Appendix A: The input functions of genes: Michaelis-Menten and Hill equations

A.1 Binding of a repressor to a promoter

This appendix provides a simplified introduction to basic models in biochemistry. We will begin with understanding the interaction of a repressor protein with DNA and with its inducer[1]. We will then turn to activator proteins. The repressor X binds to a specific DNA site, D, in a promoter. Thus, X and D bind to form a complex, [XD]. Transcription of the gene occurs only when the repressor is *not* bound, that is, when D is free. The DNA site can thus be either free, D, or bound, [DX], resulting in a **conservation equation**:

$$D + [XD] = D_T \qquad \text{(A.1.1)}$$

where D_T is the total concentration of the site. For example, a single DNA binding site per bacterial cell means that DT = 1/cell volume ~ 1/μm^3 ~ 1 nM. In eukaryotic cells, the volume of the nucleus is on the order of 10–100 μm^3.

The repressor X and its target D diffuse in the cell and occasionally collide to form the complex [XD]. This process can be described by mass-action kinetics: X and D collide and bind each other at a rate k_{on}. The rate of complex formation is thus proportional to the collision rate, given by the product of the concentrations of X and D:

$$\text{rate of complex formation} = k_{on}\, X\, D$$

The complex [XD] falls apart (dissociates) at a rate k_{off}. The rate of change of [XD] based on these collision and dissociation processes is described by

$$d[XD]/dt = k_{on}\, X\, D - k_{off}\, [XD] \quad \text{(A.1.2)}$$

The rate parameter for the collisions, k_{on}, describes how many collision events occur per second per protein at a given concentration of D, and thus has units of 1/time/concentration. It is useful to remember that k_{on} in biochemical reactions is often limited by the rate of collisions of a diffusing molecule hitting a protein-size target, and has a diffusion-limited value of about $k_{on} \sim 10^8 – 10^9\ M^{-1}\ sec^{-1}$, independent of the details of the reaction. For the case of a transcription factor and DNA, the diffusion limit is usually about ten times higher because of one-dimensional diffusion effects due to sliding of the transcription factor along the DNA (Berg, Winter and von Hippel, 1981)

The off-rate k_{off}, on the other hand, has units of 1/time and can vary over many orders of magnitude for different reactions, because koff is determined by the strength of the chemical bonds that bind X and D.

The kinetics of Equation A.1.2 approach a steady-state in which concentrations do not change with time, d[XD]/dt = 0. Solving Equation A.1.2 at steady-state, we find that the balance between the collision of X and D and the dissociation of [XD] leads to the chemical equilibrium equation:

---

[1]The theoretical treatment for the input function of simple gene regulation was initiated by Gad Yagil in the context of the lac system of *E. coli* ((Yagil and Yagil, 1971)).

$$K_d\;[XD]\;=\;XD \qquad \text{(A.1.3)}$$

where K_d is the **dissociation constant**,

$$K_d\;=\;k_{off}/k_{on}$$

The dissociation constant K_d has units of concentration. The larger the dissociation constant, the higher the rate of dissociation of the complex, that is, the weaker the binding of X and D.
Solving for the concentration of free DNA sites, D, using Equations. A.1.1 and A.1.3, we find $K_d\;(D_T - D) = X\;D$, which yields

$$\frac{D}{D_T} = \frac{1}{1+X/K_T} \quad \text{(A.1.4)}$$

For many repressors, [XD] complexes dissociate within less than 1 sec (that is, $k_{off} > 1\;sec^{-1}$). Therefore, we can average over times much longer than 1 sec and consider D/D_T as the probability that site D is free, averaged over many binding and unbinding events.
The probability that the site is free, D/D_T, is a decreasing function of the concentration of repressor X. When there is no repressor, X = 0, the site is always free, $D/D_T = 1$. The site has a 50% chance of being free, $D/D_T = 1/2$, when $X = K_d$.
When site D is free, RNA polymerase can bind the promoter and transcribe the gene. The rate of transcription (number of mRNAs per second) from a free site is given by the maximal transcription rate β . (Note that in the main text we used \beta to denote the rate of protein production. This rate is proportional to the transcription rate times the number of proteins translated per mRNA provided that there is a constant mRNA life-time and translation rate.) The maximal transcription rate depends on the DNA sequence and position of the RNA polymerase binding site in the promoter and other factors. It can be tuned by evolutionary selection, for example, by means of mutations that change the DNA sequence of the RNAp binding site. In different genes, \beta ranges over several orders of magnitude, \beta ~ 10^{-4} – 1 mRNA/sec. The rate of mRNA production, called the **promoter activity**, is b\beta times the probability that site D is free:

