BE/APh 161: Physical Biology of the Cell

Justin Bois

Caltech

Winter, 2019

Donna and Benjamin M.
Rosen Bioengineering Center

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

7 Statistical mechanics of gene expression regulation

We have been using the tools of statistical mechanics, and two-state and MWC models in particular, to study a host of problems, including ligand-receptor binding, allostery, operation of ion channels, and even single molecule experiments in the homework. We will now use the tools of statistical mechanics to study the regulation of **gene expression**. A gene is **expressed** when its gene product is produced by the cell, first by transcription of mRNA by RNA polymerase and then translation of the mRNA into protein by the ribosomes. As in the previous applications of statistical mecahanics, the power of this approach lies in

- The ease of mathematizing cartoons using states and weights.
- Dissociation constants emerge, and these can be *measured*.
- Allows identification of the "knobs" that can be used to tune gene expression.

7.1 Gene expression preliminaries

To being talking about regulation of gene expression, we need to first understand the basic architecture of a gene. We will focus on bacteria; eukaryotic gene architecture is typically more complex.

Fig. 16 show a cartoon of the promoter region of a gene. The colored rectangle represents the DNA. The light pink region to the right is the start of the coding region of the gene. Ahead of the gene is a **promoter**, which is the part of the DNA that the **RNA polymerase** binds to to start transcription. The promoter region is decorated with binding sites for other molecules generically termed **transcription factors**.

An activating transcription factor, or **activator**, may bind a region near the promoter, and then can have a favorable interaction with the polymerase. As we will see when we work out the statistical mechanics, this results in recruiting more polymerase to the promoter and therefore gives higher expression of the gene.

A repressive transcription factor, or **repressor**, may bind to a part of the promoter region, sometimes called an **operator**. When it does so, it occludes or otherwise inhibits the polymerase from binding the promoter.

It is useful to know some typical numbers about this system.

quantity	value	BNID
RNA polymerase footprint	$\approx$ 40 base pairs	107873
elongation rate	$\approx$ 60 nucleotides/second	103021
initiation rate	$\approx$ 20 transcripts/minute	111997
number of RNA polymerases per cell	$\approx$ 1000	101440

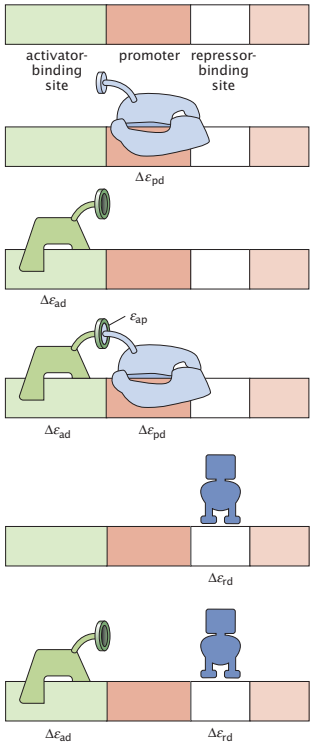

Figure 16: A sketch of the promoter region of a gene. The RNA polymerase (light blue) binds to the promoter to start transcription. It is occluded from doing so when a repressor is bound to the repressor binding site. If an activator is bound to the activated binding site, the polymerase has a favorable interaction with it when bound to the promoter. This figure is adapted from *PBoC2* Fig. 19.18. In *PBoC2*, the energies are denoted with epsilons; we will use E's.

7.2 Separation of time scales

In our modeling, we will assume that the *rate* of production of mRNA transcripts for a particular gene is proportional to the *equilibrium* probability that a polymerase molecule is bound to the promoter of the gene. This seems odd at first glance, that we would use an equilibrium thermodynamic property, p_{bound}, to describe a kinetic process, the rate of production. The key to this assumption being valid is a **separation of time scales** in the transcription process.

Getting the polymerase started is inefficient. The polymerase tends to bind and rebind to the promoter. It often generate small transcripts that are disregarded, and then rebinds and starts over. Typically after many binding and unbinding events, the polymerase gets going with transcription. The binding and unbinding of the polymerase to the promoter is very fast, so fast that it is typically not measurable. The dissociation constant, however, can be measured, and can be as small as $K_{\text{d}} = 5$ nM (BNID 103592).

