Problem Set No. 3

Due: Thursday, 11/04/10 at the start of class

Objective: To understand and develop models of protein binding and enzyme kinetics. To understand DNA and RNA structure, replication, transcription, and translation.

Review problems

You should pay special attention to these questions after reading. Note that the answers are given in the back of the book. Formulate your answers fully first and then check them. This can be a significant aid in your understanding of the material, and similar questions may be asked on the final. You do not need to provide written answers in the solutions you hand in.

- ECB 5-6
- ECB 5-8
- ECB 5-9
- ECB 5-18
- ECB 6-8
- ECB 6-9
- ECB 6-15
- ECB 6-19
- ECB 7-7
- ECB 7-17

Problem 1

Leucine zippers are a family of proteins that play important roles as transcriptional factors in eukaryotic cells: by recognizing and binding to specific nucleotide sequences in DNA they can regulate the expression of various genes (i.e., the transcription of particular proteins). The structure of the activator GCN4 of the family of bZip leucine zippers in a complex with DNA has been solved and can be found on the Protein Databank with PDB code 1YSA.

Download this file and view it with Pymol. You should view it in “cartoon” representation for more clarity.

You want to split the protein and DNA into two separate objects. Follow these steps:

1. Display > Sequence
2. Display > Sequence Mode > Chains
3. Select the A and B chains corresponding to the two DNA strands by using the letters at the top of the display window.

4. A button next to “(sele)” > Create object. You should get a new row in the selection pane (“obj1”) and a new row of chain codes at the top of the display window.
5. Select the C and D chains corresponding to GCN4 by using the letters at the top of the display window.
6. A button next to “(sele)” > Create object. You should get a new row in the selection pane (“obj2”) and a new row of chain codes at the top of the display window.
7. A button next to “1YSA” > Delete object. You should now be left with two objects, obj1 and obj2, corresponding to the DNA strands and GCN4, respectively.
8. S button next to “all” > Display As > Cartoon.

a) What is the dominant secondary structure of GCN4? Is it a monomer, dimer, trimer, or other multimeric protein? What kind of structure is it (hint: consult figure 4-13 in the book)?

b) What kinds of amino acid residues in the protein are interacting with the DNA nucleotides, and why do these interactions make sense? To answer this, you will display amino acid residues of different types according to the following color scheme: hydrophobic=gray, positively charged=blue, negatively charged=red:

1. S button next to “obj2” > Sticks. This will show the stick models of the residues in addition to the cartoon representation.
2. Display > Sequence Mode > Residue Codes
3. In the second sequence line corresponding to obj2 (GCN4), click on all strongly hydrophobic residues to select them. These codes are A, C, F, I, L, M, V, W.
4. C button next to “(sele)” > Grays > Gray 50.
5. A button next to “(sele)” > Delete selection.
6. Repeat starting at step 3 for positively charged residues, coloring them blue: K, R.
7. Repeat starting at step 3 for negatively charged residues, coloring them red: D, E.

c) Estimate the number of nucleotide base pairs with which the GCN4 motif interacts. You may want to display the protein using spheres while keeping the DNA representation as a cartoon structure.

Problem 2

Cortisol is a steroid hormone produced by the adrenal cortex often called the "stress hormone" because it is released in response to stress and anxiety. Synthetic cortisol and derivatives are used to treat various illnesses. In the blood stream, cortisol is largely bound to albumin and corticosteroid binding globulin proteins, with all but about 4% of the total cortisol bound and unavailable. Vogeser and Briegel [Clinical Biochemistry 40, 724 (2007)] have hypothesized that the elevation of physiological temperature due to a fever during illness promotes the release of cortisol by unbinding. They measured *in vitro* the fraction of free (unbound) cortisol in blood serum samples from various patients, as a function of temperature. They obtained the following results:

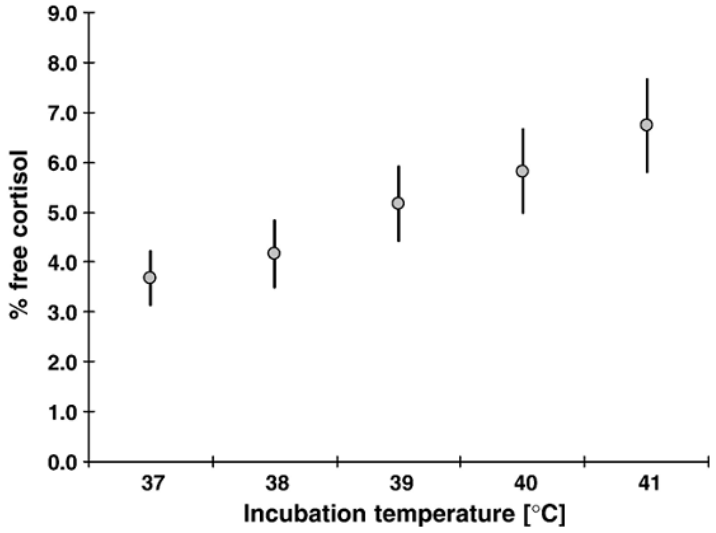

Table 2
Concentrations of cortisol in materials obtained by equilibrium dialysis at different incubation temperatures from sera of ten healthy individuals, and respective cortisol free-to-total ratios [mean (SD)]

Incubation temperature	37 °C	38 °C	39 °C	40 °C	41 °C
Dialysate cortisol concentrations (μg/dL)	0.41 (0.14)	0.46 (0.19)	0.58 (0.21)	0.65 (0.25)	0.75 (0.24)
% Free cortisol	3.7 (1.1)	4.2 (1.3)	5.2 (1.5)	5.8 (1.7)	6.7 (1.8)

Assume that the binding of cortisol (C) to a binding protein (B) is described by the reaction:

$$C + B \overset{K_A}{\leftrightarrow} CB \qquad K_A = \frac{[CB]}{[C][B]}$$

a) Derive a relationship between the fraction of *total* cortisol that is bound, f, the enthalpy of binding, ΔH, and the temperature T. Assume that the heat capacities of reactants and products are roughly equal such that the binding enthalpy is temperature independent. Your answer will also contain a reference temperature T_0 and the concentration of binding partners, $[B]$, which you will assume to be constant and in excess. Hint: consider the constant C_p model.

b) Show that a plot of $\ln\left(\frac{f}{1-f}\right)$ versus $\frac{1}{T}$ will allow you to compute ΔH. Estimate its value in kcal/mol from the data provided above.

Problem 3

In class, we discussed the idea of binding cooperativity in conjunction with the protein hemoglobin. One question that arises in such problems is how we can quantify the *degree* of cooperativity for a particular molecular process. A general model for varying degrees of cooperativity is the so-called Hill model due to Archibald Vivian Hill. The Hill model poses the following simple reaction scheme between protein (P) and ligand (L):

$$P + nL \leftrightarrow PL_n \qquad K_D \equiv \frac{[P][L]^n}{[PL_n]}$$

This equation states that the binding of ligand to protein can only happen if n different ligands all bind at once. That is, n specifies the degree of cooperativity for the system. For noncooperative transitions, $n = 1$. For cooperative ones, $n > 1$.

a) Find an expression for the fraction of bound proteins, $f = [PL_n]/[P]_0$, where $[P]_0$ it the total concentration of protein, as a function of the ligand concentration and the *dissociation* constant K_D.

b) If f were measured as a function of $[L]$, how would you estimate the degree of cooperativity n? Hint: you need to find an equation for a line in which n appears either as a slope or intercept.

c) Sketch f versus $[L]$ for noncooperative binding ($n = 1$), cooperative binding ($n > 1$), and negative or anti-cooperative binding ($0 < n < 1$). What is the slope of the curves at the origin, $[L] = 0$, in terms of K_D and n, and what is its value for the specific cases of cooperative and anti-cooperative binding?

d) Can you think of a mechanism by which binding might be anticooperative, meaning that it becomes more difficult for secondary binding events to occur on the same molecule? Hint: think about allostery.