$$promoter\;activity\;=\frac{\beta}{1+\frac{X}{K_d}} \quad \text{(A.1.5)}$$

Figure A.1 shows the promoter activity as a function of X (here X is repressor in its active, DNA binding form, denoted X* in the main text). When X is equal to K_d, transcription is reduced by 50% from its maximal value. The value of X needed for 50% maximal repression is called the **repression coefficient**.

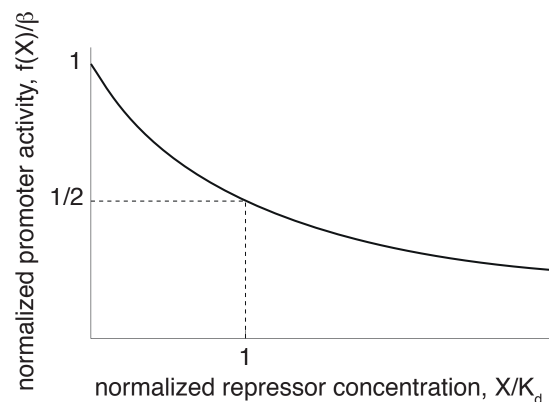


Figure A.1

For efficient repression, enough repressor is needed so that site D is almost always occupied with repressor. From Equation A.1.4, this occurs when repressor concentration greatly exceeds the

dissociation constant, such that X/Kd >> 1. This is the case for many repressors, including the lac repressor LacI.
So far we discussed how the repressor binds the promoter and inhibits transcription. To turn the gene system ON, a signal must cause X to unbind from the DNA. We will treat the simplest case, in which a small molecule (an inducer) is the signal. The inducer directly binds to protein X and causes it to assume a molecular conformation where it does not bind D with high affinity. Typically, signals can reduce the affinity of X to its DNA sites by a factor of 10 to 100. Thus, the inducer frees the promoter and allows transcription of the gene. We now consider the binding of inducer to X.

A.2 Binding of an inducer to a repressor protein: the Michaelis-Menten equation
The repressor protein X is designed to bind a small-molecule inducer S_x, which can be considered as its input signal. The two can collide to form a bound complex, $[XS_x]$. The repressor is therefore found[2] in either free form, X, or bound form $[XS_x]$, and a conservation law states that the two forms sum up to the total concentration of repressor protein X_T:

$$X_T = X + [XS_x] \qquad (A.2.1)$$

X and S_x collide to form the complex $[XS_x]$ at a rate $k_{on,}$ and the complex [XSx] falls apart (dissociates) at a rate k_{off}. Thus, the mass-action kinetic equation is:2

$$d\,[XS_x]/dt = k_{on}\, X\, S_x - k_{off}\,[XS_x] \qquad (A.2.2)$$

At steady state, $d[XS_x]/dt = 0$, and we find the chemical equilibrium relation:

$$K_x\,[XS_x] = X\, S_x \qquad (A.2.3)$$

where $K_x = k_{off}/k_{on}$ is the dissociation constant. For the lac repressor, K_x ~ 1 µM ~ 1000 inducer (IPTG) molecules/cell. Using the diffusion-limited value for k_{on} ~ 10^9/M/sec, we find the lifetime of the complex is $1/k_{off}$ ~ 1 msec.
Using the conservation of total repressor X (Equation A.2.1), we arrive at a useful equation that recurs throughout biology (this equation is known as the Michaelis–Menten equation in the context of enzyme kinetics; we use the same name in the present context of inducer binding):

$$[XS_x] = \frac{X_T S_X}{S_X + K_X} \quad \text{Michaelis–Menten equation} \qquad (A.2.4)$$

