

Biology 5357

Chemistry & Physics of Biomolecules

Examination #1

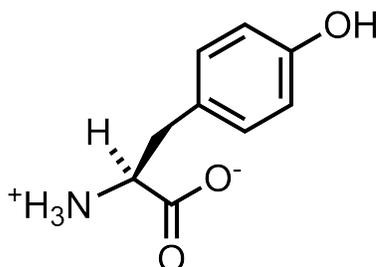
Proteins Module

September 28, 2018

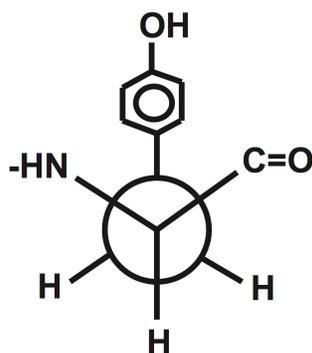
Answer Key

Question 1

(A) (5 points) The 1-letter code for tyrosine is “Y”. The pK_a of the phenolic -OH in the side chain is approximately 10 to 10.5. The structure of the zwitterionic form of tyrosine, showing the L- enantiomer at the α -carbon, is shown below:



(B) (5 points) A Newman projection along the C_α to C_β bond showing the $+60^\circ$ value of the X_I torsional angle is given below. The $+60^\circ$ conformation is relatively disfavored due to the steric clash caused by placing the side chain ring between the two other large substituents at the C_α (*i.e.*, the -NH- and the -C=O).



Question 2

(A) (5 points) For all amino acids except proline, the *trans* peptide bond is preferred over *cis* due to reduced steric crowding in the *trans* form (*i.e.*, the larger groups on either side of the peptide bond, the two α -carbons, are 180° apart). When a proline residue follows the peptide bond, the amide nitrogen is connected to both the C_α and C_δ carbons of the proline, and the *trans* and *cis* peptide bonds are nearly isosteric.

(B) (5 points) For the non-proline residues, $\Delta G = -RT \ln(1000) = 4.1$ kcal/mol in favor of the *trans* isomer at room temperature. In the case of a following proline residue, $\Delta G = -RT \ln(4) = 0.8$ kcal/mol in favor of *trans*. Thus, the $\Delta\Delta G$ for the peptide bond conformers upon switching from a proline to a non-proline C-terminal residue is roughly 3.3 kcal/mol.

Question 3

(A) (5 points) We know $\Delta G = \Delta H - T \Delta S$, or $T = (\Delta G - \Delta H) / -\Delta S = (\Delta H - \Delta G) / \Delta S$. At the melting temperature, T_m , half of the protein molecules are folded and half are unfolded. Thus, the equilibrium constant between folded and unfolded is $K = 1$, and we know from $\Delta G = -RT \ln(K)$ that $\Delta G = 0$ for $K = 1$. So, $T_m = (\Delta H - \Delta G) / \Delta S = (130 - 0) / 0.373 = 348.5K = 75.4^\circ C$

(B) (5 points) At room temperature, $\Delta G = \Delta H - T \Delta S = 60 - (298.15)(0.155) = 13.8$ kcal/mol favoring the folded form of the protein. We know $\Delta G = -RT \ln(K)$, or equivalently, $K = e^{-\Delta G/RT}$. The equilibrium constant between unfolded and folded protein is $K = e^{-13.8/0.59} = e^{-23.4} = 6.95 \times 10^{-11}$. The *percentage* of unfolded lysozyme molecules is less than 10^{-8} , very small!

Question 4

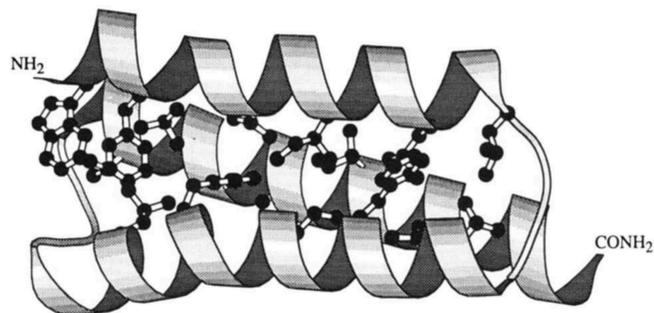
(A) (5 points) Urea is a chemical denaturant that destabilizes essentially all native protein structures. The “red” curve is 1.0 M urea since it is shifted furthest right and has the highest T_m , indicating the greatest stability of the three unfolding curves.