We can write the reaction scheme of a polymerase getting started making a transcript as

$$\text{unbound} \underset{k_-}{\overset{k_+}{\rightleftharpoons}} \text{bound} \xrightarrow{\alpha_m} \text{elongating}. \tag{7.1}$$

Using mass action kinetics, we can write the dynamical equations of the unbound (u), bound (b), and

elongating (e) states as

$$\frac{\mathrm{d}P_\mathrm{u}}{\mathrm{d}t} = -k_+ P_\mathrm{u} c_\mathrm{p} + k_- P_\mathrm{b} \tag{7.2}$$

$$\frac{\mathrm{d}P_\mathrm{b}}{\mathrm{d}t} = k_+ P_\mathrm{u} c_\mathrm{p} - k_- P_\mathrm{b} - \alpha_m P_\mathrm{b}, \tag{7.3}$$

$$\frac{\mathrm{d}P_\mathrm{e}}{\mathrm{d}t} = \alpha_m P_\mathrm{b}, \tag{7.4}$$

where P_i is the probability of being in state i and c_p is the concentration of available polymerase. If we define the dimensionless time $\tau = k_- t$, the equations are

$$\frac{\mathrm{d}P_\mathrm{u}}{\mathrm{d}\tau} = -P_\mathrm{u} c_\mathrm{p}/K_\mathrm{d} + P_\mathrm{b} \tag{7.5}$$

$$\frac{\mathrm{d}P_\mathrm{b}}{\mathrm{d}\tau} = P_\mathrm{u} c_\mathrm{p}/K_\mathrm{d} + P_\mathrm{b} - \frac{\alpha_m}{k_-} P_\mathrm{b}, \tag{7.6}$$

$$\frac{\mathrm{d}P_\mathrm{e}}{\mathrm{d}\tau} = \frac{\alpha_m}{k_-} P_\mathrm{b}, \tag{7.7}$$

where we have defined $K_\mathrm{d} = k_-/k_+$. In looking at the above, if $\alpha_m/k_- \ll 1$, then the dynamics of the second ODE (7.6) are much slower than the first (7.5). The probability P_u rapidly comes to steady state, so

$$\frac{\mathrm{d}P_\mathrm{u}}{\mathrm{d}\tau} = -P_\mathrm{u} c_\mathrm{p}/K_\mathrm{d} + P_\mathrm{b} \approx 0. \tag{7.8}$$

So, P_b is entirely determined from this equation, which is in fact an equilibrium equation. The last equation, (7.7), then states that the rate of elongation, which is the rate of production of mRNA transcripts, is proportional to the *equilibrium* probability of the promoter being bound, P_b.

So, our goal in quantifying the rate of production of mRNA for a target gene is to compute the probability that the polymerase is bound to the promoter at equilibrium. The statistical mechanical approach we have developed are well suited for this task.

7.3 Statistical mechanics of unregulated gene expression

Let us now consider computing P_b for the case where the expression is unregulated. That is, there are no repressors or activators. There are then two states to consider, the promoter is bound or the promoter is unbound. Let's write a states and weights table.

state	statistical weight
unbound	$\mathrm{e}^{-\beta E_\mathrm{u}}$
bound	$\mathrm{e}^{-\beta (E_\mathrm{b} - \mu_\mathrm{p})}$

Here, we have done what we did in past lectures, subtracting a chemical potential of the polymerase to keep track of the loss of entropic degrees of freedom when it binds the promoter. But what is the chemical potential of the unbound polymerase, μ_p? We need to think a bit more carefully about this.

It is important to know that nearly all polymerases are bound to the genome and plasmids. This is known from experiments where cells divide asymmetrically and the DNA-less cell has virtually no polymerases. So, all of the polymerases are bound to the DNA. They are just bound nonspecifically.

Let P be the number of polymerases that are available to transcribe the gene of interest.[7] Let $N_{\rm NS}$ be the number of nonspecific sites on the genome to which a polymerase can bind. Since the *E. coli* genome is about 4×10^6 base pairs, and there are only about 1000 polymerases per cell, and each polymerase is about 40 base pairs across, $N_{\rm NS}/P \approx 100$ as a lower bound[8], and we will take $N_{\rm NS} \gg P$.

With this in mind, we can rewrite the states and weights table explicitly taking into account the multiplicity of states.

state	statistical weight
unbound	$\frac{N_{\rm NS}!}{P!(N_{\rm NS}-P)!}\, {\rm e}^{-\beta P E_{\rm pd}^{\rm NS}}$
bound	$\frac{N_{\rm NS}!}{(P-1)!(N_{\rm NS}-P+1)!}\, {\rm e}^{-\beta (P-1) E_{\rm pd}^{\rm NS}}\, {\rm e}^{-\beta E_{\rm pd}^{\rm S}}$