Problem 4

A number of enzyme-catalyzed reactions in the cell stem from the action of not a single, but two proteins. One example is the protein actin. Binding of ATP to monomeric actin makes it favorable for multiple actin proteins to bind to each other and form long filaments that are an essential structural component of the eukaryotic cytoskeleton. Ultimately, however, the association of two actin proteins stimulates hydrolysis of one ATP to ADP, which eventually causes destabilization and dissociation of the filaments. By this type of mechanism, the self-assembly of actin into filaments becomes a cyclical process of assembly and disassembly. Consider a generic reaction mechanism for a process similar to this one:

$$E + S \overset{K_M}{\leftrightarrow} ES \qquad K_M = \frac{[E][S]}{[ES]}$$

$$ES + ES \overset{k_2}{\rightarrow} E_2S_2 \overset{k_3}{\rightarrow} E_2SP$$

In a very rough correspondence with the actin example, the substrate S indicates ATP and E is an actin protein. For this problem, you will make the following assumptions:

- the first reaction (binding of S to E) is rapid and at quasi-equilibrium
- the mass balance on total enzyme $[E]_0$ used for the first reaction can neglect the species E_2S_2 and E_2SP because they are initially very small
- the formation of E_2S_2 and E_2SP are irreversible steps
- the intermediate E_2S_2 is at quasi steady state
- the substrate concentration is nearly constant with time (e.g., it is in excess).

a) With the approximations above, find an expression for the equilibrium concentration $[E_2S_2]$.

b) Find an expression for the rate of formation of E_2SP, $v = d[E_2SP]/dt$, in terms of the substrate concentration $[S]$, the total enzyme concentration $[E]_0$, and the equilibrium and rate constants above.

c) You decide to perform a series of experiments in which you measure the initial rate of formation v of E_2SP as a function of $[S]$ for fixed $[E]_0$. You identify the concentration $[S]_{1/2}$ at which the rate is 50% of its maximum value, $v = 0.5v_{\max}$, as 40 nM. Find the value of K_M.

d) Consider the alternative case in which, instead of the second reaction above, the mechanism was described by the equation:

$$ES + E \overset{k_2}{\rightarrow} E_2S \overset{k_3}{\rightarrow} E_2P$$

In this case, the enzymes dimerize only when one bound and one free protein associate. You perform the same series of experiments as above, only measuring $v = d[E_2P]/dt$. Show that, in this case, you expect to find a concentration $[S]$ at which the rate is a maximum, beyond which the rate actually decreases. Why does this occur, in physical terms? How would you estimate K_M by knowing the concentration $[S]_{\max}$ that gives $v_{\max}$?

Problem 5

In class we discussed a model for the effects of pH on the activity of enzymes. Here, consider a more elaborate mechanism that allows protonation of the substrate-bound enzyme:

$$EH + S \overset{K_M}{\leftrightarrow} EHS \overset{k_c}{\rightarrow} EH + P \qquad K_M \equiv \frac{[EH][S]}{[EHS]}$$

$$EH \overset{K_3}{\leftrightarrow} E + H^+ \qquad K_3 \equiv \frac{[E][H^+]}{[EH]}$$

$$EH + H^+ \overset{K_4}{\leftrightarrow} EH_2 \qquad K_4 \equiv \frac{[EH][H^+]}{[EH_2]}$$

$$EHS \overset{K_5}{\leftrightarrow} ES + H^+ \qquad K_5 \equiv \frac{[ES][H^+]}{[EHS]}$$

$$EHS + H^+ \overset{K_6}{\leftrightarrow} EH_2S \qquad K_6 \equiv \frac{[EHS][H^+]}{[EH_2S]}$$

Note that this model reduces to the one discussed in class if $K_5 \ll 1$ and $K_6 \gg 1$.

a) Making rapid equilibrium assumptions as appropriate, show that the reaction velocity can be expressed in the form below. Find expressions for $k_{\text{c,app}}$ and $K_{M,\text{app}}$ in terms of the rate coefficients defined above and $[H^+]$.

$$v = \frac{k_{\text{c,app}}[E]_0[S]}{K_{M,\text{app}} + [S]}$$

b) Trypsin is a protease found in the digestive system of many vertebrates produced by the pancreas; it breaks large proteins into smaller peptide fragments by cleaving the polypeptide chain at basic residues. Kasserra and Laidler [Canadian J. Chem 47, 4021 (1969)] found that bovine pancreatic trypsin is active only within a narrow range of pH. The table below shows their results for extracting effective Michaelis-Menten rate constants for trypsin digestion of a model substrate (BAME). According to your expression in part (a), does their data suggest that protonation/deprotonation of substrate-bound trypsin is relevant to the binding mechanism? Explain.

c) Does the data suggest that binding of the substrate does or does not affect the protonation/deprotonation pKa's? In other words, is it true that $K_3 \approx K_5$ and $K_4 \approx K_6$? Explain.

d) The data suggests that there is a maximum in $k_{c,\text{app}}$ with pH. To which parameters in the model does this special pH relate? Find the relationship.