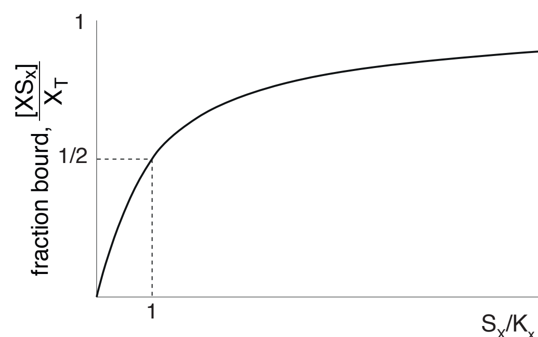


Figure A.2

The Michaelis–Menten term (Figure A.2) has three notable features:

1. It rises about linearly with S_x when S_x is low ($S_x \ll K_x$).

[2] Usually the number of Sx molecules is much larger than the number of X molecules, and so we need not worry about conservation of Sx, Sx,total = Sx + [XSx]. For example, in the lac system, the number of LacI repressors, each made of a tetramer of LacI proteins, is XT ~ 10 units/cell, which is negligible relative to Sx, which is at least 1000/cell for a detectable response.

2. It reaches saturation (stops rising) at high S_x.
3. It is equal to 0.5 when $S_x = K_x$.

The dissociation constant thus provides the scale for detection of signal: S_x concentrations far below K_x are not detected; concentrations far above K_x saturate the repressor at its maximal binding.
Recall that in cases like LacI, only X unbound to Sx, is active, X*, in the sense that it can bind the promoter D to block transcription. Because free X is active, we denote it by X*. Active repressor, $X^* = X_T - [XS_x]$, decreases with increasing inducer levels:

$$X^* = \frac{X_T}{1+S_X/K_X} \quad \text{concentration of X not bound to } S_X \quad \text{(A.2.5)}$$

A.3 Cooperativity OF Inducer binding And THE Hill Equation

Before returning to the input function, we comment on a more realistic description of inducer binding. Most transcription factors are composed of several repeated protein sub-units, for example, dimers or tetramers. Each of the protein subunits can bind inducer molecules. Often, full activity is only reached when multiple subunits bind the inducer. A useful phenomenological equation for this process can be derived by assuming that n molecules of Sx can bind X.
To describe the binding process, we assume a simple case: the protein (multimer) X can either be bound to n molecules of Sx, described by the complex [nSx X], or unbound, denoted Xo (thus, in this simple treatment, intermediate
states where fewer than n molecules are bound are neglected). The total concentration of bound and unbound X is X_T, and the conservation law is thus

$$[nS_x\, X] \;+\; X_o \;=\; X_T \quad \text{(A.3.1)}$$

The complex [nSx X] is formed by collisions of X with n molecules of Sx. Thus, the rate of the molecular collisions needed to form the complex is given by the product of the concentration of free X, Xo, and the concentration of Sx to the power n (the probability of finding n copies of Sx at the same place at the same time):

$$collision\, rate \;=\; k_{on}\, X_o\, S_x^n \qquad \text{(A.3.2)}$$

where the parameter k_{on} describes the on-rate of complex formation. The complex [nSx X] dissociates with rate k_{off}:

$$dissociation\, rate \;=\; k_{off}\, [nS_x\, X] \quad \text{(A.3.3)}$$

The parameter k_{off} corresponds to the strength of the chemical bonds between Sx and its binding sites on X. The total rate of change of the concentration of the complex is thus the difference between the rate of collisions and dissociations:

$$d[nS_x\, X]/dt \;=\; k_{on}\, X_o\, S_x n - \, k_{off}\, [nS_x\, X] \quad \text{(A.3.4)}$$

This equation reaches equilibrium within milliseconds for typical inducers. Hence, we can make a steady-state approximation, in which d[nSx X]/dt = 0, to find that dissociations balance collisions:

$$k_{off}\, [nS_x\, X] \;=\; k_{on}\, X_o\, S_x^n \quad \text{(A.3.5)}$$

We can now use the conservation equation (Equation A.3.1) to replace Xo with XT –

[nSx X], to find

$$(k_{off}/k_{on})\,[nS_x\,X] \;=\; (X_T - [nS_x\,X])S_x^n \quad \text{(A.3.6)}$$

Finally, we can solve for the fraction of bound X, to find a binding equation known as the Hill equation:

$$\frac{[nS_X\,X]}{X_T} = \frac{S_X^n}{K_X^n + S_X^n} \quad \text{Hill equation} \qquad \text{(A.3.7)}$$

where we have defined the constant Kx such that

$$Kxn \;= k_{off}/k_{on} \qquad \text{(A.3.8)}$$

Equation A.3.7 can be considered the probability that the site is bound, averaged over many binding and unbinding events of Sx.