Here, $E_{\rm pd}^{\rm NS}$ denotes the energy of nonspecific binding of the polymerase to DNA, and $E_{\rm pd}^{\rm S}$ denotes the energy of specific binding of the polymerase to DNA. If $N_{\rm NS} \gg P$, then

$$\frac{N_{\rm NS}!}{(N_{\rm NS}-P)!} \approx (N_{\rm NS})^P. \tag{7.9}$$

With this approximation, we can write P_b as

$$P_{\rm b} = \frac{\frac{N_{\rm NS}^{P-1}}{(P-1)!}\, {\rm e}^{-\beta (E_{\rm pd}^{\rm S} + (P-1)E_{\rm pd}^{\rm NS})}}{\frac{N_{\rm NS}^{P-1}}{(P-1)!}\, {\rm e}^{-\beta (E_{\rm pd}^{\rm S} + (P-1)E_{\rm pd}^{\rm NS})} + \frac{(N_{\rm NS})^P}{P!}\, {\rm e}^{-\beta P E_{\rm pd}^{\rm NS}}}. \tag{7.10}$$

Dividing top and bottom by the last term in the denominator yields

$$P_{\rm b} = \frac{\frac{P}{N_{\rm NS}}\, {\rm e}^{-\beta \Delta E_{\rm pd}}}{1 + \frac{P}{N_{\rm NS}}\, {\rm e}^{-\beta \Delta E_{\rm pd}}}, \tag{7.11}$$

where $\Delta E_{\rm pd} = E_{\rm pd}^{\rm S} - E_{\rm pd}^{\rm NS}$ is the difference in energy between specific and nonspecific counding. Typically, $\Delta E_{\rm pd} < 0$.

In looking at this expression, it is clear that our μ in our original states and weights table on page 47 is

$$\mu_{\rm p} = E_{\rm pd}^{\rm NS} + k_B T \ln \frac{P}{N_{\rm NS}}. \tag{7.12}$$

This is the same form as the chemical potential of ligands in a dilute solution, with the mole fraction replaced by $P/N_{\rm NS}$. With this convention, we also find that the statistical weight associated with the unbound state is unity. The states and weights table is then conveniently written as

[7] Some of the cell's polymerases may be transcribing genes or are bound to other promoters. We will take P to be all of those available with the correct σ factor.

[8] The ratio is even bigger, since we could consider each one-base shift to be another non-specific binding site.

state	statistical weight
unbound	1
bound	$e^{-\beta(E_{pd}^{S}-\mu_p)} = \frac{P}{N_{NS}} e^{-\beta\Delta E_{pd}}$.

Because it comes up so often, for convenience going forward, we define

$$\rho = \frac{P}{N_{NS}} e^{-\beta\Delta E_{pd}}, \tag{7.13}$$

such that the probability that an unregulated promoter is bound is $P_b = \rho/(1+\rho)$.

7.4 Simple repression

Now, we will consider the case where a repressor can bind to the promoter region and occlude the polymerase from binding. As we write our states and weights table, we are again faced with how to write a chemical potential, this time for repressors. In fact, most repressors are also bound to DNA, either specifically or nonspecifically. We can see this by considering that the dissociation constant for nonspecific binding of repressors to DNA is about 10 µM.[9] The number of nonspecific binding sites, accounting for possible overlap, is about 10^5 per cell, for a concentration of about 200 µM. The equilibrium expression for receptor-nonspecific site binding is

$$K_d = \frac{c_{NS}\, c_R}{c_{R\cdot NS}} = \frac{(c_{NS} - c_{R\cdot NS})(c_R^0 - c_{R\cdot NS})}{c_{R\cdot NS}} \approx \frac{c_{NS}^0 (c_R^0 - c_{R\cdot NS})}{c_{R\cdot NS}}. \tag{7.14}$$

In the last approximation, we have used that fact that there are far fewer repressors than nonspecific binding sites, since repressor copy numbers range from 10 to 10,000 per cell (BNID 102632). We can rearrange this to get

$$c_{R\cdot NS} = \frac{c_R^0\, c_{NS}^0}{K_d + C_{NS}^0} = \frac{c_R^0}{1 + K_d/c_{NS}^0}. \tag{7.15}$$

Because $K_d \ll c_{NS}^0$, we have $c_{R\cdot NS} \approx c_R^0$, so nearly all repressors are bound to DNA.

We therefore know that the chemical potential term in the states and weights table for repressors is $\mu_r = E_{rd}^{NS} + k_B T \ln R/N_{NS}$. So, our states and weights table for repressor-mediated transcription is

state	statistical weight
unbound	1
polymrerase bound	ρ
repressor bound	$e^{-\beta(E_{rd}^{S}-\mu_r)} = \frac{R}{N_{SS}} e^{-\beta\Delta E_{rd}}$

From the states and weights table, we get

$$P_b = \frac{\rho}{1 + \rho + \frac{R}{N_{SS}} e^{-\beta\Delta E_{rd}}}. \tag{7.16}$$

[9] I got this number from Bintu, et al., Curr. Op. Genet. Dev., **15**, 116–124, 2005.

