Biology 5357

Chemistry & Physics of Biomolecules

Examination #1

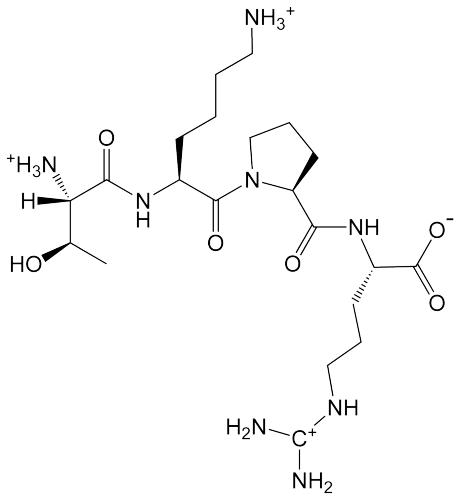
Proteins Module

October 1, 2021

Answer Key

Question 1

(A) (5 points)



3 points for correct amino acid structures/connectivity, 1 point for correct protonation states, and 1 point for having correct chirality at α-carbons

(B) (5 points)

Structure ***1*** on the left of the figure is the correct fold, as it has the canonical twist observed in essentially all β-sheet structures. Structure ***2*** is the enantiomeric, mirror image fold. (3 points)

The two commonly observed β-sheet packing motifs are "parallel" and "perpendicular". This jellyroll structure is an example of "parallel" packing as the strands on the green sheet are roughly parallel to strands on the purple sheet in the figure. (2 points)

Question 2

(A) (5 points)

A urea concentration of 3.5M is the midpoint of the equilibrium folding-unfolding curve, and at this concentration the free energy difference between folded and unfolded protein is zero. At the lower concentration of 1M urea, a higher proportion of the protein will be folded. At 1M urea, we have $\Delta\Delta G = 1.80 \times (3.5 - 1.0) = 4.5$ kcal/mol favoring the folded form.

(B) (5 points)

If only the $[\text{urea}]_{50\%}$ concentration increases with *m* remaining the same, then the difference in stability of the folding and unfolded forms of the protein will be greater, *i.e.*, $\Delta\Delta G$ will increase. (3 points)

Similarly, if the *m* value increases while the $[\text{urea}]_{50\%}$ concentration remains constant, then the $\Delta\Delta G$ value will be larger. (2 points)

(C) (5 points)

Assuming a 2-state folding model, the *m* value is a rough measure of the cooperativity of the protein folding process. Thus, the higher the m value, the greater the cooperativity. This cooperativity arises from accumulation of the large number of small interactions that typically favor folding. (2 points)

These interactions include formation of hydrogen bonds and salt bridges, dispersion from close packing, and burial of hydrophobic surface area in accord with the hydrophobic effect. In particular, *m* values often correlate well with hydrophobic surface burial in the folded protein. (3 points)

Question 3

(A) (5 points)

Metropolis Monte Carlo is often vastly more efficient than simple "random" Monte Carlo since it focuses sampling in the "important" regions of the potential energy surface. For example, the Metropolis method can be tuned to reduce time spent sampling the very high energy portions of the surface. For this reason, Metropolis Monte Carlo is often referred to as one type of "importance sampling".

(B) (5 points)

The Metropolis accept/reject criterion keeps a system at equilibrium provided it starts at equilibrium. Consider an equilibrium system of many molecules with each molecule in one of two states, A or B, where state A is lower in energy than B. According to the rules for Metropolis sampling, a Monte Carlo trial that move a molecule from the B to A state will always be accepted since A is lower in energy. What about trial moves from A to B, which are uphill in energy? If we reject all those moves, then eventually all the molecules will be in the A state. So, we much accept at least some uphill energy moves. Since A is lower in energy than B, there will be more A than B at equilibrium, so we don't want to accept all uphill A to B moves. But how many should we accept? The correct idea is to accept just enough uphill energy moves to maintain equilibrium between the number of molecules in the A state and the B state. Equilibrium is maintained if the "flux" in both directions is equal, *i.e.*, if the number of accepted A to B moves is equal to the number of accepted B to A moves.

Thus, we must have:

[probability of trying an B to A move]
× [probability of accepting an B to A move once we choose to make it]

equal to

[probability of trying an A to B move]
× [probability of accepting an A to B move once we choose to make it]

We want to solve this equation for the odds of accepting an uphill energy move:

[probability of accepting an A to B move once we choose to make it]

But from the rules for Metropolis sampling, downhill energy moves are always accepted, so

[probability of accepting an B to A move once we choose to make it] = 1

And we also know the ratio of trying a B to A vs. a A to B move, since this is just the equilibrium constant between the concentrations of A and B:

[probability of trying an B to A move] /[probability of trying an A to B move] = [B] / [A],

And the value of this equilibrium constant is just $e^{-\Delta E/kT}$, where ΔE is the energy of state A minus the energy of state B. Finally, this means that the probability of accepting an uphill trial move from A to B must also be $e^{-\Delta E/kT}$, and this exactly what the Metropolis rule, comparing $e^{-\Delta E/kT}$ to a random number on [0,1] generates!

