Biology 5357

Chemistry & Physics of Biomolecules

Examination #1

Proteins Module

September 29, 2023

Answer Key

Question 1 (16 points; A-D = 4 pts each)

Close to its melting temperature (T_m), lysozyme undergoes reversible thermal unfolding at pH 7 with $\Delta H° = 130$ kcal/mol and $\Delta S° = 373$ cal/mol/K. At room temperature of approximately 25°C, $\Delta H° = 60$ kcal/mol and $\Delta S° = 155$ cal/mol/K. Remember that 0°C = 273.15 K.

(A) Estimate the melting temperature of lysozyme.

At the T_m the protein is half folded and half unfolded, so $K_{eq} = [N] / [U] = 1$. Then $\Delta G = -RT \ln (K_{eq}) = 0$, since $\ln(1) = 0$.

In addition, $\Delta G = \Delta H - T\Delta S$, or $T_m = (\Delta H - \Delta G) / \Delta S = (13000 - 0) / 373 = 348.5K = 75.4°C$.

(B) What is the stability of folded lysozyme in kcal/mol, relative to its unfolded form, at room temperature?

At room temperature, $\Delta G = \Delta H - T\Delta S = 60000 - (298.15)(155) = 13787$ cal/mol = 13.8 kcal/mol.

Note this is ΔG for unfolding, N → U, so the native (folded) form is 13.8 kcal/mol more stable than unfolded.

(C) What percentage of lysozyme molecules are unfolded, on average, at room temperature?

At 298.15 K, we have $\Delta G = -13.8$ kcal/mol $= -RT \ln (K_{eq})$, so $K_{eq} = e^{13.8/RT} = e^{23.27} = 1.28 \times 10^{10}$.

Thus, effectively all molecules are folded, and fewer than 1 in 10 billion is unfolded.

(D) Explain the change in the enthalpy and entropy of unfolding between the two temperatures. For example, rationalize the large increase in enthalpy of unfolding from 60 to 130 kcal/mol upon shifting from 25°C to T_m.

When a globular, folded protein unfolds, it exposes a large additional amount of hydrophobic surface area to water. The signature of the hydrophobic effect is that as temperature increases, nonpolar surface exposure becomes more unfavorable enthalpically, and relatively more favorable entropically. At lower temperature water forms rigid "clathrates" surrounding hydrophobic surfaces resulting in an entropic penalty. At higher temperature, clathrate structures melt leading to poorer hydrogen bonding, and an unfavorable enthalpy change.

Question 2 (15 points; A-C = 5 pts each)

Chymotrypsin inhibitor 2 (CI2) is a small globular protein of 83 residues, whose folding has been much studied by Alan Fersht's group and others in the biophysics community.

(A) Which of the diagrams above depicts the correct structure of CI2? Describe at least two features that justify your choice.

The structures shown are mirror images, and that shown on the right is correct. The correct structure has a right-handed α-helix, and a negative dihedral ("twist") angle between adjacent strands of β-sheet (*i.e.*, the closer yellow strand must be rotated counterclockwise to align it with the more distant orange strand).

(B) Does the CI2 structure contain a β-α-β motif? Explain.

The 3D arrangement is similar, but technically CI2 does not have a β-α-β motif. The helix connects directly to the yellow strand but not to the orange strand, so topologically it is not an example of the motif.

