**Problem Set No. 2**

Due: 10/21/10 at the start of class

**Objective**: To develop a understanding of the thermodynamics and kinetics of protein folding, binding, and enzyme catalysis.

**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 4-10
- ECB 4-14
- ECB 4-15
- ECB 4-17
- ECB 4-21

**Problem 1**

Name the amino acid:

a) contains sulfur but cannot form disulfide bonds; always the first amino acid synthesized in newly-translated proteins

b) not often found in helices or beta sheets; main constituent of collagen

c) biochemical precursor to serotonin; thought to be effective as a sleep aid

d) side chain is an analogue of methanol; it's -OH group is often the site of the attachment of sugars (glycosylation) or phosphate groups (phosphorylation)

e) small shifts in pH near neutral conditions can cause large shifts in net charge due to protonation/deprotonation; as a result, often present in catalytic sites

f) negatively-charged amino acid often used as a flavor-enhancer in food (in a salt form)

**Problem 2**

Consider a two-state folding protein described by the reaction:

$$U \overset{K}{\leftrightarrow} F \qquad K = \frac{[F]}{[U]}$$

where "U" stands for unfolded and "F" folded. Imagine now that a denaturant "D" is added to the system, such as the small molecule urea, that destabilizes the folded state. One possible (approximate) mechanism by which this might occur is through binding to the unfolded state:

$$U + D \overset{K'}{\leftrightarrow} UD \qquad K' = \frac{[UD]}{[U][D]}$$

a) Find an expression for the fraction of protein that is folded, $x$, in terms of $K, K'$, and $[D]$. Hint: consider that the total concentration of protein, regardless of state, must be some constant, say $C_0$.

b) Find an expression for the ratio of the folded to the unfolded protein concentration, $r = [F]/([U] + [UD])$, in terms of $K, K'$, and $[D]$.

c) In the limit that no denaturant is added, $K = r$. On the other hand, for finite denaturant concentration, one might define an "effective" equilibrium constant $K_{\text{eff}} = r$. On this basis, find an expression for the effective folding free energy $\Delta G_{\text{eff}}$ in terms of the actual (zero-denaturant) folding free energy $\Delta G$, $K'$, $[D]$, and $T$.

d) A series of spectroscopic experiments measures $r$ for various denaturant concentrations $[D]$, giving the data below. Using these and an appropriate *linear regression*, determine both $K$ and $K'$. If the temperature is $300\ K$, also estimate the folding free energy $\Delta G$ in kcal/mol.

| $[D]$ (M) | $r$ |
|---|---|
| 0.4 | 238.3 |
| 0.8 | 119.6 |
| 1.2 | 79.8 |
| 1.6 | 59.9 |
| 2 | 48.0 |
| 2.4 | 40.0 |
| 2.8 | 34.3 |
| 3.2 | 30.0 |
| 3.6 | 26.7 |
| 4 | 24.0 |
| 4.4 | 21.8 |
| 4.8 | 20.0 |
| 5.2 | 18.5 |
| 5.6 | 17.1 |
| 6 | 16.0 |

**Problem 3**

In this exercise, you will examine the structure of a small protein obtained from the online Protein Databank, using the popular molecular visualization software Pymol. Pymol should be installed on ECI lab computers, but you can also download a free version for your own use at:

http://pymol.org/educational/

Thrombin is a ~240 amino acid protein that plays a critical role in coagulation and blood clot formation: it converts soluble fibrinogen (another protein) into fibrin, which in turn assembles into long fibers that cross-link to form a mesh over a clotted wound. Thrombin is an important drug target; its activity can be suppressed by binding of a therapeutic ligand to an active catalytic site, and this inactivation can be used to regulate clotting. It has been found that the 12-residue leach peptide hirudin is a natural inhibitor of thrombin, which no doubt aids the leach in extracting blood from its host organism.

Here, you will examine the structure of hirudin bound to thrombin. Go to the Protein Databank website at www.pdb.org. Protein structures are stored as .pdb files in this large repository of structures. You will download the structure with pdb code 1A5G. Enter this code to go to the page for hirudin. Click on Download Files > PDB File (text) to download and save the file to your hard drive.