(C) (5 points)

In a replica exchange protocol several molecular dynamics simulations are started in parallel at a series of different, closely spaced temperatures. The exact setup used depends on the system being studied, but something like 20 simulations spaced 5-10°C apart is typical. Then at points along the MD trajectories, attempts are made to "swap" the temperatures of simulations adjacent in temperature. Acceptance or rejection of attempted temperature swaps is decided via the Metropolis Monte Carlo criterion applied to the total (potential plus kinetic) energy. Replica exchange "enhances" sampling at the lower simulation temperatures since the broader sampling done at the higher temperatures filters down to lower temperatures via successful temperature swaps. Use of the Metropolis criterion ensures maintenance of "detailed balance" which guarantees that the final sampling at the lower temperatures is "correct" in the statistical mechanical sense.

Question 4 (15 points; A-C, 5 points each)

(A) Given that $\Delta G = -RT \ln(K_d)$, then at T = 25°C = 298K we have:

$$\Delta G = -(1.9872 \text{ cal/mol/K}) \times (298\text{K}) \times \ln(1\times10^{-8}) = -10.9 \text{ kcal/mol}$$

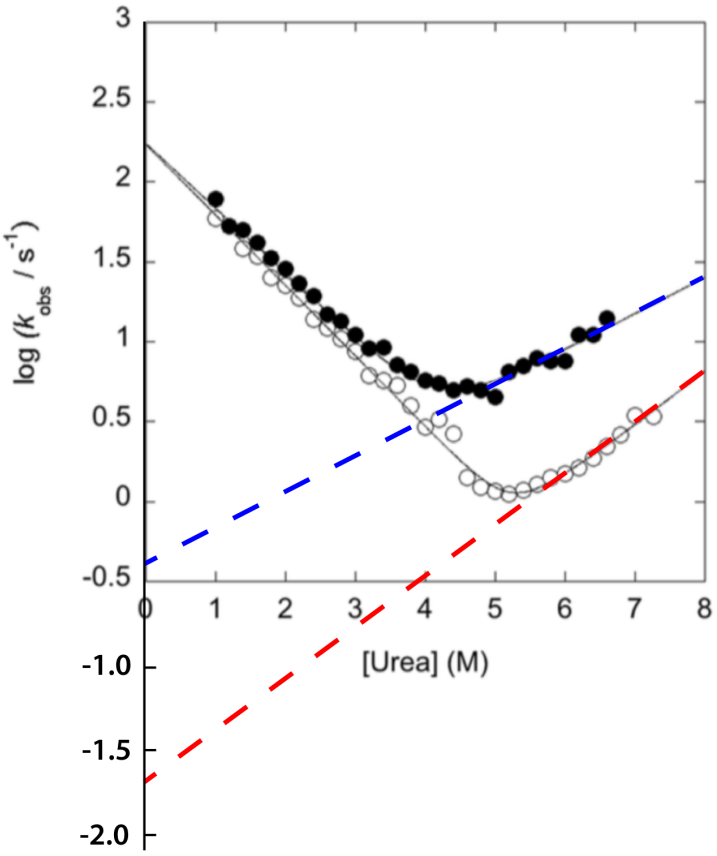
Thus, the drug (D) has a binding energy to the protein (P) of 10.9 kcal/mol, and this binding energy is a free energy (ΔG).

(B) The overall binding energy of a ligand or drug is determined by how strongly it interacts with a protein minus how tightly it "binds" or interacts with water. Adding a hydrogen bonding group to a ligand can provide an extra interaction with its protein target. But in most cases the added hydrogen bonding group can also interact with water, and the extra interaction in the bound and unbound states will approximately cancel. On the other hand, adding an additional hydrophobic group to a ligand, which is removed from exposure to water when inside a protein binding pocket, will clearly favor ligand binding.