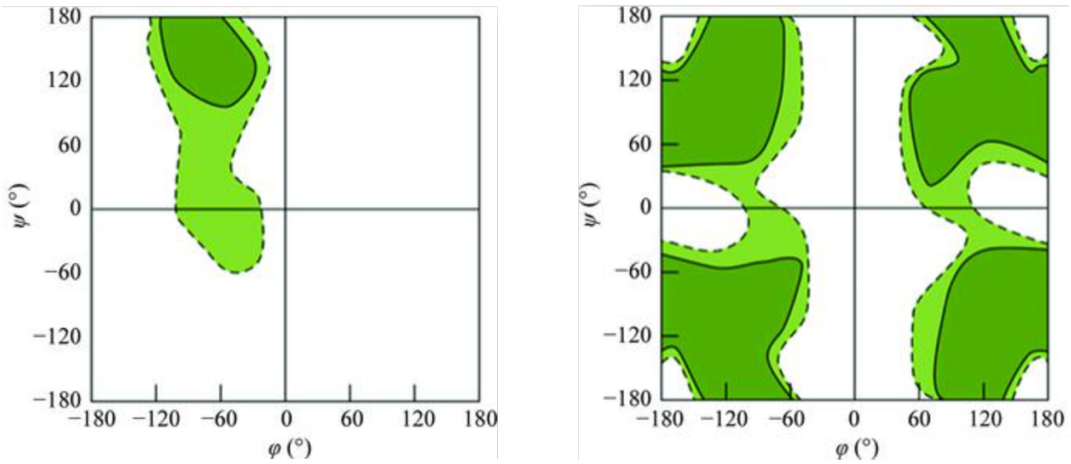
(C) What is the generally preferred arrangement for an α-helix packing against a β-sheet? Does the CI2 structure appear to follow this preference?

An α-helix prefers to lie diagonally across a face of a β-sheet, such that the helix packs along the high diagonal ridge. In CI2 the helix connects at one end of the yellow strand, and runs diagonally toward the "high" end of the orange strand.

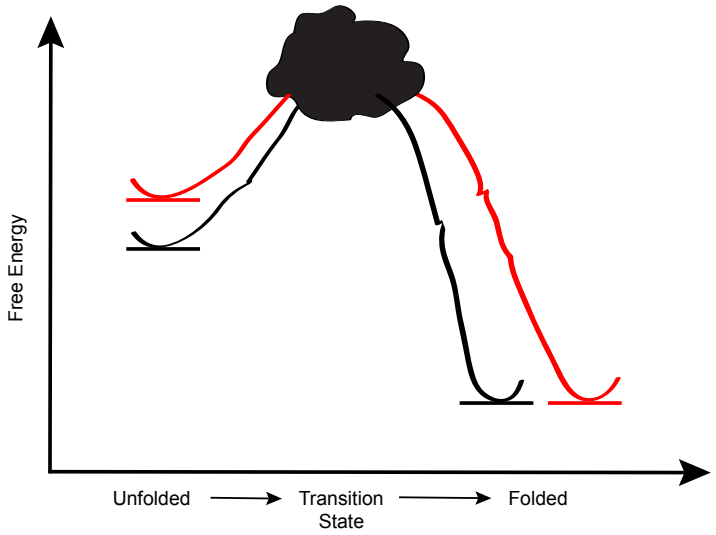
Question 3 (16 points; A-D = 4 pts each)

(A) Provide a rough sketch of the Ramachandran map for an Ala residue and for a Gly residue. Label the axes and indicate regions of the maps corresponding to α-helix and β-sheet structure. How would the map differ for a Pro residue?

The plot on the left is for Ala, and all residues except Gly and Pro. The plot on the right is for Gly. The α-helix region is around (ϕ, ψ) of (-60°, -40°), and β-sheet structure is the broad upper left quadrant around (ϕ, ψ) of (-120°, +120°). The ϕ value for Pro is restricted due residing in a 5-membered ring, and often adopts a value of around -20°, with ψ values centered near -30° or +130°.

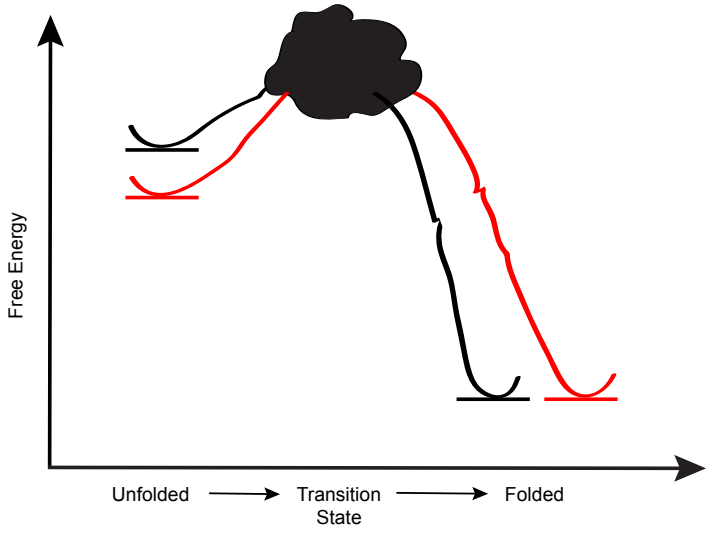


(B) At some position in a folded globular protein of known structure you have made a Gly-to-Ala mutation. Assume the Ala mutant is fully compatible with the tertiary fold of the protein. Sketch a free energy diagram showing how the native and denatured states of the wild type and mutant proteins compare.



