Biology 5357

Chemistry & Physics of Biomolecules

Examination #1

Proteins Module

September 29, 2017

Answer Key

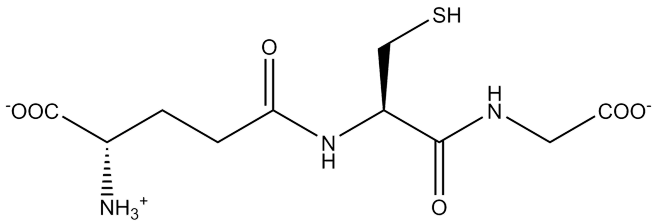
Question 1

(A) (5 points) Structure (b) is more common, as it contains the shorter connection between strands, one that is compatible with the standard right-handed twist between adjacent strands of β□sheet.

(B) (5 points) The simple Crick knobs-into-holes model of helix packing predicts a preferred packing angle of +23° between α□helices motif, which would lead to a left-handed helical bundle. However, the helix-to-helix connections at the end of a 3-helix bundle prefer a negative interhelical angle which is more consistent with the twist of the connection of the two helices. Thus, we have competing effects, and the preference for right- *vs*. left-handed 3-helix bundles is quite small.

Question 2

(A) (4 points)



(B) (3 points) The free –SH groups of the Cys residues of two glutathione molecules react via oxidation (net removal of the two sulfhydryl hydrogens is an "oxidation" event) to form a disulfide bridge, R–S–S–R.

(C) (3 points) Glutathione is able to shuttle between a monomeric, reduced sulfhydryl form, and an oxidized disulfide bridged dimer. Glutathione catalyzes protein folding by facilitating the equilibration of reduced and various oxidized forms of the Cys residues in a folding protein, thus increasing the rate of equilibration to the final correctly folded form with native disulfide bonds. The same function if provided in vivo by protein disulfide isomerases such as thioredoxin.

Question 3

(A) (5 points) There are several ways to get this estimate. One method is outlined here. First, note the data in panel (c) the overnight refolding data is already transformed to have flat baselines and y-axis values from 0 to 100. It is generally possible to transform raw data into this format. From inspection of the plot, the midpoint of the folding transition is at [urea] = 3.8M, and the slope of the curve at

the midpoint is approximately 80 [urea]$^{-1}$. For some points chosen in the linear portion of the curve, near the midpoint, we have:

[urea]	f_N	$K_{([N]/[U])}$	$\Delta G_{(N\ vs.\ U)}$
3.6	66	1.94	-0.39
3.8	50	0	0
4.0	34	0.515	0.39

where f_N is the %fraction of native, and K converts to ΔG via $\Delta G = -RT \ln(K)$. Then, assuming the ΔG value is linear with [urea], and extrapolating to [urea] = 0M, gives $\Delta G\,(0M) = -0.39 \times (3.8 / 0.2) = -7.4$ kcal/mol. Thus, we estimate the folded form of pseudoazurin is 7.4 kcal/mol more stable than the unfolded form. This is a fairly typical value for a wild type, monomeric, globular protein.

(B (5 points) Taking T_1 as 5°C and T_2 as 25°C, we are given that $k_{T2}/k_{T1} = 14$. Substituting these values into $\ln (k_{T2} / k_{T1}) = (E_a / R)\,(T_2 - T_1) / (T_1\,T_2)$ and solving for E_a gives an activation enthalpy of 21.7 kcal/mol.

(C) (5 points) The stability of the protein relative to its unfolded state is typical, but the rate of folding is very slow. And the enthalpic barrier is also large at over 20 kcal/mol. In lecture, we discussed two features that can result in anomalous folding of simple monomeric globular proteins: formation of correct disulfide bonds, and cis vs. trans isomerization of peptide bonds preceeding Pro residues. Since the pseudoazurin sequence contains lots of Pro, but no Cys, we might suspect Pro isomerization as the slow folding phase. This could be tested via sequential mutation of the Pro sites to Ala or another residue, and measuring folding rates for the mutant proteins.

