BE 159: Signal Transduction and Mechanics in Morphogenesis

Justin Bois

Caltech

Winter, 2019

This document was prepared at Caltech with financial support from the Donna and Benjamin M. Rosen Bioengineering Center.

2 Biochemical kinetics in signaling

When we look at a picture of how signaling works in a cell, as in Fig. 1, we see that a variety of processes occur along the signaling pathway. There is ligand binding to receptors that are embedded in a two-dimensional surface, the cell membrane. There is the transport, either passive by diffusion or active by motor proteins, of signaling molecules or transcription factors through the cytoplasm. Then, the transcription factor needs to get into the nucleus via nuclear pore complexes. From there, it needs to find the appropriate promoter to bind on the genome, an interesting transport problem by itself. There are plenty of interactions with the machinery involved in transcription, post transcriptional modifications, and eventual export from the nucleus. There are lots and lots of kinetic processes! We might throw our hands up in the air and scream that we do not know how to model *all* of that.

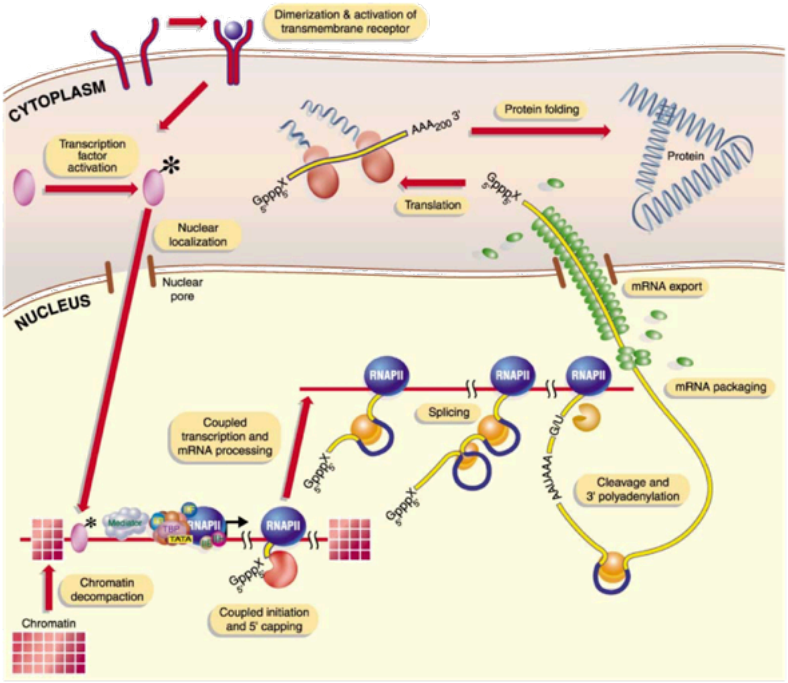

Figure 1: Schematic of a generic signaling pathway. Taken from Orphanides and Reinberg, *Cell*, **108**, 429–451, 2002.

So how do we overcome this modeling paralysis and proceed to develop physical description of these processes? There are a few main ideas we can consider to deal with this issue.

1. **Separation of time scales.** Some of the processes that happen along a signaling pathway are very fast compared to others. So, if we are interested in the dynamics of the entire pathway, say in terms of the more global response of a

cell to a variation in ligand concentration, we can ignore the fast processes, or at least assume that fast dynamics reach equilibrium rapidly.

2. **Assumption of Poisson processes.** In this context, a Poisson process may be thought of as a series of well-defined, separate events that occur randomly, without memory of what has occurred before. This is often the case for things like molecular collisions. If we model the events along a signaling pathway as Poisson processes, we can at least write down equations to describe the dynamics.

3. **Consideration only of average properties.** Instead of keeping track of what each molecule in the cell is doing, we can instead only consider how the *concentrations* of molecular species.

In what follows, we will put these approximations to use to arrive a **mass action kinetics** to describe the dynamics of molecules involved in cell signaling, and indeed in many other cellular processes. These ideas are central to the Goentoro and Kirschner paper, and will come into play throughout the rest of the course.

2.1 Thinking probabilistically: Master equations

Let us define very broadly a state s of a system to include all molecular species. Whenever there is a change of state, say from s' to s, there is a unit change in molecular species. For example, two proteins molecules that are bound to each other can separate, and this would lead to a state change. You can imagine that the state space available to all molecules in a cell is enormous. Nonetheless, let us move forward to write down a **master equation** to describe the dynamics of the *probability* that the molecular species of a cell are in state s at time t, which we denote as $P(s,t)$.