Wild type (Gly) protein shown in black, mutant (Ala) protein shown in red. The relative energies of the two transition states are unknown, hence the "black cloud".

(C) Repeat part B above, only this time for a Pro-to-Ala mutation.



Wild type (Pro) protein shown in black, mutant (Ala) protein shown in red. The relative energies of the two transition states are unknown, hence the "black cloud".

(D) For the above two cases, what happens to the stability of the mutant protein when compared to the stability of the wild type protein? Explain.

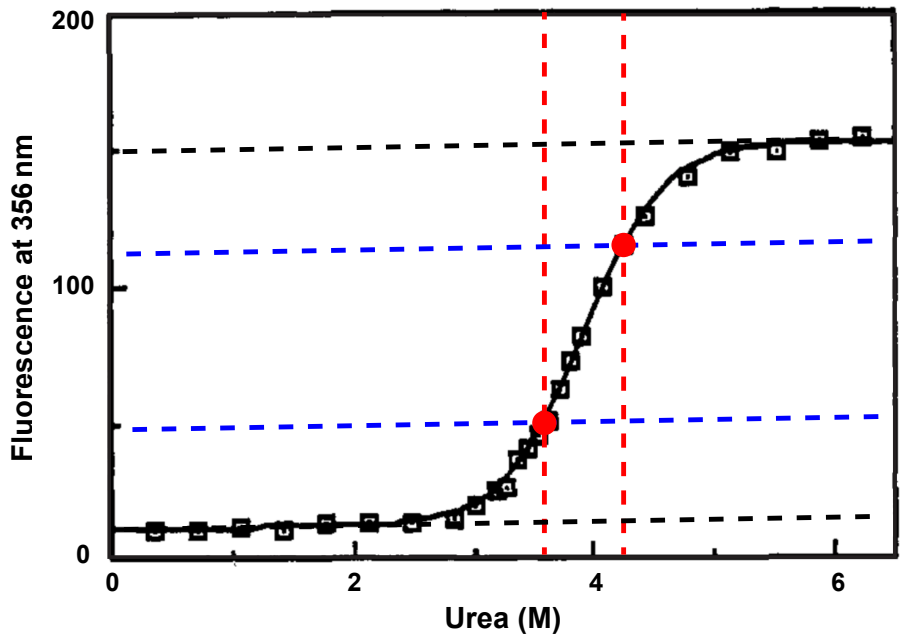
For both mutations, the biggest change is in the stability of the unfolded form. The increased flexibility of the Gly mutant results in a relative stabilization of the unfolded form, compared to Ala wild type. Thus, the folded *vs.* unfolded stability of the Gly mutant will be decreased.

For the Pro mutant, the argument is reversed. In this case the unfolded mutant is destabilized relative to the unfolded wild type. This is due to restricted motion and flexibility in the unfolded Pro mutant, which is unfavorable entropically. As a result, the folded *vs.* unfolded stability of the Pro mutant is increased relative to wild type.

Question 4 (18 points; A-C = 5 pts each, D = 3 pts)

Shown below are an equilibrium folding-unfolding curve and a kinetic "chevron" plot for the CI2 protein. In both cases, urea was used as the denaturing agent, and experiments were performed at room temperature of 25°C.

(A) Analyze the equilibrium folding-unfolding curve to estimate a numerical value for the stability of folded CI2 relative to its unfolded form.



