Cooperativity & Protein–DNA interactions

1. *Cooperative Binding in Hemoglobin (MWC Model):* Consider the Monod, Wyman, Changeux (MWC) model of cooperativity in hemoglobin. Hemoglobin has $n = 4$ oxygen-binding sites that can exist in a tense (T) or relaxed (R) state, the T state favored by a factor $L \equiv [T_0]/[R_0]$. Given normalized oxygen fugacities of α and $c\alpha$ per site in the two states, the overall fractional occupancy of oxygen is

$$Y = \alpha \frac{(1+\alpha)^{n-1} + Lc(1+c\alpha)^{n-1}}{(1+\alpha)^n + L(1+c\alpha)^n}\,.$$

(a) Estimate the effective Hill coefficient n_h for hemoglobin, either from the maximal slope of $\log[Y/(1-Y)]$ vs $\log\alpha$ in a Hill plot, or from the slope at half-saturation. Comment on the relative importance of parameters $c \ll 1$ and $L \gg 1$ in determining n_h, by considering the limits of $c \to 0$ and $L \to \infty$.

(b) How does the probability of being in the T state change as more oxygen molecules are bound? Calculate the equilibrium probability P_T of being in the T state, and plot it as a function of the normalized oxygen pressure α. Next calculate the conditional probability $P_T(i)$ of being in the T state, given that i molecules of oxygen are bound. Plot $P_T(i)$ as a function of i. What is the effect of oxygen on hemoglobin conformation?

(c) Different species have different mechanisms of adaptation to hypoxia at high altitude. In humans, adaptation to high altitude involves, among other factors, rapid elevation of the level of $2,3-DPG$ (aka $2,3-BPG$) molecule, which is synthesized in red-blood cells and binds preferentially to the T state of hemoglobin. Some birds that fly at high altitude are adapted by having the hemoglobin with a mutation at the interface between its α and β domains, making the transition between the two conformational states of hemoglobin easier, i.e. reducing the free energy difference between the R and T states. How do these two mechanisms of adaptation affect the saturation curve of hemoglobin? Do you expect them to have similar or different effects on oxygen uptake in the lungs, and release in the tissues?

*****

2. (Optional) *Heterogeneous Binding Sites and Partial Cooperativity:* Consider a variant of the MWC model, with the sites that have different affinities, $K_1, K_2, \cdots, K_n$, for oxygen.

(a) Show that at both very low, and very high pressures, the model can be approximated by the classical MWC model with identical sites, but with an effective binding constant K_{eff}. In either regime, which sites contribute most to K_{eff}, the strongest or the weakest?

(b) Can the above model, in the simplified limit of only one state – say for $L = 0$, lead to a sigmoidal Hill plot with $n_h < n$?

(c) Consider another variant of the model, with different affinities but with extreme cooperativity as considered by Hill, i.e. with either no site occupied or all sites occupied. Can such a model provide a sigmoidal Hill plot with $n_h < n$?

*****

3. *Nucleosome-mediated transcription factor binding:* Consider a variant of the Monod, Wyman, Changeux (MWC) model for describing cooperativity in transcription factors (TF)-DNA binding (Mirny, PNAS 2010). In this model, nucleosomal DNA, which contains n TF binding sites, can be in two states: O (open) and N (closed by nucleosome), the latter favored by a factor L, with normalized affinities of TFs K_O and $K_O = cK_N$ per site in the two states. The overall fractional occupancy of a TF binding site is then

$$Y = \alpha \frac{(1+\alpha)^{n-1} + Lc(1+c\alpha)^{n-1}}{(1+\alpha)^n + L(1+c\alpha)^n},$$

where $\alpha = [TF]/K_O$.

(a) Without rederiving, use the results of Problem 1 to discuss how $c \ll 1$ and $L \gg 1$ control cooperativity and accessibility. Explain that for large L (closed chromatin) TFs bind independently, while for small L (open chromatin) binding is cooperative.