7.4.1 Fold change

A more convenient metric to measure experimentally is the **fold change** in gene expression, defined as

$$\text{fold change} = \frac{P_b}{P_b(R=0)}. \tag{7.17}$$

The unregulated probability of bound polymerase is always $\rho/(1+\rho)$, so it is convenient to write

$$P_b = \frac{\rho}{1+\rho}\,(\text{fold change}) = \frac{\rho}{1+\rho}\,\frac{1}{1+\frac{R}{(1+\rho)N_{\text{SS}}}\,\mathrm{e}^{-\beta\Delta E_{\text{rd}}}} \tag{7.18}$$

The fold change is then

$$\text{fold change} = \frac{1}{1+\frac{R}{(1+\rho)N_{\text{SS}}}\,\mathrm{e}^{-\beta\Delta E_{\text{rd}}}}. \tag{7.19}$$

The value of ρ will vary from promoter to promoter. The term P/N_{NS} is close to the same for all bacterial cells, with

$$\frac{P}{N_{\text{NS}}} \approx \frac{10^3}{10^6} \approx 10^{-3}. \tag{7.20}$$

For the lac promoter, $\Delta E_{\text{pd}} \approx -3k_BT$, and for the T7 promoter, which codes for the protein of the T7 phage, $\Delta E_{\text{pd}} \approx -8k_BT$. Thus, for lac, $\rho \approx 10^{-2}$, and for T7, $\rho \approx 1$. For the former cass, ρ is small, and we have a **weak promoter**. A weak promoter allows for easier regulation; it takes less repressors to see a change in expression levels, since for weak promoters,

$$\text{fold change} \approx \frac{1}{1+\frac{R}{N_{\text{SS}}}\,\mathrm{e}^{-\beta\Delta E_{\text{rd}}}}. \tag{7.21}$$

In this form, we see that the quantity $N_{\text{NS}}\mathrm{e}^{\beta\Delta E_{\text{rd}}}$ is akin to a dissociation constant in ligand-receptor binding. Defining $K_r \equiv N_{\text{NS}}\mathrm{e}^{\beta\Delta E_{\text{rd}}}$, we can write the fold change as

$$\text{fold change} \approx \frac{1}{1+R/K_{\text{d}}}. \tag{7.22}$$

A cell can tune R by regulating the expression of the repressor itself, and evolution can tune ΔE_{rd}.

7.5 Simple activation

Let us now turn our attention to simple activation. In this case, there is no repressor; just an activator that has a favorable interaction with the polymerase. We can again write our states and weights, and can do so taking shortcuts we have already worked out. Specifically, we know that it is always the *difference* in energy between specific and nonspecific binding that comes into the statistical weights. We also know that most activators, like repressors, are bound to promoter regions or to nonspecific sites on the DNA. The only added wrinkle in this example is the extra energy, ΔE_{pa}, in the state where both the activator and polymerase are bound that is due to the favorable interaction between the activator and polymerase.

state	statistical weight
unbound	1
polymerase bound	ρ
activator bound	$\frac{A}{N_{\mathrm{SS}}}\,\mathrm{e}^{-\beta\Delta E_{\mathrm{ad}}}$
activator and polymerase bound	$\rho\,\frac{A}{N_{\mathrm{SS}}}\,\mathrm{e}^{-\beta\,(\,\Delta E_{\mathrm{ad}}+\Delta E_{\mathrm{pa}})}$

The numerator in the expression for P_{b} contains the weights where the polymerase is bound, in this case two of entries from the states and weights table.

$$
\begin{aligned}
P_{\mathrm{b}} &= \frac{\rho + \rho\,\frac{A}{N_{\mathrm{SS}}}\,\mathrm{e}^{-\beta\,(\Delta E_{\mathrm{ad}}+\Delta E_{\mathrm{pa}})}}{1+\rho+\frac{A}{N_{\mathrm{SS}}}\,\mathrm{e}^{-\beta\Delta E_{\mathrm{ad}}}+\rho\,\frac{A}{N_{\mathrm{SS}}}\,\mathrm{e}^{-\beta\,(\Delta E_{\mathrm{ad}}+\Delta E_{\mathrm{pa}})}} \\
&= \frac{\rho}{1+\rho}\,\frac{1+\frac{A}{N_{\mathrm{SS}}}\,\mathrm{e}^{-\beta\,(\Delta E_{\mathrm{ad}}+\Delta E_{\mathrm{pa}})}}{1+\frac{A}{(1+\rho)N_{\mathrm{SS}}}\,\mathrm{e}^{-\beta\Delta E_{\mathrm{ad}}}+\rho\,\frac{A}{(1+\rho)N_{\mathrm{SS}}}\,\mathrm{e}^{-\beta\,(\Delta E_{\mathrm{ad}}+\Delta E_{\mathrm{pa}})}} \\
&= \frac{\rho}{1+\rho}\,\frac{1+(A/K_{\mathrm{d,a}})\mathrm{e}^{-\beta\Delta E_{\mathrm{pa}}}}{1+A/K_{\mathrm{d,a}}+\frac{\rho}{1+\rho}\,(A/K_{\mathrm{d,a}})\mathrm{e}^{-\beta\Delta E_{\mathrm{pa}}}}.
\end{aligned}
\tag{7.23}
$$