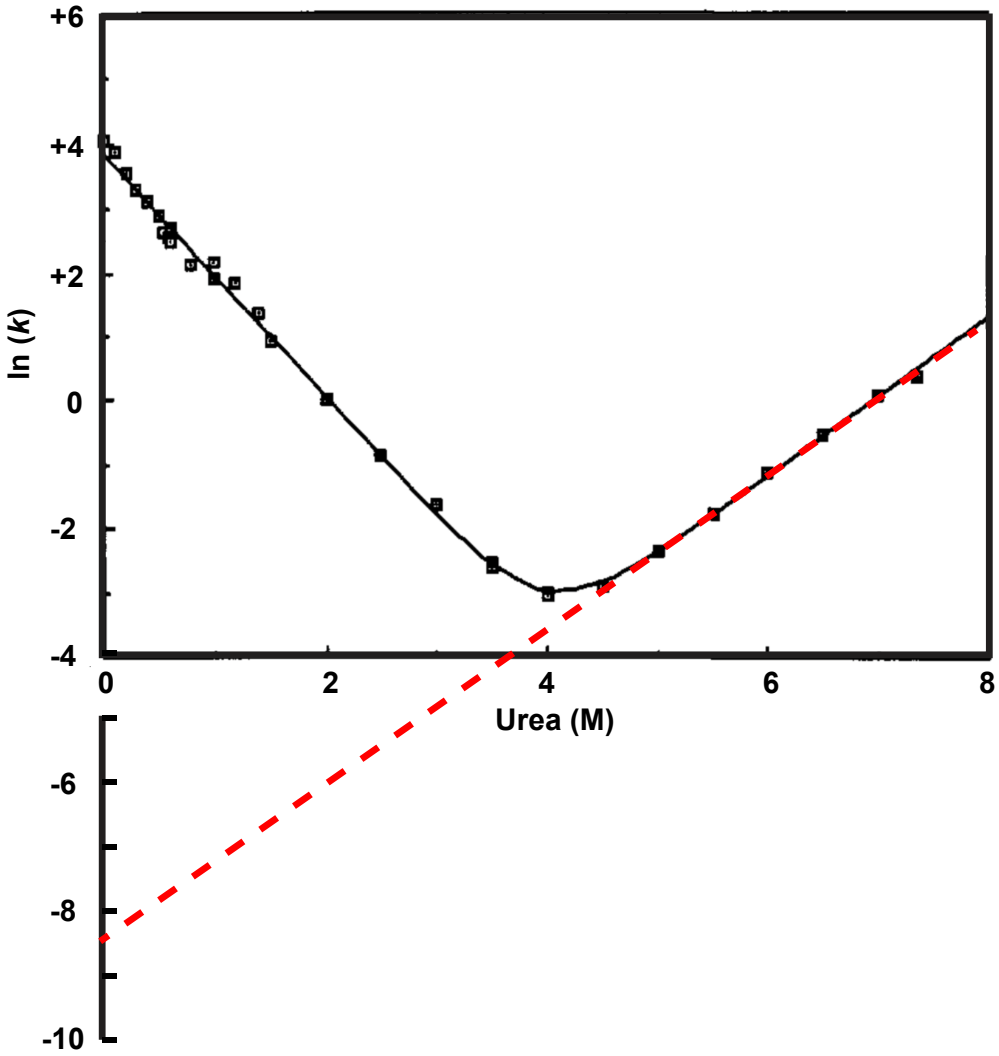
Consider the two red dots in the above plot. The lower left red dot is at about 3.58M urea, and interpolating between the two black dashed baselines gives approximately 74% folded protein. The upper right red dot is at 4.25M urea and 28% folded.

Then the lower left red dot has K_{eq} = 74 / 26 = 2.85, and $\Delta G = -RT \ln (K_{eq})$ = (-1.987) (298.15) ln (2.85) = -620 cal/mol = -0.62 kcal/mol. Similarly, for the upper right red dot, K_{eq} = 28 / 72 = 0.39, and $\Delta G = -RT \ln (K_{eq})$ = (-1.987) (298.15) ln (0.39) = +560 cal/mol = +0.56 kcal/mol.

Assuming that ΔG is linear with respect to urea concentration, we can use the slope between the two red dots to extrapolate to 0M urea, starting from red dot with ΔG = -0.62 at 3.58M urea. Then ΔG at 0M urea is:

ΔG = -0.62 - (3.58) ([0.56-(-0.62)] / [4.25-3.58]) = -6.93 kcal/mol

(B) Analyze the chevron plot to estimate a numerical value for the stability of folded CI2 relative to its unfolded form. How does your value compare to that determined in part A? Note "ln" is the logarithm in base *e*, not the logarithm in base 10 or "log", where ln(x) = 2.303 log(x).