Generally, a master equation is a loss-gain equation for probabilities of states governed by a Markov process.[1] Specifically,

$$\frac{\mathrm{d}P(s,t)}{\mathrm{d}t} = \sum_{s'} \left[W(s \mid s')P(s',t) - W(s' \mid s)P(s,t) \right]. \tag{2.1}$$

Here, $W(s \mid s')$ is the transition probability per unit time of going from s' to s. Note that there is an ODE for *each* of the many many many states s.

The master equation makes sense by inspection and appears simple. The nuance lies in the definition of the transition rates, $W(s \mid s')$. There is also the computational difficulty that state space is enormous. In general, solving the master equation is difficult and is usually intractable analytically.

[1] A good reference for studying master equations is *Stochastic Processes in Physics and Chemistry* by N. G. van Kampen.

To make some more progress, let's restrict what we call a "state." We will define a state to be a set of copy numbers of molecular species. At this point, it helps to be less abstract and think of a concrete example. Consider the case where two signaling molecules, a and b may bind and unbind to each other, and these are the only molecules we are considering. There are then three molecular species a, b, and ab. We then define a state by the copy numbers of these respective species.

$$s \to \mathbf{n} \equiv (n_{\mathrm{a}}, n_{\mathrm{b}}, n_{\mathrm{ab}}). \tag{2.2}$$

Then, we can re-write the master equation as

$$\frac{\mathrm{d}P(\mathbf{n},t)}{\mathrm{d}t} = \sum_{\mathbf{n}'} \left[W(\mathbf{n} \mid \mathbf{n}')P(\mathbf{n}',t) - W(\mathbf{n}' \mid \mathbf{n})P(\mathbf{n},t) \right]. \tag{2.3}$$

2.2 Assigning the transition rates

Since the events that change the state are binding of an a and a b molecule or the dissociation of an ab complex, we know that very many of the transition rates, $W(\mathbf{n} \mid \mathbf{n}')$ are zero. Specifically, $W(n_{\mathrm{a}}, n_{\mathrm{b}}, n_{\mathrm{ab}} \mid n'_{\mathrm{a}}, n'_{\mathrm{b}}, n'_{\mathrm{ab}}) = 0$ for all cases except:

$$\begin{aligned} n_{\mathrm{ab}} &= n'_{\mathrm{ab}} - 1,\ n_{\mathrm{a}} = n'_{\mathrm{a}} + 1,\ n_{\mathrm{b}} = n'_{\mathrm{b}} + 1 \quad \text{(dissociation)} \\ \text{or } n_{\mathrm{ab}} &= n'_{\mathrm{ab}} + 1,\ n_{\mathrm{a}} = n'_{\mathrm{a}} - 1,\ n_{\mathrm{b}} = n'_{\mathrm{b}} - 1 \quad \text{(binding).} \end{aligned} \tag{2.4}$$

What value should we assign $W(\mathbf{n} \mid \mathbf{n}')$ for these two cases? Consider first dissociation. The probability per unit time of getting a dissociation event should be dependent on the number of ab complexes there are. If there are no ab complexes, the probability of getting a dissociation is zero. If we further assume that the complexes are all independent of each other, valid in the *dilute limit*, then the probability of observing a transition should be proportional to the number of ab complexes. Finally, since we model all processes as Poisson processes, there is no memory, so therefore no temporal dependence. So, we have

$$W_{\mathrm{dissoc}} = k_{-1}\, n_{\mathrm{ab}}, \tag{2.5}$$

where k_{-1} is a constant. (The subscript -1 denotes dissociation; we will use the subscript 1 for binding.)

Now, let's consider binding. Again, the transition rate for binding should be independent of time because we are dealing with Poisson processes. In order for a binding event to happen, two molecules need to collide. The probability of collision should scale with the copy number of each species, a and b. It should also scale inversely with the available volume (or surface area if we are talking about binding

events on a membrane). In other words the bigger the volume, the less likely it is to observe a collision.[2] So, we have

$$W_{\text{binding}} = k_1 n_{\text{a}} n_{\text{b}} / V, \tag{2.6}$$

where V is the volume of the cell or system of interest.