Before closing your browser, look at the structure of thrombin on the PDB page for 1A5G. Click on the "Sequence" tab. You should see three panels for three different chains. Thrombin consists of two protein chains, a "large subunit" and a "small subunit", and the third chain is hirudin.

a) What is the 1-letter sequence of hirudin? What fraction of it contains negatively charged residues, positively charged residues, and hydrophobic residues, respectively?

b) What is the secondary structure content of the large subunit of thrombin, in terms of percentage alpha helix and beta sheet? Hint: find "DSSP Secondary Structure".

Now close your browser and open Pymol. Notice that Pymol has two windows, a menu window and a display window. Load the 1a5g.pdb structure that you saved in Pymol by using the File > Open command in the menu window.

You will want to first remove water molecules from the structure. To do so, click the "A" button next to 1A5G in the display window and go to Remove Waters.

Next, you will change the display to be a bit more informative. To have Pymol show a backbone cartoon description of the structure, go to the 1A5G S button > as > cartoon. Then, show any disulfide bonds by going to 1A5G S button > disulfides > sticks. Finally, to color the chains differently, go to 1A5G C button > by chain > by chain(*/CA).

At this point, you should see three different chains and any disulfide bonds. The largest chain is the large subunit of thrombin. The smallest is hirudin. You can rotate the structure by dragging in the display window.

c) Looking at the structure, how many intrachain disulfide bonds does thrombin have? How many interchain disulfide bonds does it have?

d) Does hirudin bind to the interior or exterior of thrombin? What interactions might make it a high-affinity binder for thrombin? Hint: consider the sequence of hirudin, and your answer in part a.

On the menu window, go to Display > Sequence. Text should appear at the top of the display window, with one-letter codes for the amino acids in each chain (with the same color scheme). Scroll over the sequence until you reach the 1-letter codes corresponding to hirudin. Click on the one-letter codes for any of the amino acids in hirudin that are hydrophobic, to select them. You should obtain a new row to the right named (sele). On the S button to the right of that, go to Show > Sticks. This will show the stick versions of the hydrophobic residues in hirudin.

e) Do hirudin's hydrophobic residues point towards or away from thrombin? Why do you think this is the case?

Finally, you will make an informative picture of hirudin binding to thrombin. First, click the A button next to (sele) and go to Delete Selection. You need to make two separate objects, one with thrombin and one with hirudin, from the one object that exists. Go to 1A5G A Button > Duplicate Object. You should now have a obj01 item that is a duplicate of your original system.

Then, go to Display > Sequence Mode > Chains. Your 1-letter residue codes should disappear and now there should be three letters, one for each chain, and two lines, one for each object. Click on the "I" code at the top of the display window for the 1A5G object, corresponding to hirudin, to select it. Then go to (sele) A button > remove atoms. Next, click on the "H" and "L" chains in the obj01 object to select them. Then, again, (sele) A button > remove atoms. At this point you should have two objects, one with only the two thrombin chains (1A5G) and one with only the hirudin chain (obj01).

On the 1A5G S button, go to Show as > surface. On the obj01 S button, go to Show as > sticks. Zoom into the binding cavity by clicking the right mouse button and dragging.

f) Change the background to white by going to Display > Background > White. Then, save a picture by going to File > Save Image. Load the image into your favorite word processing software and print it out and attach it to your solutions.