The folding arm of the chevron curve crosses the Y-axis of 0M urea at roughly a ln(k_F) of +3.75, or $k_F = e^{+3.75} = 42.52$. Extrapolating the unfolding arm (red dashed line) to 0M urea gives a ln(k_U) of -8.35, or $k_U = e^{-8.35} = 2.36 \times 10^{-4}$. And $K_{eq} = k_F / k_U = (42.52 / 2.36 \times 10^{-4}) = 1.63 \times 10^5$. Then, $\Delta G = -RT \ln (K_{eq}) = (-1.987)\ (298.15)\ \ln\ (1.63 \times 10^5) = -7109$ cal/mol = -7.11 kcal/mol. This stability values is close to the corresponding value from the equilibrium folding-unfolding curve.

(C) A simple Arrhenius model from transition state theory, given by the equation $k = D \exp(-\Delta G^{\#} / RT)$, often provides an excellent approximation of the protein folding rate. Here, k is the rate, $\Delta G^{\#}$ is the free energy of the transition state for the process, and R is the gas constant of 1.987 cal/mol/K. Note D is a prefactor set to 10^{10} s^{-1}, a typical rate for the elementary step of adding a residue at the end of a helix. Use this equation to estimate the barrier to be crossed when CI2 folds.

Using the k_F value from part A of 42.52, which is equal to $D \times e^{-\Delta G^{\#}/RT}$. So, $\Delta G^{\#} = -RT \ln (42.52 / 10^{10}) = (-1.987)(298.15) \ln (4.25 \times 10^{-9}) = 11420$ cal/mol = 11.42 kcal/mol. For a simple 2-state folding model, this is a barrier that would be crossed, from the unfolded to the folded state, approximately once every 40 microseconds.

(D) What assumptions are made in the analysis you performed in parts A-C? How might these assumptions be tested or validated?

The analyses in parts A-C all assume in part or in whole that the protein folding event conforms to the 2-state folding hypothesis, i.e., that the folding reaction is well described as a simple equilibrium between unfolded and folded forms separated by a single simple transition state barrier.

There are several ways a folding reaction can violate 2-state behavior: (1) significant populations of intermediates accumulate during folding, (2) the reaction is not reversible, (3) oligomerization or aggregation can distort the analysis, (4) large, multi-domain proteins can fold in multiple stages and at different levels of organization, *etc.*

Tests for 2-state behavior include: (1) checking for consistent equilibrium folding results for different X-axis variables, for example temperature *vs.* denaturant unfolding, (2) similarly check for consistency for varied Y-axis properties, such as CD ellipticity *vs.* viscosity of the protein solution, (3) check for reversibility, for example by ramping temperature up and down repeatedly, (4) comparing enthalpy from equilibrium temperature unfolding to the enthalpy measured by differential scanning calorimetry (DSC), *etc.*

Question 5 (35 points total; individual point values given below)

On the following pages, provide a *brief* answer and discussion for each of the following questions:

(A) What is a disordered protein or protein region? (1 pt)

Anything vaguely correct; no fixed structure, many different states, in an ensemble

(B) When talking about disordered proteins, what is an ensemble? (2 pts)

A description that captures the idea that it's (1) a way to think about the 3D manifestation of an IDR and (2) describing many different conformations that are exchanging between one another; e.g. "A collection of interconverting conformations" would be full credit as it captures both ideas; basically it's impossible to describe the second idea without encapsulating the 1st.

