APh161: Physical Biology of the Cell
Homework 3
Due Date: Tuesday, January 30, 2007

"Facts do not cease to exist because they are ignored." – Aldous Huxley

A. Reading and Writing:

Read Chap. 19 of PBOC and write a referee report using the usual rules for writing such reports.

1. The Hemoglobin Story

We have adopted hemoglobin as one of our molecules of interest to discuss the statistical mechanics of binding reactions. In this problem, you will work out the characteristics of hemoglobin binding from a number of different perspectives. You may find it useful to refer to chap. 7 of PBOC to do this problem (chap. 7 is posted along with this homework).

(a) A feeling for the numbers: in class we discussed the number of red blood cells and the number of hemoglobin molecules in each red blood cell. Note also that these results are implicit in the measurements of the CBC blood test discussed in the first lecture and posted on the website. In this part of the problem, I want you to take it a little further. First, make those estimates concrete by estimating the number of red blood cells in an adult human and the number of hemoglobins per red blood cell. Then, 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?

(b) Revisit the problem of dimoglobin described in class. In particular, compute $p_0(c)$, $p_1(c)$ and $p_2(c)$, the probability of finding 0, 1 and 2 O_2 molecules bound to the dimoglobin molecule. Write expressions for these three probabilities and then plot them as a function of the concentration (all on the same graph). Make sure to explain physically and intuitively what is going on in this system as revealed by these graphs. To make these plots,

you will need the parameters ϵ and J which characterize the binding energy and the cooperativity. Assume that $\epsilon = -5k_BT$ and make several choices of the cooperativity parameter to see how it changes the binding curve.

(c) Work out the average number of oxygen molecules bound on hemoglobin by using equilibrium constants K_1, K_2, K_3 and K_4. K_1 is the equilibrium constant for binding the first oxygen. K_2 is the equilibrium constant for binding the second one and so on. Make sure you write down the definitions of each such binding constant in terms of $[O_2]$ (the oxygen concentration) and $[Hb]$ (the hemoglobin concentration) and introduce $[nO_2 - Hb]$ as the concentration of hemoglobins with n bound O_2. With these definitions in hand, write a full expression for $\langle N_{bound} \rangle$ as a ratio, where $\langle N_{bound} \rangle$ is the average number of bound O_2 on a hemoglobin molecule. Make a plot of the result using the parameters:

$$K_1 = 16.6\ mmHg \tag{1}$$

$$K_2 = 43.9\ mmHg \tag{2}$$

$$K_3 = 4.3\ mmHg \tag{3}$$

$$K_4 = 1.25\ mmHg. \tag{4}$$

Note: the source of this data is the paper by Imai (1990) and is posted on the webpage.

(d) Derive the average number of oxygens bound on hemoglobin using a scheme like that we used for dimoglobin in class. That is, rederive the average number bound using a microscopic model and statistical mechanics. The total energy can be written as

$$E = \epsilon \sum_{\alpha=1}^{4} \sigma_\alpha + J \sum_{(\alpha \neq \gamma)=1}^{4} \sigma_\alpha \sigma_\gamma + K \sum_{\alpha,\beta,\gamma}{}' \sigma_\alpha \sigma_\beta \sigma_\gamma + L \sum_{\alpha,\beta,\gamma,\delta}{}' \sigma_\alpha \sigma_\beta \sigma_\gamma \sigma_\delta, \tag{5}$$

where the parameter σ_i is the occupancy variable for the i^{th} binding site on hemoglobin and the sums are over distinct sites (i.e. there are no repetitions like $\sigma_1\sigma_1\sigma_2$. Find a relation between the parameters used here and those used in part (b) - that is, in this part of the problem you have the parameters ϵ, J, K and L and in part (b) you have $K_1 - K_4$. The idea is to use the K_is to

determine the parameters used in the statistical mechanics model.