The parameter n is known as the Hill coefficient. When n = 1, we obtain the Michaelis–Menten term (Equation A.2.4). As shown in Figure A.3, both the Michaelis–Menten and Hill equations reach half-maximal binding when Sx = Kx.
The steepness of the Hill curve is greater the larger the Hill coefficient n (Figure A.3). In the lac system, n = 2 with the inducer IPTG (Yagil and Yagil, 1971).
Reactions described by Hill coefficients n > 1 are often termed cooperative reactions.
The concentration of unbound repressor X is given by:

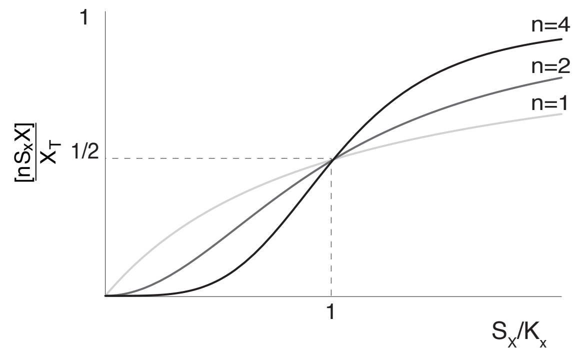


Figure A.3

$$\frac{X^*}{X_T} = \frac{1}{1+\left(\frac{S_X}{K_X}\right)^n} \quad \text{(A.3.9)}$$

A.4 The Monod-Changeux-Wyman model

We note that a more rigorous and elegant analysis of cooperative binding based on symmetry principles is due to Monod, Changeux, and Wymann, in a paper well worth reading ((Monod, Wyman and Changeux, 1965)), usually also described in biochemistry textbooks. In this model X switches to an active state X* and back. The signal Sx binds X with dissociation constant Kx, and binds X* with a lower dissociation constant Kx*. Up to n molecules of Sx can bind to X. The two states, X and X* spontaneously switch such that in the absence of Sx, X is found at a probability larger by L than X*. The result is:

$$\frac{X^*}{X_T} = \frac{\left(1+\frac{S_X}{K_X^*}\right)^n}{L\left(1+\frac{S_X}{K_X}\right)^n + \left(1+\frac{S_X}{K_X^*}\right)^n}$$

Interesting extensions to this model make analogies to Ising models in physics (Duke et al., 2001). One difference between the rigorous models and the Hill curve is that binding at low

concentrations of Sx is linear in Sx rather than a power law with coefficient n, as in Equation A.3.7. This linearity is due to the binding of a single site on X, rather than all sites at once.

A.5 The input function of a gene regulated by a repressor

We can now combine the binding of inducer to the repressor (Equation A.2.5) and the binding of the repressor to the DNA (Equation A.1.4) to obtain the input function of the gene. The input function in this case describes the rate of transcription as a function of the input inducer concentration Sx:

$$f(S_X) = \frac{\beta}{1+X^*/\mathrm{K_d}} = \frac{\beta}{1+\frac{X_T}{k_d}\frac{1}{1+\left(\frac{S_X}{K_X}\right)^n}} \qquad \text{(A.5.1)}$$

Figure A.4 shows how the transcription rate of a gene repressed by X increases with increasing inducer concentration Sx. Note, when no inducer is present, there is a **leakage transcription rate**, f(Sx = 0) = b/(1 + XT/Kd), also called the **basal promoter activity**. This leakage is smaller the stronger X binds its DNA site. In Figure A.4, the parameter values are XT/Kd = 10 (top curve) and XT/Kd = 50 (bottom curve), both with n = 2. Half-maximal induction is reached at Sx = 3 Kx and Sx = 7 Kx, respectively. The ihalf-maximal induction point, Sx = S1/2, is approximately (when XT >> Kd)