(B) (5 points) Assuming the protein exhibits 2-state folding, we need to determine at least two points in the transition region. For the blue curve (1.5 M urea), we estimate the protein is 0.8 fraction native at $55^\circ C$, and 0.2 fraction native at $77^\circ C$. Then, 0.8 fraction native is an equilibrium constant of $K = [\text{folded}] / [\text{unfolded}] = 4.0$, and 0.2 fraction native is $K = 0.25$. Using $\Delta G = -RT \ln(K)$ and $K = 4.0$ gives a free energy of $\Delta G = -(1.987)(273.15+55)(1.386)/1000 = -0.90$ kcal/mol, indicating the native state is 0.90 kcal/mol more stable than the unfolded state. For $K = 0.25$ at $77^\circ C$, the corresponding free energy difference is $\Delta G = +0.96$ kcal/mol favoring the unfolded form. Finally, we assume that ΔH for unfolding is independent of temperature, which is equivalent to assuming $\Delta C_p = 0$. (Note this is not fully correct true, as there is a positive heat capacity change typically associated with protein unfolding, which is roughly proportional to the number of residues in the protein.) If ΔC_p is 0, we can perform a linear extrapolation from our two data points, $\Delta G = +0.96$ kcal/mol at $77^\circ C$ and $\Delta G = -0.90$ at $55^\circ C$, to yield $\Delta G = -3.44$ at $25^\circ C$.

(C) (5 points) Analysis exactly analogous to that for the blue curve in part (B) gives ΔG at $25^\circ C$ of -2.33 kcal/mol for the brown (2.0 M) curve. So, we have values of $\Delta G = -2.33$ kcal/mol at 2.0 M urea and $\Delta G = -3.44$ kcal/mol at 1.5 M urea. Then linear extrapolation to 0 M urea gives $\Delta G = -6.77$ kcal/mol at $25^\circ C$. Note that the analysis process is more important than the exact numbers, as the -6.77 kcal/mol value is sensitive to the estimates of data point locations on the unfolding curves and

the assumptions described in part (B). Obtaining accurate protein folding thermodynamic values is exacting work, requiring precise data and careful analysis!

(D) (5 points) The alpha 3C three-helix bundle protein was designed and studied in Bill DeGrado's lab (*Protein Science*, 7, 1404-1414 '98). A model of the protein showing the side chain packing in the hydrophobic core between the helices is below. Similar to four-helix bundles discussed in class, the helix packing angle for each pair of helices is at approximately the $+23^\circ$ suggested by Crick's "knobs-into-holes" model.



Question 5

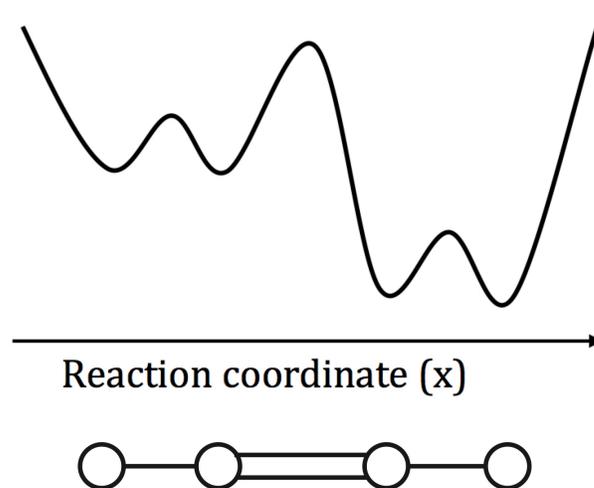
(A) (5 points) The boxed data point is in the high denaturant concentration region of the chevron plot. This point would be obtained by starting with a sample of folded protein, pulsing the sample (using a stop flow apparatus or similar) at time $t = 0$ into roughly 4 M denaturant, and then using a fast measurement technique (fluorescence or similar) to follow the decay of some signal characteristic of the folded form. Once the unfolding data is obtained, a numerical fit can be used to find an exponential term that best accounts for the early stages of the unfolding. The rate corresponding to this exponential term is taken as k_{obs} and used to generate the data point.