Now that we know our transition rates, we can rewrite the master equation.

$$\begin{aligned}\frac{\mathrm{d}P(n_{\text{a}}, n_{\text{b}}, n_{\text{ab}}, t)}{\mathrm{d}t} = &\frac{k_1}{V}(n_{\text{a}}+1)(n_{\text{b}}+1)P(n_{\text{a}}+1, n_{\text{b}}+1, n_{\text{ab}}-1, t) \\ &+ k_{-1}(n_{\text{ab}}+1)P(n_{\text{a}}-1, n_{\text{b}}-1, n_{\text{ab}}+1, t) \\ &- \left(\frac{k_1}{V} n_{\text{a}} n_{\text{b}} + k_{-1} n_{\text{ab}}\right) P(n_{\text{a}}, n_{\text{b}}, n_{\text{ab}}, t),\end{aligned} \tag{2.7}$$

where it is understood that $P(n_{\text{a}}, n_{\text{b}}, n_{\text{ab}}, t)$ is zero if any of n_{a}, n_{b}, or n_{ab} are less than zero.

2.3 Dynamics of averages

We now have a workable master equation, but there still many, many equations. If we instead consider instead how the *average* number of each species changes over time, we can greatly reduce the number of equations. In doing this, we are throwing out much of the information contained in the probability distribution $P(\mathbf{n}, t)$, considering only its first moment. With that caveat in mind, let's compute the first moment. Recall that the average number of a molecules is

$$\langle n_{\text{a}} \rangle(t) = \sum_{n_{\text{a}}=0}^{\infty} n_{\text{a}} \, P(n_{\text{a}}, t), \tag{2.8}$$

with

$$P(n_{\text{a}}, t) = \sum_{n_{\text{b}}=0}^{\infty} \sum_{n_{\text{ab}}=0}^{\infty} P(n_{\text{a}}, n_{\text{b}}, n_{\text{ab}}, t), \tag{2.9}$$

thereby giving

$$\langle n_{\text{a}} \rangle(t) = \sum_{n_{\text{a}}=0}^{\infty} \sum_{n_{\text{b}}=0}^{\infty} \sum_{n_{\text{ab}}=0}^{\infty} n_{\text{a}} \, P(n_{\text{a}}, n_{\text{b}}, n_{\text{ab}}, t). \tag{2.10}$$

[2] That the transition rate is proportional to V^{-1} and not, say, V^{-2} requires some careful analysis we will not go into here.

So, we will multiply both sides of equation (2.7) by n_a and apply the triple sum

$$\sum_{n_\mathrm{a}=0}^{\infty}\sum_{n_\mathrm{b}=0}^{\infty}\sum_{n_\mathrm{ab}=0}^{\infty} \tag{2.11}$$

to the resulting equation. Evaluation of the left hand side of equation (2.7) is trivial.

$$\sum_{n_\mathrm{a}=0}^{\infty}\sum_{n_\mathrm{b}=0}^{\infty}\sum_{n_\mathrm{ab}=0}^{\infty} n_\mathrm{a}\,\frac{\mathrm{d}P(n_\mathrm{a},n_\mathrm{b},n_\mathrm{ab},t)}{\mathrm{d}t} = \frac{\mathrm{d}\langle n_\mathrm{a}\rangle}{\mathrm{d}t}. \tag{2.12}$$

The last two terms on the right hand side are the easiest to evaluate.

$$\begin{aligned}-\sum_{n_\mathrm{a}=0}^{\infty}\sum_{n_\mathrm{b}=0}^{\infty}\sum_{n_\mathrm{ab}=0}^{\infty}\frac{k_1}{V}\,n_\mathrm{a}^2 n_\mathrm{b}\,P(n_\mathrm{a},n_\mathrm{b},n_\mathrm{ab},t) &= -\frac{k_1}{V}\sum_{n_\mathrm{a}=0}^{\infty}\sum_{n_\mathrm{b}=0}^{\infty} n_\mathrm{a}^2 n_\mathrm{b}\,P(n_\mathrm{a},n_\mathrm{b},t)\\ &= -\frac{k_1}{V}\,\langle n_\mathrm{a}^2 n_\mathrm{b}\rangle. \end{aligned} \tag{2.13}$$