(C) What features of a protein sequence might influence or determine if it is folded or disordered? Name two (2) features and briefly explain why each feature could influence if the protein is disordered or folded. (4 pts)

Two of the following at two points each:

(1) Low hydrophobicity; no folded core
(2) High charge; inability to fold/highly soluble/electrostatic repulsion
(3) Presence of secondary structure breaking residues; impede the formation of helices and sheets

(D) What are some challenges in studying disordered proteins experimentally? Name two challenges, explain the molecular origin of each challenge, and indicate what kinds of problems it causes. (5 pts)

Two of the following at 2 points each:

(1) Low solubility; hard to work with proteins in vitro, which makes many techniques hard to use
(2) No fixed structure; cannot rely on techniques that make use of constructive interference so obtain ensemble average measurements OR need to work with single-molecule experiments
(3) Some of the most powerful techniques used to study IDPs are technically really tricky; can make it difficult to obtain high-resolution information
(4) Degradation when expressed in e. coli; hard to get large-yields of protein

(5) Often require multiple techniques because no one technique gives enough information; requires expertise in many things

The extra +1 is for ensuring at least one of them has a mechanistic explanation for the issue and/or if additional detail worthy of +1 is present.

(E) How can charged residues influence conformational behavior in disordered regions? Name two effects and how they might influence IDR ensemble behavior. (4 pts)

Two of the following at two points each:

(1) Like charges repel; leads to expansion for IDRs with many residues of the same charge
(2) Opposite charges attract; leads to attractive interactions and/or local or long-range attraction
(3) Solvent loving, so can contribute to chain expansion because charged residues want to be solvent-exposed

There are probably additional answers that could be receive credit here, such as ion chelation, charge regulation, *etc*.

(F) What are the three main ways that a disordered region can bind to a partner? (3 pts)

Folding upon binding, forming a fuzzy complex, and fully disordered complex.

Note that if students misunderstand and name the types of partners (*e.g.* proteins, RNA, DNA, small-molecules lipids, *etc.*) they should receive 1 or 2 points out of 3 depending on how reasonable the suggested things are.

(G) Choosing one of the answers from part F, provide a molecular description of how this binding can occur? (2 pts)

Something that makes sense given the answer to (F); ideally, the answer here would be a description of one of the three modes. Note if (F) is wrong, then as long as what they say here makes sense, it can receive full credit.

(H) How can transient helices in a disordered region influence binding? (2 pts)

The answer should combine the following:

(1) Pre-organizing binding interface / pre-pay entropic penalty [or something to this general effect]

(2) Some consequence of the above (*e.g.* tighter binding, faster k_{on}, slower k_{off}, reduce transition state diffusion time by organizing the transition complex, *etc.*)

(I) What is conformational selection? (2 pts)

Mode of molecular interaction in which a macromolecule (*i.e.* a folded partner) binds to specific conformations (or set of conformations) in the disordered protein's ensemble. No penalty for not mentioning IDPs.

(J) Why might IDRs be sensitive to their solution environment? (2 pts)

A complete answer should contain some combination of: (1) large surface area, lots of exposed residues, (2) no fixed 3D structure holding the amino acids in a specific topology/exist in an ensemble, and (3) many residues that like to interact with the solvent.

(K) What is a Short Linear Motif (SLiM) and what do they do? (2 pts)

A stretch of 5-12 (as long as lower bound < 10 and upper bound between 10 and 20 this is fine, we say 5-12 but this definition varies quite a bit) residue elements that enable molecular recognition with a partner.

(L) What mechanisms allow IDRs to engage in specific molecular recognition in the absence of a 3D structure? (3 pts)

Some combination of the following:

(1) SLiMs
(2) Folding upon binding in a specific 3D structure
(3) Many different SLiMs together on a single IDR
(4) SLiMs surrounded by a specific chemical context (enabling chemical specificity)
(5) Different types of chemistries combined on the same IDR
(6) Working on concert with a folded domains; not really in the spirit of the question but 100% true, and maybe one of or the major mechanisms!

(M) How might post-translational modifications influence IDR conformational behavior? Feel free to choose whatever modifications you want. (3 pts)

Some combination of the following:

(1) Phosphorylation → Add negative charge → any reasonable consequence of that charge change

(2) Lysine acetylation → Remove positive charge → any reasonable consequence of that charge change

(3) Arginine methylation → add methyl to guanidinium group → any reasonable consequence of that in terms of sterics or altering pi-system (Note: This does not remove a charge. But if proposed along with a cogent explanation under the assumption only deduct 1 point)

(4) Ubiquitination → adds a small protein! → any reasonable consequence

(5) Any other PTM, physical description, and reasonable consequence that makes sense (even protonation/deprotonation of titratable sidechains, who's to say gain/loss of H is not a PTM)