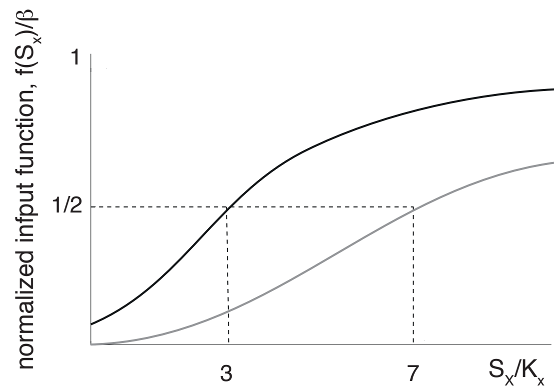


Figure A.4

$$S_{1/2} \sim \left(\frac{X_T}{K_d}\right)^{\frac{1}{n}} K_x \qquad \text{(A.5.2)}$$

The halfway inducer concentration S1/2 can be significantly larger than Kx (Figure A.4). For LacI, for example, XT/Kd ~ 100 and n = 2, so that S1/2 ~10 Kx.
We now turn to describe transcription activators.

A.6 Binding of an activator to its DNA site

In the decade following the discovery of the *lac* repressor, other gene systems were found to have repressors with a similar principle of action. It is interesting that it took several years for the scientific community to accept evidence that there also existed transcriptional activators.
An activator protein increases the rate of transcription when it binds to its DNA site in the promoter. The rate of transcription is thus proportional to the probability that the activator X is bound to D. Using the same reasoning as above, the binding of X to D is described by a Michaelis–Menten function:

$$promoter\ activity = \frac{\beta X^*}{X^*+K_d} \qquad \text{(A.6.1)}$$

Many activators have a specific inducer, Sx, such that X is active, X*, in the sense that it can bind DNA to activate transcription, only when it binds Sx[3].1 Thus, we obtain

Figure A.5 Input function for a gene regulated by an activator as a function of the inducer level.

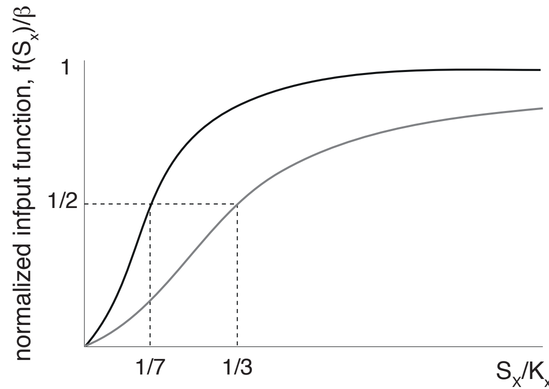


Figure A.5

$$X^* = \frac{X_T S_X^n}{K_X^n + S_X^n} \quad (A.6.2)$$

The genes input function is

$$f(S_x) = \beta \; X^*/(K_d + X^*) \quad (A.6.3)$$

This function, shown in Figure A.5, is an increasing function of signal. The basal transcription level is zero in this regulation function, f(Sx = 0) = 0. Simple activators thus can have lower leakage than repressors. If needed, however, a nonzero basal level can be readily achieved by allowing RNAp to bind and activate the promoter to a certain extent even in the absence of activator.
The inducer level needed for half-maximal induction of an activator can be much smaller than Kx:

$$S_{1/2} \sim \left(\frac{K_d}{X_T}\right)^{\frac{1}{n}} \; K_x \qquad (A.6.4)$$

in contrast to the repressor case (Equation A.5.2). In Fig A.5, for example, Sx ~ 1/3 Kx and Sx ~ 1/7 Kx for the cases of XT/Kd = 10 (bottom curve) and XT/Kd = 50 (top curve), both with n = 2.
Overall, however, similar input function shapes as a function of inducer Sx can be obtained with either activator or repressor proteins. Rules that seem to govern the choice of activator or repressor for a given gene are discussed in Chapter 7.
In this appendix we described a simplified model that captures the essential behavior of a simple gene regulation system, in which proteins are transcribed at a rate that increases with the amount of inducer Sx. Many real systems have additional important details that make them tighter and sharper switches. The present description is sufficient, however, to understand basic circuit elements in transcription networks.

---

[3]In other cases the activator is active when it is unbound to Sx and inactive when it is bound. In such cases, Sx is an inhibitor of X. Similarly, some repressors can be activated by binding Sx. These cases can be readily described using the reasoning in this appendix.

