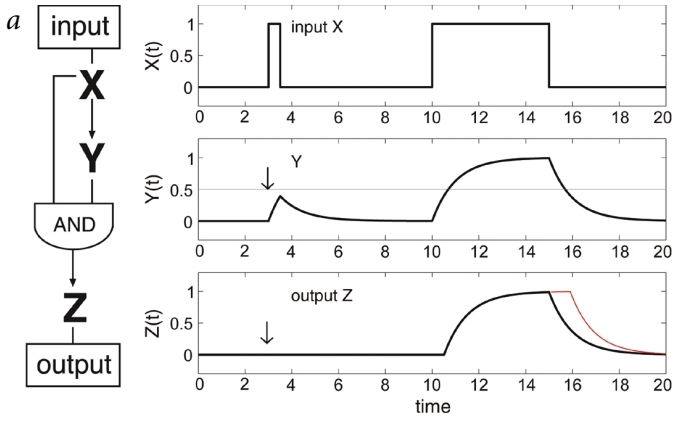

Figure 7: Dynamics of a coherent FFL, figure from [7].

5.2 An Incoherent FFL

Consider the FFL illustrated in Fig. 5b. In this incoherent FFL, the logic function that regulates Z is 'X and NOT Y'. That is Z turns on once X is above a given threshold, but only stays on fully as long as Y is below another threshold. Again for simplicity we assume $K_{xy} = K_{yz}$. As before, the dynamics of Y are described by:

$$\frac{dY}{dt} = \beta_y \, \Theta(X > K_{xy}) - \alpha_y Y$$

and

$$Y(t) = Y_{st}(1 - e^{-\alpha_y t})$$

To describe Z we consider two phases - 1) while $Y < K_{yz}$ and 2) while $Y > K_{yz}$. For the first phase:

$$\frac{dZ}{dt} = \beta_z \, \Theta(X > K_{xz}) - \alpha_z Z$$

and

$$Z(t) = Z_m(1 - e^{-\alpha_z t})$$

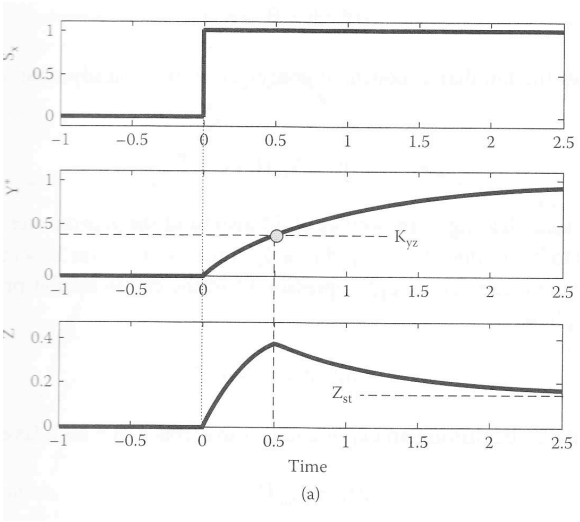

Figure 8: The dynamics of an incoherent FFL illustrated in Fig. 5b. Figure from [2].

As we did in the case of the coherent FFL, we can calculate the time until Y reaches the treshold K_{yz}. We'll call this T_{rep} and it is the same formula we found for T_{on} previously.

$$T_{\text{rep}} = \frac{1}{\alpha_y \log[\frac{1}{1-K_{yz}/Y_{st}}]}$$

After a delay, T_{rep}, Y starts to repress the transcription of Z and Z decays to a new lower steady state, $Z_{st} = \beta_z^{'}$. The value of $\beta_z^{'}$ depends on how leaky the repression of Z is by Y. The dynamics of Z in Phase 2 is given by:

$$Z(t) = Z_{st} + (Z_0 - Z_{st})e^{-\alpha_z(t-T_{\text{rep}})}$$

where

$$Z_0 = Z_m(1 - e^{-\alpha_z T_{\text{rep}}})$$

The dynamics of X, Y, and Z for a FFL of this type are illustrated in Fig. 8. Note that the stimulus amount of Z in the system initially increases, but then decreases to a lower steady even though the initial stimulus persists. This system thus generates pulse-like dynamics to a persistent signal. How pulse-like the signal is depends on the ratio of β_z to $\beta_z^{'}$. We define the repression factor, F, as follows:

$$F = \frac{\beta_z}{\beta_z^{'}} = \frac{Z_m}{Z_{st}}$$

Fig. 9 illustrates how varying the repression factor F affects the dynamics of Z.

5.3 Multiple FFLs

Multiple FFLs can be link together to generate very complicated dynamics, such as cascades of gene activity such as those often observed in development. One example, involving both coherent and incoherent FFLs is illustrated in Fig. 10.

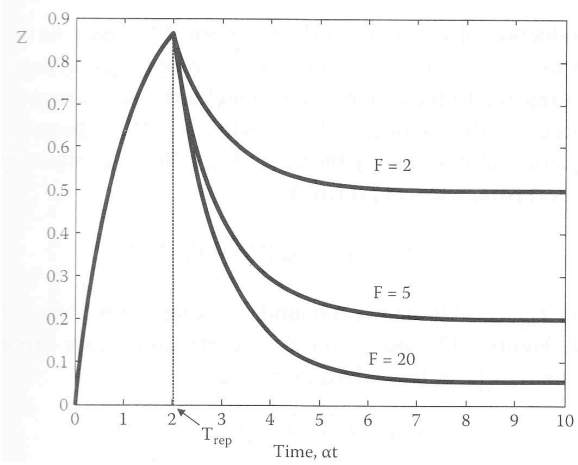

Figure 9: The effect of varying the repression factor, F, on the dynamics of an incoherent FFL. Figure from [1].

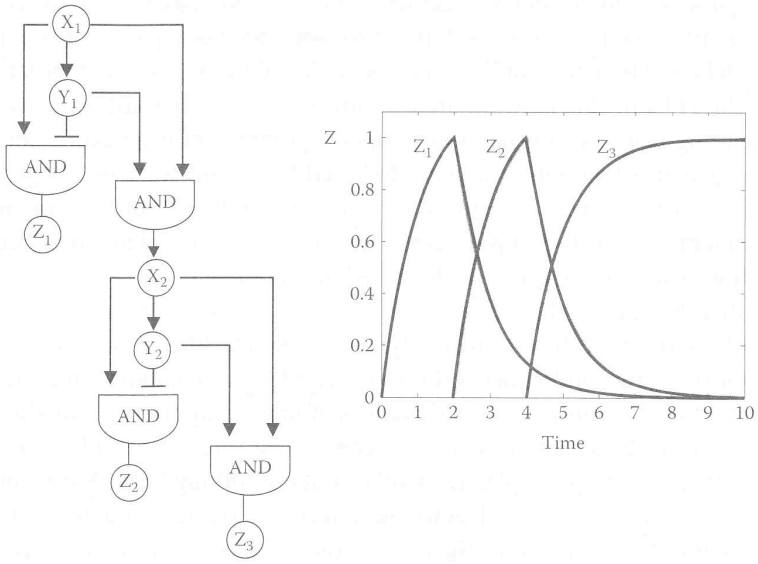

Figure 10: A complex network involving multiple FFLs, both coherent and incoherent, chained together. Figure from [1].