(B) (5 points) Again, we will assume simple 2-state folding, which is suggested in this case by the linearity of the folding and unfolding legs of the chevron plots. Both folding legs cross 0 M denaturant near $\ln(k_{fold}) = 7.0$. Extrapolation of the unfolding legs to 0 M denaturant gives $\ln(k_{unfold}) = 3.8$ for the isolated domain, and $\ln(k_{unfold}) = 2.3$ for the domain in the context of the multidomain construct. Then combining the relations $K = [\text{folded}] / [\text{unfolded}] = k_{fold} / k_{unfold}$ and $\Delta G = -RT \ln(K)$, we have $\Delta G = -RT [\ln(k_{fold}) - \ln(k_{unfold})]$. For the isolated domain, the resulting free energy differences are $\Delta G = 1.9$ kcal/mol at 25°C for the isolated domain, and $\Delta G = 2.8$ kcal/mol for the same domain in the multidomain construct.

(C) (5 points) A simple explanation of the data would be to suggest the domain is stabilized in the multidomain construct via interdomain interactions. Then the unfolding rate of the domain in question would be slower in the context of the multidomain construct. If the domain folds independently, prior to multidomain assembly, then the folding rate would be similar for the isolated domain and the domain in the full construct.

Question 6 (10 points)

As shown below, there will be four MSM states corresponding to the four minimum energy basins. The slowest interconversion is associated with the highest barrier, as shown by the double line connecting the two middle states along [x].



Question 7

(A) (5 points) The free energy difference is $\Delta G = -RT \ln(0.1/0.01) = 1.38 \text{ kcal/mol}$.

(B) (5 points) The probability of the folded form at equilibrium is 1, since there is flux into the folded state, but folded molecules cannot return for the unfolded state.

Question 8 (5 points)

Any one of the following: (1) implicit solvent, which reduces the number of degrees of freedom, but cannot describe short-range solvent interactions involving explicit molecular positions, orientations and interactions (2) coarse-graining, which reduces the number of degrees of freedom, but limits the resolution and often the accuracy of the underlying simulations, (3) metadynamics, which adds a biasing potential that encourages the generation of new conformations, (4) accelerated molecular

dynamics, which reduces the well depths of low energy regions to enhance configurational sampling, but modifies the underlying potential energy surface such that it is difficult to extract correct thermodynamic values, (5) replica exchange, which increases temperature to facilitate crossing of large energetic barriers, but require multiple simulations and does not scale well with system size, and (6) Markov state models, which use many short trajectories to build a kinetic model of the potential surface, but require large amounts of data and careful analysis to ensure a statistically valid model.

Question 9 (10 points)

$$\begin{aligned}
 p_{u_1}(x_i) p_{u_2}(x_j) p(i \rightarrow j) &= p_{u_2}(x_i) p_{u_1}(x_j) p(j \rightarrow i) \\
 \frac{p(i \rightarrow j)}{p(j \rightarrow i)} &= \frac{p_{u_2}(x_i) p_{u_1}(x_j)}{p_{u_1}(x_i) p_{u_2}(x_j)} \\
 &= \frac{e^{-u_2(x_i)/kT} e^{-u_1(x_j)/kT}}{e^{-u_1(x_i)/kT} e^{-u_2(x_j)/kT}} \quad \left. \begin{array}{l} \\ \end{array} \right\} \text{Z's cancel} \\
 &= e^{-(\Delta u(x_i) - \Delta u(x_j))/kT} \quad \Delta u(x_i) = u_2(x_i) - u_1(x_i)
 \end{aligned}$$

$$p(i \rightarrow j) = \min(1, e^{-[\Delta u(x_i) - \Delta u(x_j)]/kT})$$