A.6.1 Comparison of dynamics with logic and Hill Input Functions

How good is the approximation of using logic input functions (see Section 2.3.4) instead of graded functions like Hill functions? In Figure A.6, the dynamics of accumulation of a simple one-step transcription cascade are shown, using three different forms of the input function f(X). The input functions are Hill functions with n = 1 and n = 2, and a logic input function. At time t = 0, X* starts to be produced, and its concentration increases gradually with time. The graded input functions show expression as soon as X* appears, whereas the logic input function shows expression only when X* crosses the threshold K. Overall, the dynamics in this cascade are quite similar for all three input functions.

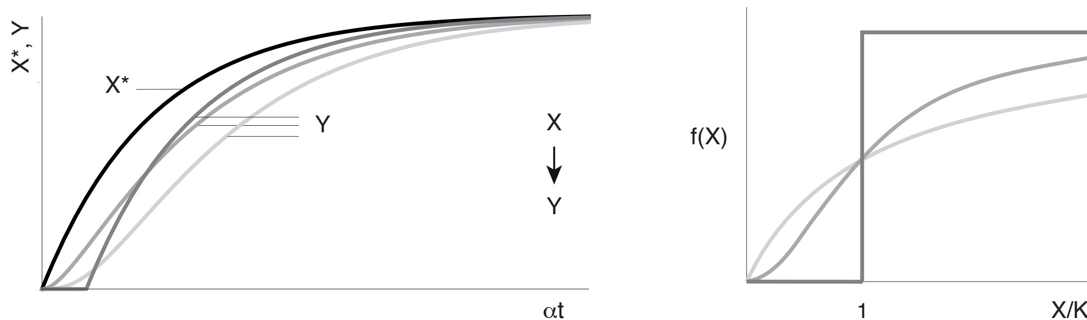
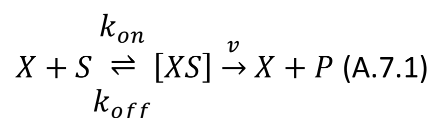


Figure A.6

A.7 Michaelis-Menten enzyme kinetics

We now briefly describe a useful model of the action of an enzyme X on its substrate S, to catalyze formation of product P. Enzyme X and substrate S bind with rate k_{on} to form a complex [XS], which dissociates with rate k_{off}. This complex has a small rate *v* to form product P, so that



$$X + S \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} [XS] \overset{v}{\rightarrow} X + P \quad \text{(A.7.1)}$$

The rate equation for [XS], taking into account the dissociation of [XS] into X + S, as well as into X + P, is

$$\mathrm{d}[XS]/dt = k_{on}\, X\, S - k_{off}\, [XS] - v[XS] \quad \text{(A.7.2)}$$

At steady-state, we obtain

$$[XS] = k_{on}/(v + k_{off}) XS \quad \text{(A.7.3)}$$

If substrate S is found in excess, we need only worry about the conservation of enzyme X:

$$X + [XS] = X_T \quad \text{(A.7.4)}$$

Using this in Equation A.7.3, we find the Michaelis-Menten equation:

$rate\ of\ production = v[XS] = v\, XT \frac{S}{Km + S}$ Michaelis–Menten enzyme kinetics (A.7.5)

where the Michaelis–Menten coefficient of the enzymee is:

$$K_m \; = \; (v \; + \; k_{off}\;)/k_{on} \qquad \text{(A.7.6)}$$

This constant has units of concentration and is equal to the concentration of substrate at which the production rate is half maximal. When substrate is saturating, S >> Km, production is at its maximal rate, equal to $v\,X_T$. Thus, the production rate does not depend on S (that is, it depends on S to the power zero) and is known as **zero-order kinetics**:

$$production\,rate \; = \; v\,X_T \qquad zero - order\;kineticss \qquad \text{(A.7.7)}$$

In the main text we will sometimes make an approximation to this function, in which the substrate S is found in low concentrations, S << Km. In this case, the production rate becomes linear in S, as can be seen from Equation A.7.5 by neglecting S in the denominator. This regime is known as **first-order kinetics**:

$$production\,rate \; = \; v\,X_T\,\frac{S}{K_m} \qquad first - order\;kinetics$$

Further reading

(Ackers, Johnson and Shea, 1982) “Quantitative model for gene regulation by lambda phage repressor.”

(Monod, Wyman and Changeux, 1965) “On the nature of allosteric transitions: a plausible model. “

(Ptashne, 2004) “Genetic Switch: Phage Lambda Revisited”

(Setty *et al.*, 2003) “Detailed map of a cis-regulatory input function.”