Here, we have defined $K_{\mathrm{d,a}}$ analogously to $K_{\mathrm{d,r}}$ from before,

$$
K_{\mathrm{d,a}} = N_{\mathrm{SS}}\,\mathrm{e}^{\beta\Delta E_{\mathrm{ad}}}. \tag{7.24}
$$

This is the dissociation constant activator binding to the promoter region.

We can immediately extract the expression for the fold change,

$$
\text{fold change} = \frac{1+(A/K_{\mathrm{d,a}})\mathrm{e}^{-\beta\Delta E_{\mathrm{pa}}}}{1+A/K_{\mathrm{d,a}}+\frac{\rho}{1+\rho}\,(A/K_{\mathrm{d,a}})\mathrm{e}^{-\beta\Delta E_{\mathrm{pa}}}}. \tag{7.25}
$$

The fold change can actually be less than one if the promoter is strong (or if ΔE_{pa} is large and positive. That means that presence of the activator can actually decrease expression. If we want good control of expression by an activator, then, we need to have a favorable interaction between the polymerase and the activator ($\Delta E_{\mathrm{pa}} < 0$) and a weak promoter ($\rho \ll 1$). Provided this is the case, such that $\rho/(1+\rho) \approx \rho \ll 1$, the maximum possible fold change can be found by taking the limit of large A. We get a maximum fold change of $\mathrm{e}^{-\Delta E_{\mathrm{pa}}}$.

7.6 Cooperative repression

Now imagine a situation where two repressors can bind to the operator. We may get additional energetic contribution if there are two repressors,[10] say ΔE_{rr}. We can again directly write the states and weights table.

[10] *PBoC2* uses the notation $\Delta E_{\mathrm{rr}} = J$.

state	statistical weight
unbound	1
polymrerase bound	ρ
one repressor bound	$2R/K_{\mathrm{d,r}}$
two repressors bound	$(R/K_{\mathrm{d,r}})^2 \mathrm{e}^{-\beta \Delta E_{\mathrm{rr}}}$

The probability of having the polymerase bound is then

$$P_{\mathrm{b}} = \frac{\rho}{1+\rho} \, \frac{1}{1 + \frac{1}{1+\rho}(R/K_{\mathrm{d,r}}) \left(2 + (R/K_{\mathrm{d,r}}) \mathrm{e}^{-\beta \Delta E_{\mathrm{rr}}}\right)}. \tag{7.26}$$

For a weak promoter, this reduces to

$$\begin{aligned} P_{\mathrm{b}} &= \frac{\rho}{1+\rho} \, \frac{1}{1 + (R/K_{\mathrm{d,r}}) \left(2 + (R/K_{\mathrm{d,r}}) \mathrm{e}^{-\beta \Delta E_{\mathrm{rr}}}\right)} \\ &= \frac{\rho}{1+\rho} \, \frac{1}{\left(1 + R/K_{\mathrm{d,r}}\right)^2 + \left(\mathrm{e}^{-\beta \Delta E_{\mathrm{rr}}} - 1\right) (R/K_{\mathrm{d,r}})^2}. \end{aligned} \tag{7.27}$$

The case where there is no enhanced binding of the second receptor, i.e, $\Delta E_{\mathrm{rr}} = 0$, reduces to

$$P_{\mathrm{b}} = \frac{\rho}{1+\rho} \, \frac{1}{\left(1 + R/K_{\mathrm{d,r}}\right)^2}. \tag{7.28}$$

So, cooperative binding, with $\Delta E_{\mathrm{rr}} < 0$, gives greater repression than without cooperative binding.

The analyses in this lecture demonstrate how carefully considering the statistical mechanics of gene expression reveals what parameters, usually energetics of binding interactions, may be adjusted to tune the properties of regulation of gene expression.