(b) *Two TF species A and B with n_A, n_B sites:* Assume identical affinities K and concentrations [A], [B]. If either A or B can evict the nucleosome, show that the nucleosome occupancy is

$$Y_N([A],[B]) = \frac{L(1+c\alpha_A)^{n_A}(1+c\alpha_B)^{n_B}}{(1+\alpha_A)^{n_A}(1+\alpha_B)^{n_B} + L(1+c\alpha_A)^{n_A}(1+c\alpha_B)^{n_B}}, \quad \alpha_{A,B} = [A,B]/K.$$

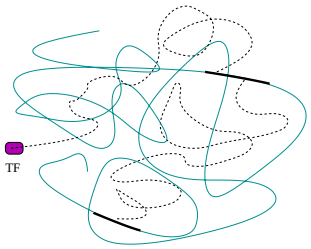
(c) *TF knock-out:* To explore how loss of one factor affects chromatin accessibility, consider a knock-out experiment in which TF A is deleted from the system. (See. e.g. Fig. 2D,G in *bioRxiv 2018 344168*). Using the model, determine how removal of TF A changes the nucleosome occupancy compared with the wild-type case.

*****

4. *Target site location:* Complex transcription machinery in cells is regulated by a set of protein molecules–*transcription factors* (TFs) whose functions can be described as:

• *Receiving a control signal*- This can be the binding or unbinding of a ligand, resulting in initiation or shutting down of the transcription machinery.

• *Finding a specific site on the DNA and binding to it.*



(a) Suppose the protein has to locate a unique binding site on a genome of length M. It may do so by alternately diffusing in solution, and sliding along the DNA, as depicted in Fig. 1.

Given a typical TF diameter of 10nm and cytoplasm dynamic viscosity of approximately 0.1 g $s^{-1}cm^{-1}$, estimate D_{3d} for a TF in cytoplasm. (For 1D sliding, one can assume $D_{1d} \approx 0.1 * D_{3d}$.)

(b) Consider the mechanism of search by sliding and 3D diffusion discussed on the lecture. In addition to theses processes, a protein can make occasional hops, i.e. once it dissociates from DNA, it associates again *at the same place*. Calculate the total search time, assuming that upon dissociation a protein will make a hop with a probability $p \approx 0.9$.

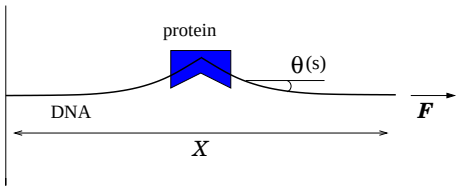
(c) Given the 1D diffusion coefficient D_{1d}, obtain the optimal target location time t_{loc}. The dissociation rate of the proteins from DNA is controlled by the nonspecific binding energy E_{ns}. Estimate E_{ns} for the optimal target location time. Assume $D_{1d} = 1\mu m^2/sec$, $\tau_{3d} = 10^{-3} sec$. Find the location time for $M = 10^6$ base-pairs.

*****

5. (Optional) *Protein-DNA interaction through bending:* In the worm-like chain (WLC) model, the energy cost of deformed piece of DNA of length L is

$$H = \frac{1}{2}\int_0^L ds\ \frac{\kappa}{R^2(s)},$$

where κ is the bending modulus and $R(s)$ is the local curvature radius. Proteins specifically bound to DNA introduce local "kinks" in the DNA structure. Consider the experimental setup in figure below.



(a) Express the local curvature radius through the local inclination angle $\theta(s)$. Modify the above Hamiltonian to include the applied force F and write it down as $H[\theta(s)]$.

(b) By minimizing H, find the equation for $\theta(s)$. Assuming θ is small, solve the equation and calculate the extension X of the DNA as a function of F. Invert the relation and plot the function $F(X)$.

(c) Calculate the energy cost of the DNA deformation. Given that near the protein, $\theta = 0.5$ and that the energy of specific binding is 20 k_BT, estimate the force at which the protein will "pop" from the DNA. Plot the modified force–extension curve.

*****