(Berg *et al.*, 2002) “Biochemistry Enzymes: Basic Concepts and Kinetics. “

Exercises

1.A.1. Given a simple repressor with parameters \beta, XT, Kd, Kx, and n, design an activator that best matches the performance of the repressor. That is, assign values to β, XT, Kd, and Kx for the activator so its input function will have the same maximal expression, and the same $S_{1/2}$, and the same slope around $S_{1/2}$ as the repressor input function.

1.A.2. Derive the approximate value of diffusion-limited k_{on} based on dimensional analysis. Dimensional analysis seeks a combination of the physical parameters in the problem that yields the required dimensions. If only one such combination exists, it often supplies an intuitive approximate solution to otherwise complicated physical problems. Assume a target protein with a binding site of area a = 1 nm^2, and a small molecule ligand that diffuses with diffusion constant D = 1000 μm^2/sec. The affinity of the site is so strong that it binds all ligand molecules that collide with it.

Solution:

To study the on-rate k_{on}, place a single protein in a solution of 1 M ligand L (concentration of ligand is ρ = 1M = $6\cdot10^{23}$ molecules/liter ~ 10^9 mol/μm^3). The number of L molecules colliding with the binding site of the protein has dimensions of molecules/sec and should be constructed from ρ, D, and a. The combination with the desired dimensions is $k_{on} \sim \rho\, D\, a^{1/2}$,

because D has units of [length]^2/[time] and a has units of [length]^2. This combination makes sense: it increases with increasing ρ, a, and D as expected. Inserting numbers, we find $k_{on} \sim \rho$ D a ~ 10^9 mol/μm^3·1000 μm^2/sec·10^{-3} μm = 10^9 mol/sec, hence $k_{on} \sim 10^9$/M/sec. Note that dimensional analysis neglects dimensionless prefactors and is often only accurate to within an order of magnitude.

1.A.3. What is the expected diffusion-limited k_{on} for a protein sliding along DNA to bind a DNA site. The protein is confined to within r = 1 nm of the DNA. The total length of DNA in a bacterium such as *E. coli* is on the order of 1 mm (!), and the volume of the *E. coli* cell is about ~1 μm^3. Discuss the biological significance of the increase in k^{on} relative to free diffusion in space.

1.A.4. Off-times

(a) Estimate the off-time ($1/k_{off}$) of a diffusion-limited repressor that binds a site with Kd = 10^{-11} M.

(b) What is the off-time of a small-molecule ligand from a receptor that binds it with $K_d = 10^{-6}$ M (bacterial chemotaxis attractants), $K_d = 10^{-12}$ M (mammalian hormone binding to receptors)?

(c) Mammalian ligands that bind a receptor on the cell surface are often taken up into the cell and destroyed or recycled together with the receptor, in a process called endocytosis. Explain how the ligand can remain bound for long enough if endocytosis takes minutes?

Ackers, G. K., Johnson, A. D. and Shea, M. A. (1982) ‘Quantitative model for gene regulation by lambda phage repressor.’, *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.79.4.1129.
Berg, J. M. (Jeremy M. *et al.* (2002) *Biochemistry*. W.H. Freeman.
Berg, O. G., Winter, R. B. and von Hippel, P. H. (1981) ‘Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 1. Models and Theory’, *Biochemistry*. doi: 10.1021/bi00527a028.
Monod, J., Wyman, J. and Changeux, J. P. (1965) ‘On the nature of allosteric transitions: A plausible model’, *Journal of Molecular Biology*. doi: 10.1016/S0022-2836(65)80285-6.
Ptashne, M. (2004) *A genetic switch : phage lambda revisited*, *Cold Spring Harbor Laboratory Press Cold Spring Harbor NY*. doi: 10.1016/j.molcel.2017.02.028.
Setty, Y. *et al.* (2003) ‘Detailed map of a cis-regulatory input function’, *Proceedings of the National Academy of Sciences of the United States of America*, 100(13). doi: 10.1073/pnas.1230759100.
Yagil, G. and Yagil, E. (1971) ‘On the Relation between Effector Concentration and the Rate of Induced Enzyme Synthesis’, *Biophysical Journal*. doi: 10.1016/S0006-3495(71)86192-1.