(e) One of the favorite tools for dealing with cooperative interactions is the use of the so-called Hill function. In this problem, you will derive the Hill equation and then see to what extent it is possible to use it for thinking about hemoglobin. For a reaction $nA + R \rightleftharpoons nAR$, we need n copies of the ligand A in order to form the complex nAR. In this case, imitate the treatment of binding given in class to find an expression for the fraction of receptors that are in the complexed state by using the definition of the dissociation constant

$$K_d = \frac{[A]^n[R]}{[nAR]}. \tag{6}$$

In particular, show that

$$\theta([A]) = \frac{[A]^n/K_d}{1 + [A]^n/K_d}. \tag{7}$$

The number n is called the Hill coefficient. Make a plot of the binding probability for the cases $n = 1, 2$ and 4 using a sensible value of K from the earlier problems. Here too, using the data provided on the HW website, see if you can find a best fit to the Hb data using a Hill function. This data has been extracted from Imai (1990).

(f) (Extra credit) Fit the data from Imai (1990) to: (i) a hemoglobin model without any cooperativity, where each oxygen binds independently. (ii) Your results from part (c). What we have in mind here is for you to determine the constants K_1, etc. by fitting to the binding curve.

2. Statistical Mechanics of Gene Regulation.

(a) In class, I derived an expression for the probability that RNA polymerase will be on the promoter of interest in the absence of any transcription factors. Reproduce the entirety of that argument including the missing algebraic steps that were glossed over in class and show that p_{bound} may be written as

$$p_{bound} = \frac{1}{1 + \frac{N_{NS}}{P} e^{\beta \Delta \epsilon_{pd}}}. \tag{8}$$

Make a log-log plot of the probability that polymerase is bound as a function of the number of polymerase molecules using $N_{ns} = 5 \times 10^6$ and $\Delta\epsilon_{pd} = -5k_BT$. Also, make sure to discuss the implications of this result for a weak promoter, in particular, comment on the basal transcription rate.

(b) Now generalize the problem you did above to the case in which there is a second promoter competing with the promoter of interest. Assume that the binding energy for that site is identical to that of the promoter of interest and derive an expression for p_{bound} for this promoter. Comment on the relative importance of the nonspecific sites and the competing promoter in inhibiting the binding to the promoter of interest.

(c) Müller-Hill, in Oehler *et al.* (1994), performed a series of impressive measurements of repression in the *lac* operon for the case in which only a single repressor binding site (the primary operator) was present. In this part of the problem you will reproduce the derivation given in class for the problem of repression, culminating in an expression for p_{bound}. Once you have that expression, use your algebraic expressions to fit the repression measured by Müller-Hill - their results are in the paper posted with this homework. Note that for the purposes of this analysis, we define repression as

$$\text{repression}(R) = \frac{p_{bound}(R = 0)}{p_{bound}(R \neq 0)}. \tag{9}$$

To effect the fit, you will use the measured value of repression and the number of repressors (remember that Oehler *et al.* report the number of repressor monomers - divide by 4 to find the number of active repressors) - this leaves as the only unknown the value $\Delta\epsilon_{rd}$ since you already know the value of $\Delta\epsilon_{pd}$ from our earlier treatment of the problem in the absence of repression and its role when calculating fold-activities. Once you have all of these numbers in hand, make a plot of the "fold-activity" as a function of the number of repressors. In fact, "fold-activity" is the inverse of what we mean by repression and is defined as

$$\text{fold-activity}(R) = \frac{p_{bound}(R \neq 0)}{p_{bound}(R = 0)}. \tag{10}$$

Make a log-log plot of the fold-activity in the case of pure repression. What features of the curve change by varying parameters such as the binding en-

ergy? Examine your expression in the limit when the promoter binding is weak and show that in this case the fold change is given approximately by $F_{reg}(R)$ itself. Work the numbers for the case of interest in this problem and show that this is the appropriate limit.

d) Calculate p_{bound} in the case of an activator "recruiting" RNAP. As we saw in class this process can be described using a binding energy of the activator to DNA ($\Delta\epsilon_{ad}$) and an interaction energy between the activator and RNAP (ϵ_{ap}). Demonstrate that the regulation factor is greater than one and show what happens in the limit that the interaction energy between activator and the polymerase goes to zero. Make a log-log plot of the fold activity in this pure activation case. What features of the curve change by varying the parameters?

3. Bioinformatics and HIV.

Go through the bioinformatics tutorial and carry out the associated analysis that is posted with this homework on the webpage.