Similarly,

$$-\sum_{n_\mathrm{a}=0}^{\infty}\sum_{n_\mathrm{b}=0}^{\infty}\sum_{n_\mathrm{ab}=0}^{\infty} k_{-1} n_\mathrm{a} n_\mathrm{ab}\,P(n_\mathrm{a},n_\mathrm{b},n_\mathrm{ab},t) = -k_{-1}\langle n_\mathrm{a} n_\mathrm{ab}\rangle. \tag{2.14}$$

In these expressions, for notational convenience, we have not written the explicit time dependence of the averages. Now, we will work on the first sum on the right hand side.

$$\begin{aligned}&\sum_{n_\mathrm{a}=0}^{\infty}\sum_{n_\mathrm{b}=0}^{\infty}\sum_{n_\mathrm{ab}=0}^{\infty}\frac{k_1}{V}\,n_\mathrm{a}(n_\mathrm{a}+1)(n_\mathrm{b}+1)P(n_\mathrm{a}+1,n_\mathrm{b}+1,n_\mathrm{ab}-1,t)\\ &\quad= \frac{k_1}{V}\sum_{n_\mathrm{a}=0}^{\infty}\sum_{n_\mathrm{b}=0}^{\infty}\sum_{n_\mathrm{ab}=0}^{\infty} n_\mathrm{a}(n_\mathrm{a}+1)(n_\mathrm{b}+1)P(n_\mathrm{a}+1,n_\mathrm{b}+1,n_\mathrm{ab},t)\\ &\quad= \frac{k_1}{V}\sum_{n_\mathrm{a}=0}^{\infty}\sum_{n_\mathrm{b}=0}^{\infty} n_\mathrm{a}(n_\mathrm{a}+1)(n_\mathrm{b}+1)P(n_\mathrm{a}+1,n_\mathrm{b}+1,t)\\ &\quad= \frac{k_1}{V}\sum_{n_\mathrm{a}=0}^{\infty}\sum_{n_\mathrm{b}=0}^{\infty}(n_\mathrm{a}-1)n_\mathrm{a} n_\mathrm{b}\,P(n_\mathrm{a},n_\mathrm{b},t)\\ &\quad= \frac{k_1}{V}\left(\langle n_\mathrm{a}^2 n_\mathrm{b}\rangle - \langle n_\mathrm{a} n_\mathrm{b}\rangle\right). \end{aligned} \tag{2.15}$$

And finally, the second sum on the right hand side.

$$\sum_{n_\mathrm{a}=0}^{\infty}\sum_{n_\mathrm{b}=0}^{\infty}\sum_{n_\mathrm{ab}=0}^{\infty} k_{-1} n_\mathrm{a}(n_\mathrm{ab}+1)P(n_\mathrm{a}-1,n_\mathrm{b}-1,n_\mathrm{ab}+1,t)$$

$$= k_{-1} \sum_{n_a=0}^{\infty} \sum_{n_b=0}^{\infty} \sum_{n_{ab}=0}^{\infty} n_a(n_{ab}+1)P(n_a-1, n_b, n_{ab}+1, t)$$

$$= k_{-1} \sum_{n_a=0}^{\infty} \sum_{n_{ab}=0}^{\infty} n_a(n_{ab}+1)P(n_a-1, n_{ab}+1, t)$$

$$= k_{-1} \sum_{n_a=0}^{\infty} \sum_{n_{ab}=0}^{\infty} (n_a+1)n_{ab}P(n_a, n_{ab}, t)$$

$$= k_{-1} \left(\langle n_a n_{ab} \rangle + \langle n_{ab} \rangle \right). \tag{2.16}$$

Now that we have computed all of the sums, let's put it all together.

$$\frac{\mathrm{d}\langle n_a \rangle}{\mathrm{d}t} = \frac{k_1}{V} \left(\langle n_a^2 n_b \rangle - \langle n_a n_b \rangle \right) + k_{-1} \left(\langle n_a n_{ab} \rangle + \langle n_{ab} \rangle \right) - \frac{k_1}{V} \langle n_a^2 n_b \rangle - k_{-1} \langle n_a n_{ab} \rangle$$

$$= -k_1 \frac{\langle n_a n_b \rangle}{V} + k_{-1} \langle n_{ab} \rangle. \tag{2.17}$$