**Problem 4**

An aqueous solution of a protein is initially denatured with the addition of urea. The total concentration of protein is $C_0$. The solution is then suddenly diluted to low urea concentrations by the addition of aqueous buffer. The proteins in this solution, initially unfolded, now begin to fold with time. In addition, folded monomeric proteins then self-associate to form dimers, driven by specific interactions. The process can be described by the set of reactions

$$U \underset{k_u}{\overset{k_f}{\leftrightarrows}} F \qquad K = \frac{k_f}{k_u}$$

$$2F \underset{k_d}{\overset{k_a}{\leftrightarrows}} F_2 \qquad K_A = \frac{k_a}{k_d}$$

a) Write a coupled set of differential equations governing the time evolution of the concentrations $[U], [F], [F_2]$. Be sure to indicate the initial conditions for these equations.

b) Find expressions for the fraction of protein in each species at $t \to \infty$ (i.e., at equilibrium).

c) Assume that the first reaction, folding and unfolding, happens very fast. In addition, assume that dimerization is nearly irreversible, i.e., $k_a \gg k_d$. Using these approximations, find an expression for the fraction of folded monomeric protein, $[F]/C_0$, as a function of time, in terms of $C_0$ and the rate coefficients and equilibrium constants above.

d) Define the "half-life" as the amount of time that it takes for 50% of the proteins in the solution to dimerize. If the half life for a 1 mM solution is 1 minute, what is it for a 5 mM solution? Assume that $K$ is large.

**Problem 5**

Consider the constant heat capacity model for protein folding,

$$\Delta G_{\text{fold}} = \Delta H_{\text{fold},T_f}\left(1 - \frac{T}{T_f}\right) + \Delta C_P\left[T - T_f - T\ln\left(\frac{T}{T_f}\right)\right]$$

a) Show that, according to this model, proteins have a maximum stability at a temperature $T_s$. Find an expression for $T_s$ and for the maximum stability $\Delta G_{\text{fold}}(T_s)$.

b) Prove that $\Delta S_{\text{fold}}(T_s) = 0$.

c) A given protein of 200 amino acids in length is found to have maximum stability around 1 °C and a folding temperature of 65 °C. In addition, to good approximation for many proteins, the <u>*per-residue*</u> change in heat capacity upon folding is about $0.015\ kcal/mol$ °C. Estimate the entropy and enthalpy of folding at $T_f$.

**Problem 6**

The cancer therapeutic Avastin, developed by Genentech, is a humanized monoclonal antibody that binds to and inhibits vascular endothelial growth factor (VEGF). VEGF is a signaling protein that stimulates the growth of new blood vessels (a process called angiogenesis), and that is often overexpressed in nutrient-hungry tumors. Through its inhibition by Avastin, therefore, solid tumors can be, in effect, starved with limited blood supply. Presta et al. [Cancer Research 57, 4593 (1997)] performed a series of experiments using solutions of Avastin with varying concentrations of VEGF; in each, they measured spectroscopically the fraction of Avastin bound to VEGF. They found that approximately 50% of the Avastin was bound to at least one VEGF protein when the *free* concentration of the latter reached 1.6 nM.

a) Assume that the binding mechanism for this process is the following,

$$A + V \underset{K_D}{\leftrightarrow} AV$$

where $K_D$ is the dissociation constant. Find the value of $K_D$ based on the experimental results reported above. Hint: Start with a mass balance on Avastin, assuming that its total concentration is $C_0$.

b) Antibodies actually have two active binding regions that, in principle, might each independently bind to a VEGF protein (so-called *divalent* binding). In this case, we might write the mechanism as:

$$A + V \underset{K_D}{\leftrightarrow} AV$$

$$AV + V \underset{K_D}{\leftrightarrow} AV_2$$

where we have assumed the dissociation constant to be the same for each binding site. According to this mechanism and the results above, find the value of the dissociation constant. Is it higher or lower than before? Does this make sense?

c) Presta et al. found that mutation of a single amino acid in Avastin in the binding interface (glycine to alanine) could reduce its binding affinity by a factor of ~150. By how much must the binding free energy have changed for this mutant? Compare this free energy change to the strength of an average hydrogen bond in water. Assume $T = 300K$.

**Problem 7**

Aggregation of proteins in to large, insoluble, and nonfunctional structures is a major problem in processing protein- and peptide-based therapeutics in the pharmaceutical industries. As an engineer, what strategies might you suggest for reducing the formation of aggregates in such processes? Describe and motivate at least three.