1 Simple repression

In class, we derived a mathematical model of how simple repression dictates gene expression levels. In particular, we showed that the fold-change in gene expression is given by

\[
\text{fold-change} = \frac{1}{1 + R N_{NS} e^{-\Delta \varepsilon}}.
\]

where \( R \) is the number of repressors, \( N_{NS} \) is the number of non-specific sites, and \( \Delta \varepsilon \) is the repressor binding energy. Experimentally, we can create bacterial strains that express a YFP reporter under the control of the repressor. If the YFP fluorescence signal is \( F_{\text{reporter}} \), then the fold-change is measured by calculating

\[
\text{fold-change} = \frac{F_{\text{reporter}}(R)}{F_{\text{reporter}}(R = 0)}.
\]

However, there is an extra subtlety that has to be taken into account when obtaining such fluorescence measurements. In particular, because of the intrinsic fluorescence of the cells themselves, there is a spurious contribution to the total fluorescence we measure, namely, \( F_{\text{total}} \). This quantity is given by

\[
F_{\text{total}} = F_{\text{reporter}} + F_{\text{cell}},
\]

where \( F_{\text{cell}} \) is the autofluorescence of the cell. As a result, we need to be able to subtract the cells’ average autofluorescence if the want to report only on \( F_{\text{reporter}} \). Thus, the fold-change is obtained using

\[
\text{fold-change} = \frac{F_{\text{total}}(R) - F_{\text{cell}}}{F_{\text{total}}(R = 0) - F_{\text{cell}}}.
\]
We already learned how to extract fluorescence levels from microscopy images of bacteria using Matlab. In this problem, you will use the code you wrote in order to test the prediction made by Equation 4. You can get our final version of that code under “Final version of image analysis code” in the materials corresponding to the March 9th lecture on the course website. Note that you can also learn more about this protocol from “Computational Exploration: Extracting Level of Gene Expression from Microscopy Images” in chapter 19 of PBoC2.

(a) You can find several images of bacteria under “Full data set” in the materials corresponding to the March 9th lecture. This data set corresponds to bacteria containing varying Lac repressor copy numbers, and a reporter containing a Lac repressor binding site (called O2) that controls YFP expression. The repressor copy numbers for each strain are given in the following table.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Repressor number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delta</td>
<td>0</td>
</tr>
<tr>
<td>WT</td>
<td>22 ± 4</td>
</tr>
<tr>
<td>RBS1147</td>
<td>60 ± 20</td>
</tr>
<tr>
<td>RBS446</td>
<td>124 ± 30</td>
</tr>
<tr>
<td>RBS1027</td>
<td>260 ± 40</td>
</tr>
<tr>
<td>RBS1</td>
<td>1220 ± 160</td>
</tr>
</tbody>
</table>

If you want to learn more about the measurements of the repressor copy number, you can refer to Garcia2011c, which is provided on the website. In addition, a strain called “Auto”, lacking a YFP reporter, is provided with the data set. This strain can be used to measure \( F_{cell} \). Use your code to calculate the fold-change in gene expression as function of repressor copy number by performing the various fluorescence measurements prescribed by Equation 4. Plot your results on a log-log plot.

(b) Estimate the binding energy by fitting Equation 4 “by eye” by trying a reasonable range of parameters for \( \Delta \varepsilon \) as you did in problem 2 of Homework 5. Alternatively, you can perform a least-squares minimization as discussed in class. If you want to go down this route, you might find it better to fit your data to the fold-change expressed in the language of dissociation constants

\[
\text{fold-change} = \frac{1}{1 + \frac{[R]}{K_d}}.
\]

Here, \( [R] \) is the concentration of repressors inside the cell, and \( K_d \) is the dissociation constant of repressor to the DNA.

2 A feeling for the numbers: Hemoglobin

We have adopted hemoglobin as one of our molecules of interest to discuss the statistical mechanics of binding reactions. In this problem, you will perform several estimates to get a feeling for the numbers for hemoglobin.

(a) Do problem 4.1(c) from PBoC2.
(b) Figure out roughly how many $O_2$ molecules you bring in with each breath and how many Hemoglobin molecules it would take to use each and every one of those oxygens. How does this compare with the total number of Hemoglobins in your body calculated in (a)? Hint: You will have to figure out our lung capacity and how many $O_2$ molecules are contained within that volume using the ideal gas law.

3 The MWC model of protein allostery

In class, we introduced dimoglobin as a toy model for describing the binding of oxygen to hemoglobin. This model gave us the opportunity to discuss how an interaction between two bound oxygen molecules can increase the sharpness of the binding curve and lead to cooperativity. As a reminder, the states and weights for this model are shown in Figure 1.