If we assume that the particle counts of species a and species b are independent, then $\langle n_a n_b \rangle = \langle n_a \rangle \langle n_b \rangle$. Then, we have

$$\frac{\mathrm{d}\langle n_a \rangle}{\mathrm{d}t} = -k_1 \frac{\langle n_a \rangle \langle n_b \rangle}{V} + k_{-1} \langle n_{ab} \rangle. \tag{2.18}$$

The thermodynamic *concentration* of species i is $c_i = \langle n_i \rangle / V$. So, if we divide both sides of the above equation by V, we get

$$\frac{\mathrm{d}c_a}{\mathrm{d}t} = -k_1 c_a c_b + k_{-1} c_{ab}. \tag{2.19}$$

If we do the same averaging technique with n_b and n_{ab}, we get

$$\frac{\mathrm{d}c_b}{\mathrm{d}t} = -k_1 c_a c_b + k_{-1} c_{ab}, \tag{2.20}$$

$$\frac{\mathrm{d}c_{ab}}{\mathrm{d}t} = k_1 c_a c_b - k_{-1} c_{ab}. \tag{2.21}$$

We now have three equations in terms of concentrations that we derived from the master equation.

2.4 The law of mass action

Chemical rate equations like those we just derived, in which the rate of a chemical reaction is proportional to the products of the concentrations of the participating

molecular species follow the **law of mass action**, often referred to as **mass action kinetics**.

It is important to recall all of the assumptions we made to get here.

1. Events that change state are Poisson processes. Implicit in this assumption is that the binding or dissociation (or any other event) happens essentially instantaneously with well-defined pauses between them. This is one instance of where a separation of time scales is important.
2. All molecular species are independent of each other; i.e., we are in the dilute limit.
3. In our last step, by taking $\langle n_\mathrm{a} n_\mathrm{b} \rangle \approx \langle n_\mathrm{a} \rangle \langle n_\mathrm{b} \rangle$, we tacitly assumed that $P(n_\mathrm{a}, n_\mathrm{b}, t) \approx P(n_\mathrm{a}, t) P(n_\mathrm{b}, t)$, i.e., that n_a and n_b are independent of each other, or that they have small covariance. Note that this is not always necessarily the case, especially at small copy number.

In addition to these assumptions, we are willfully throwing out all information about the probability distributions of the states we're modeling, except for the first moment (the mean).

Going forward in the course, we will use and abuse the law of mass action extensively. Especially when the copy number of molecules are small, this could lead to trouble. We might miss the important of noise (since we are neglecting fluctuations), or some of the underlying assumptions might not be valid. Nonetheless, the law of mass action is one of those approximate theories that is "unreasonably effective" in the sense that we are surprised at how well it tends to work in matching experimental observation.

2.5 Sigmoidal rate dependence?

If you have studied systems biology, you often find expressions for rates that are sigmoidal in shape, such as

$$\frac{\mathrm{d}c}{\mathrm{d}t} = \frac{k_1}{k_2 + c^2}, \tag{2.22}$$

a famous Hill function. This does not look like mass action at face. Where does this come from?

You may be familiar with Michaelis-Menten kinetics for enzymatic activity. In this scheme, a substrate s reacts with an enzyme e to form an intermediate which then forms a product p according to the mechanism

$$e + s \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} se \xrightarrow{q_1} e + p. \tag{2.23}$$

Now, let's consider the case where the complex se might bind another substrate in the reaction

$$se + s \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} s_2e \xrightarrow{q_2} se + p. \tag{2.24}$$

There are now six reactions in total and a total of five species. We could write master equations, perform averages, and then get the mass action expressions, but we will just directly write the mass action ODEs directly.

$$\frac{\mathrm{d}c_s}{\mathrm{d}t} = -k_1 c_e c_s + k_{-1} c_{se} - k_2 c_{se} c_s + k_{-2} c_{s_2e}, \tag{2.25}$$

$$\frac{\mathrm{d}c_e}{\mathrm{d}t} = -k_1 c_e c_s + (k_{-1} + q_1) c_{se}, \tag{2.26}$$

$$\frac{\mathrm{d}c_{se}}{\mathrm{d}t} = k_1 c_e c_s - (k_{-1} + q_1) c_{se} - k_2 c_{se} c_s + (k_{-2} + q_2) c_{s_2e}, \tag{2.27}$$

$$\frac{\mathrm{d}c_{s_2e}}{\mathrm{d}t} = k_2 c_{se} c_s - (k_{-2} + q_2) c_{s_2e}, \tag{2.28}$$