Question 4

(A) (5 points) The reported values can be obtained via analysis of an equilibrium temperature unfolding experiment. Some spectroscopic or other property is measured as a function of temperature, typically resulting in a sigmoid-shaped plot. After fitting baselines to the low and high temperature regions, the factions folded and unfolded are computed for the transition region. These fractions are converted to equilibrium constants and then on to free energies at each temperature in the transition region. A plot of ΔG *vs*. T is then curve fit (it would be linear if ΔC_p were zero, which is usually not the case for protein folding). The ΔS value will then be the slope of this plot at each T, and ΔH can be obtained via $\Delta G = \Delta H - T\Delta S$.

(B) (5 points) What percentage of the time is the peptide in the β-hairpin conformation in each of these two solvents? In pure water, $\Delta G = +7.2 - (298)(23/1000) = +0.35$ kJ/mol. Then using $\Delta G = -RT \ln(K)$ yields an equilibrium constant for the folded hairpin of 0.87, which converts to 46.5% hairpin. Corresponding calculations for 50% methanol give $\Delta G = -4.53$ KJ/mol, K = 6.27, and an 86.3% hairpin content.

(C) (5 points) Upon changing the solvent from pure water to 50% methanol, the values exhibit a reduced "hydrophobic effect". The values in 50% methanol are more like those expected in the "gas phase", with the ΔH being strongly favorable due to hydrogen bonds in the folded form, and the ΔS unfavorable due to greater rigidity in the folded vs. unfolded state. In pure water, the ΔC_p is much larger, also indicative of a significant hydrophobic effect (*i.e.*, the large temperature dependence of ΔH and ΔS is a hallmark of the effect).

Question 5

(A) (5 points) For the F45W mutant with the affinity tag (red dots), the folding arm of the chevron plot crosses the y-axis with a $\ln(k_{fold})$ of about +5.7. Exptrapolation of the unfolding arm to give its rate in pure water (*i.e.*, the y-intercept) gives $\ln(k_{unfold})$ of -5. Then the equilibrium constant is calculated as $\ln(K) = \ln(k_{fold}) - \ln(k_{unfold}) = 5.7 - (-5) = +11.7$, or a value of $K = 4.4 \times 10^4$. Substitution into $\Delta G = -RT \ln(K)$ gives a free energy -6.3 kcal/mol.

(B) (5 points) Since the mutant without the affinity tag is less soluble, it is possible the nonlinearity of its chevron plot at low denaturant concentration is due to aggregration or insolubility (which would lead to a deviation from ideal 2-state behavior).

(C) (5 points) The aggregation hypothesis from (B) could be tested by redetermining the chevron plot at a series of lower protein concentrations.

Question 6

(A) (5 points)

$$\rho(x) = \exp(-U/kT)/Z \quad or \quad \exp(-\Delta G/kT)/Z$$

where Z is the "partition function", which is the sum of Boltzmann factors over all states

(B) (10 points) Curve B (blue) is preferred since it provides a better, more consistent, fit to the probability values in the higher probability regions of the reference distribution. Since the high probability regions are weighted more heavily, a good fit to relative values in those regions will more effectively lower the relative entropy.

Question 7

(A) (5 points) System A will reach a reasonable equilibrium.

(B) (5 points) The systems in B and C are not normalized (i.e., the probabilities of leaving a state do not sum to one for each state). This will lead to source and sinks, and these systems will not reach equilibrium.

Question 8

(A) (5 points) Any two of the following: (1) the requirement of small time steps, (2) large number of atoms in biological systems, (3) difficult of modeling electrostatic interactions, and (4) pairwise nature of energy calculations.

(B) (5 points) Any two of the following, where a brief description of the reason for the speedup is required in addition to the name: (1) implicit solvent, which reduces the number of degrees of freedom, (2) coarse-graining, which reduces the number of degrees of freedom, (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, (5) replica exchange, which increases temperature to cross large energetic barriers, and (6) Markov state models, which use many short trajectories to build a kinetic model of the potential surface.