(a) Rederive the result from class that states that the oxygen occupancy (the average number of oxygen molecules bound to dimoglobin) is given by

$$\langle N_{int} \rangle = \frac{2[L]c_0 e^{-\beta \Delta \varepsilon} + 2 \left(\frac{[L]}{c_0}\right)^2 e^{-\beta(2\Delta \varepsilon + \varepsilon_{int})}}{1 + 2[L]c_0 e^{-\beta \Delta \varepsilon} + \left(\frac{[L]}{c_0}\right)^2 e^{-\beta(2\Delta \varepsilon + \varepsilon_{int})}},$$

(6)

where $[L]$ is the oxygen partial pressure, $c_0 = 760$ mmHg is the standard state partial pressure, $\Delta \varepsilon = \varepsilon_b - \varepsilon_{sol}$ is the difference between the ligand binding energy when bound to dimoglobin and when in solution, and $\varepsilon_{int}$ is the interaction energy between two bound oxygen molecules. To make this possible, you will have to invoke the result from class and Homework 5 that states that the energy to take a ligand molecule out of the solution reservoir is given by the chemical potential

$$\mu = \varepsilon_{sol} - k_B T \ln \left(\frac{c_0}{[L]}\right).$$

(7)

Make sure to include and explain all steps in your derivation.

(b) Plot occupancy vs. oxygen partial pressure for $\varepsilon_{int} = -5$ $K_B T$ and for $\varepsilon_{int} = 0$ on a linear-log plot (semilogx in Matlab) in order to show the effect of $\varepsilon_{int}$ on the sharpness of the occupancy curve. Use $\Delta \varepsilon = -5$ $K_B T$ for both curves.

Sharp binding curves do not always imply a direct interaction between bound molecules. An alternative model to understand cooperativity in ligand binding is based on the MWC (Monod-Wyman-Changeux) model of protein allostery. This model is described in detail in Chapter 7 of PBoC. Here, dimoglobin can exist in two distinct conformational states known as “tense” ($T$) and “relaxed” ($R$). In the absence of ligand, the $T$ state of the protein is favored over the $R$ state. We represent this unfavorable energy cost to access the $R$ state with the energy $\varepsilon$. However, the interesting twist is that the ligand binding reaction has a higher affinity for the $R$ state. This has the effect that, with increasing ligand concentration, the balance will be tipped towards the $R$ state, despite the cost, $\varepsilon$, of accessing that state.
We label the binding energies $\varepsilon_T$ and $\varepsilon_R$, which signify the favorable energy upon binding to the molecule when it is in the $T$ and $R$ states, respectively. The states and weights for this model are shown in Figure 2.

(c) Show that the oxygen occupancy in the MWC model is given by

$$\langle N_{MWC} \rangle = \frac{2 \left[ L \right] c_0 e^{-\beta \Delta \varepsilon_T} + 2 \left( \frac{\left[ L \right]}{c_0} \right)^2 e^{-\beta 2 \Delta \varepsilon_T} + e^{-\beta \varepsilon} \left[ 2 \left[ L \right] c_0 e^{-\beta \Delta \varepsilon_R} + 2 \left( \frac{\left[ L \right]}{c_0} \right)^2 e^{-\beta 2 \Delta \varepsilon_R} \right]}{1 + 2 \left[ L \right] c_0 e^{-\beta \Delta \varepsilon_T} + \left( \frac{\left[ L \right]}{c_0} \right)^2 e^{-\beta 2 \Delta \varepsilon_T} + e^{-\beta \varepsilon} \left[ 1 + 2 \left[ L \right] c_0 e^{-\beta \Delta \varepsilon_R} + \left( \frac{\left[ L \right]}{c_0} \right)^2 e^{-\beta 2 \Delta \varepsilon_R} \right]},$$

where $\Delta \varepsilon_T = \varepsilon_T - \varepsilon_{sol}$ and $\Delta \varepsilon_R = \varepsilon_R - \varepsilon_{sol}$.

(d) Show that the MWC model can lead to the the same sharpness in the Oxygen binding curve as direct interaction mechanisms. To make this possible, plot the predictions from both models, $\langle N_{int} \rangle$ and $\langle N_{MWC} \rangle$, as a function of $[L]$. Use $\Delta \varepsilon = -5 \, K_B T$ and $\varepsilon_{int} = -5 \, K_B T$ for $\langle N_{int} \rangle$, and $\Delta \varepsilon_T = 0$, $\Delta \varepsilon_R = -9.5 \, K_B T$ and $\varepsilon = 4 \, K_B T$ for $\langle N_{MWC} \rangle$. Use linear-log axis for your plot (semilogx in Matlab).

Figure 1: Cooperativity model of dimoglobin. States and weights for a cooperativity model of dimoglobin. $\varepsilon_b$ is the binding energy of an oxygen molecule, $\mu$ is the chemical potential (the energy to be paid to take an oxygen molecule out of the solution), and $\varepsilon_{int}$ is the interaction energy between bound molecules.
Figure 2: MWC model of dimoglobin. States and weights for an MWC model of dimoglobin. The energy for dimoglobin to adopt the relaxed state is $\varepsilon$. $\varepsilon_T$ and $\varepsilon_R$ are the oxygen binding energies in the tense and relaxed states, respectively.