TABLE I

Kinetic parameters for the trypsin-catalyzed hydrolysis of BAME at 25.0° and $I = 0.10$

pH	$\tilde{k}_c$ (s^{-1})	$\tilde{K}_m$ (M)	$\tilde{k}_c/\tilde{K}_m$ ($M^{-1}\ s^{-1}$)
6.0	—	—	0.175 ± 0.005
6.5	0.109 ± 0.009	0.243 ± 0.020	0.448 ± 0.009
7.0	0.185 ± 0.010	0.171 ± 0.012	1.08 ± 0.07
7.5	0.232 ± 0.009	0.129 ± 0.006	1.79 ± 0.10
7.8	0.248 ± 0.008	0.112 ± 0.006	2.22 ± 0.11
8.0	0.265 ± 0.006	0.111 ± 0.003	2.40 ± 0.07
8.5	0.248 ± 0.015	0.098 ± 0.003	2.19 ± 0.21
9.0	0.184 ± 0.012	0.078 ± 0.004	2.37 ± 0.16
9.2	0.168 ± 0.008	0.077 ± 0.002	2.19 ± 0.20
9.5	0.122 ± 0.005	0.052 ± 0.003	2.35 ± 0.18
9.7	0.105 ± 0.005	0.047 ± 0.003	2.24 ± 0.17
10.0	0.070 ± 0.011	0.035 ± 0.003	1.98 ± 0.22

In their notation, $K_{M,\text{app}} = \tilde{K}_m$ and $k_{c,\text{app}} = \tilde{k}_c$.

Problem 6

The melting temperature T_m of hybridized, complementary DNA strands indicates the temperature above which the strands are in higher concentrations in the dissociated than the associated state, with the opposite true below T_m. Consider a solution in which strand A is in dilute concentrations, while the complementary strand B is in excess. Here we define the melting temperature as that which half of the A's are hybridized. One approach to estimating T_m is to use the following equation:

$$T_m = \frac{\Delta H^0}{\Delta S^0 + R \ln[B]}$$

a) Derive this relationship. Here, ΔH^0 and ΔS^0 are the standard-state changes in enthalpy and entropy for hybridizing at T_m, i.e., for the reaction $A + B \rightarrow AB$. In line with standard conditions, $[B]$ is given in molarity. Hint: Note keep in mind the dissociation constant for this reaction, $K_D = \exp(\Delta G^0/RT)$.

b) The melting temperature is sequence-dependent. One way to estimate T_m from the above equation is to find ΔH and ΔS from the sum of all neighboring base pair interactions (i.e., sum of every two base pairs neighboring in sequence). The table below gives approximate values. Here, assume that order doesn't matter, i.e., that AT base paired with TA is the same as TA base paired with AT (although, in reality, it does make some difference). Note that each base pair will appear in two sets of neighbors in this calculation. Estimate the melting temperature of a dilute solution of oligonucleotide ATTCGTGA if its complementary strand is at excess concentration $[B] = 1M$. Be sure to indicate the complementary sequence as well.

Nearest-neighbor sequence (5'-3'/3'-5')	ΔH (kcal/mol)	ΔS (cal/mol/K)
AA base paired with TT	-7.9	-22.2
AT base paired with TA	-7.2	-20.9
CA base paired with GT	-8.5	-22.6
CT base paired with GA	-8.0	-21.6
CG base paired with GC	-10.2	-25.8
GG base paired with CC	-8.0	-19.9
terminal A paired with T	2.3	4.1
terminal G paired with G	0.1	-2.8

c) If you increased $[B]$, would you expect the melting temperature to increase or decrease? Explain.