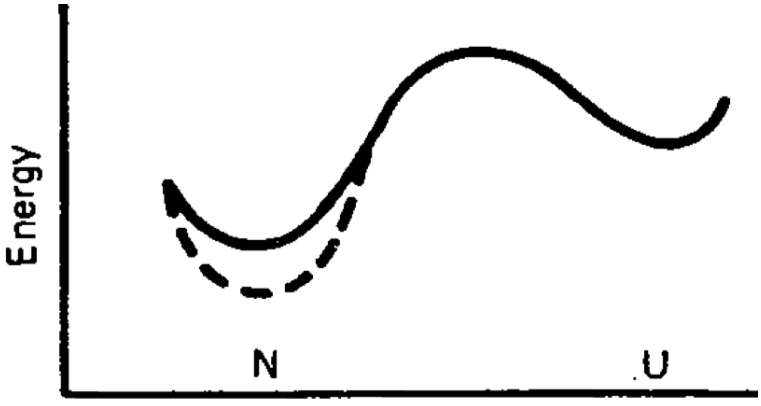
(C) When a drug or ligand make a favorable enthalpic interaction with its target or receptor, such as a hydrogen bond or burial of hydrophobic surface area, this often restricts the location and/or motion of the ligand. This restriction of conformational or configurational freedom is unfavorable entropically. The usual explanation of the hydrophobic effect itself is an example of enthalpy-entropy compensation. The exposure of a hydrocarbon to water is unfavorable by a fairly constant free energy over a broad temperature range. At lower temperature, the unfavorable ΔG is largely enthalpic in origin, while at higher temperature it is mostly entropic.

Question 5 (15 points; A=10 points, B=5 points)

(A) From the data in the chevron plot, the folding rates for both the full-length PDZ domain and the $\Delta\alpha3$ truncated domain (solid black line extrapolated to 0M urea) are very similar, with a value close to $10^{2.2} = 158$. The extrapolated unfolding rate for the full PDZ domain (dashed red line) is $10^{-1.65} = 0.0224$, while the unfolding rate of the $\Delta\alpha3$ form (dashed blue line) is $10^{-0.4} = 0.398$. The equilibrium constant between folded and unfolded of full PDZ is K = 158 / 0.0224 = 7054. Then $\Delta G = -RT \ln(K) = -1.9872 \times 298 \times \ln(7054) = 5.25$ kcal/mol, in favor of the folded form. The equilibrium constant between folded and unfolded of the $\Delta\alpha3$ truncated domain is K = 158 / 0.398 = 397, and $\Delta G = -RT \ln(K) = 3.54$ kcal/mol, again favoring the folded form. So, the difference in folded-unfolded stability between full PDZ and the $\Delta\alpha3$ form is 5.25 – 3.54 = 1.71 kcal/mol greater relative stability for full-length PDZ.



(B) We will assume a simple 2-state model for PDZ domain folding. The 2-state model is not contradicted by the linearity of both folding and unfolding legs of the chevron plot for both full PDZ and the Δα3 truncated protein. The folding rates for the full and Δα3 forms are identical, and the difference between the two is in the slower unfolding rate for the full PDZ form. This suggests that the collection of unfolded states and the transition states for both the full and Δα3 forms are similar, but the folded PDZ domain is lower in energy than the folded/native Δα3 domain. A reaction coordinate diagram showing conversion of folded/native (N) and unfolded (U) states is shown below for the full PDZ domain (dashed line) and the Δα3 truncated form (solid line):



Question 6 (10 points; A-B, 5 points each)

(A) For full credit of 5 points, provide three or more of:

- Charged residues
- High net charge
- Depletion for hydrophobic residues
- Presence of polar residues
- Enrichment for glycine or proline
- Presence of hydrophilic residues
- Residues that can't form secondary structure
- Residues that prefer to interact with solvent
- Lack of disulfide bonds

(B) For full credit of 5 points, provide three or more of:

- No fixed 3D structure so cannot use crystallography
- No folded structure so cannot examine protein folding/unfolding
- Most useful techniques (NMR/SAXS/smFRET) are hard to do
- Single point mutations often have little effect
- Aggregation prone
- Low signal-to-noise ratio
- Hard to express

Question 7 (5 points)

For full credit of 5 point, mention at least two of the following:

- MC – good for building big ensembles
- MC – needs only energy calculation, does not require forces
- MC – can be hard to find efficient trial move sets
- MD – allows you to get dynamics (which MC does not)
- MD – can have forcefield issues

Question 8 (5 points)

For full credit of 5 points, must mention the following points:

- Each peak reports on a single residue/backbone amide
- The peaks provide information on local chemical environment
- Phosphorylation changes the chemical environment, such that residues near the phosphorites have altered chemical environments

Question 9 (10 points; A-B, 5 points each)

(A) This is something of an open-ended question that is not itself clearly answered, so credit is given for any "reasonable" discussion or ideas

(B) For full credit of 5 points, mention any two of the following:

- Flexible linker between folded domains
- Defining protein organization
- Control local concentration of motifs and domains
- House short linear motifs
- Molecular recognition via folding upon binding
- Multivalent hub proteins
- Forming biomolecular condensates
- Binding many different partners
- Improving solubility/prevent aggregation