$$\frac{\mathrm{d}c_p}{\mathrm{d}t} = q_1 c_{se} + q_2 c_{s_2e}. \tag{2.29}$$

We are primarily interested in the rate of consumption of substrate, so we seek a simple expression for $\dot{c}_s$ in terms of the total enzyme and substrate concentration. Toward this end, we make a **pseudo steady state approximation** that $\dot{c}_{se} - \dot{c}_{s_2e} = 0$. This means that the concentrations of the intermediates do not change appreciably on the time scale of product formation. Again, this is an instance where separation of time scales allows us to make useful approximations to simplify the mathematics. As a result, we have

$$k_1 c_e c_s = (k_{-1} + q_1) c_{se}, \tag{2.30}$$

$$k_2 c_{se} c_s = (k_{-2} + q_2) c_{s_2e}. \tag{2.31}$$

We can rewrite the first of these equation as

$$c_e c_s = K_{M,1} c_{se}, \tag{2.32}$$

where

$$K_{M,1} \equiv \frac{k_{-1} + q_1}{k_1} \tag{2.33}$$

is the Michaelis constant for the first reaction. We also get

$$c_e c_s^2 = K_{M,1} K_{M,2} c_{s_2e}, \tag{2.34}$$

with

$$K_{M,2} \equiv \frac{k_{-2} + q_2}{k_2}. \tag{2.35}$$

Since the total amount of enzyme is conserved, we define the constant amount of enzyme

$$c_e^{\text{tot}} \equiv c_e + c_{se} + c_{s_2e}. \tag{2.36}$$

Using this relation along with equations (2.34) and (2.36) allows us to write an expression for c_e in terms of c_e^{tot} and c_s.

$$c_e = c_e^{\text{tot}} \left(1 + \frac{c_s}{K_{M,1}} + \frac{c_s^2}{K_{M,1}K_{M,2}}\right)^{-1}. \tag{2.37}$$

Substituting this expression, along with equations (2.34) and (2.36) into the expression for $\dot{c}_s$ (equation (2.25)) gives, after simplification

$$\frac{\mathrm{d}c_s}{\mathrm{d}t} = -\frac{c_e^{\text{tot}}\left(q_1 K_{M,2} c_s + q_2 c_s^2\right)}{K_{M,1}K_{M,2} + K_{M,2}c_s + c_s^2}. \tag{2.38}$$

This equation has a sigmoidal form, and it looks like a typical phenomenological Hill function (2.22) in certain limits. In particular, if $q_1 \approx 0$, that is if only the doubly-complexed substrate can produce product, the numerator becomes $q_2 c_e^{\text{tot}} c_s^2$. Further, if $K_{M,2} \ll c_s \ll K_{M,1}$, the denominator becomes $K_{M,1}K_{M,2} + c_s^2$. This means that once one substrate molecules is bound to an enzyme, the second binds much faster (its Michaelis constant is smaller). These limits are hallmarks of cooperativity. The result is

$$\frac{\mathrm{d}c_s}{\mathrm{d}t} = -\frac{q_2 c_e^{\text{tot}} c_s^2}{K_{M,1}K_{M,2} + c_s^2}. \tag{2.39}$$

which has the same form as the Hill equation (2.22) with Hill coefficient of 2.

The important lesson here is that many molecular mechanisms can give kinetics that relate to phenomenological Hill equations. But the Hill equation by itself says very little about the underlying mechanism. In my view, if you have a molecular mechanism in mind, it is best to derive the actual expressions you want to use. In fact, because it is not difficult to numerically solve a system of ODEs, you are often better off just directly solving the original mass action ODEs you write down without approximation. In the Goentoro and Kirschner paper, however, you will see that the big advantage of carefully doing analytical work, nondimensionalizing, and making reasonable approximations is that you can draw more general conclusions and sometimes expose structure to the system that might otherwise be difficult to see.

2.6 Sigmoidal rates of gene expression

If we are interested in the kinetics of actual gene products, not just intermediates along the signaling pathway, we need to model the gene expression levels. This is often down by throwing Hill functions around. However, I encourage you to not do this, but rather to think carefully about the structure of the promoter region and the transcription factors that bind it. If you take BE/APh 161, this will be covered in detail, and I omit it in this course.