Problem Set Answer Key

a) Draw and label the complete structures of the cis and trans prolyl-isomers that may be found at pH 7 for the tripeptide:

Ala – Pro – Ser

must be drawn correctly, including proper charges on the termini. Your drawing must make clear that the rotation of the peptide bond preceding the proline residue is the difference between the cis and trans isomers, with the alpha carbons of alanine and proline being opposite each other across the peptide bond in the trans form and on the same side in the cis form.

b) (6 pts) Given a polypeptide of 100 amino acids in length, how long would the folded conformation be if:

i) (2 pts) Folded into an \( \alpha \)-helix, as found in globular proteins?

As stated in the legend of Figure 2.2 from B&T, there are 3.6 residues per turn and 5.4Å per turn (3.69 res/turn; 5.44Å/turn according to Pauling & Corey paper). Therefore:

\[
\text{Rise/residue} = \frac{5.4\text{Å/turn}}{3.6\text{ residues/turn}} = 1.5\text{Å/res} \\
1.50\text{Å/res} \times 100 \text{ res} = 150\text{Å}
\]

OR

\[
\text{Rise/residue} = \frac{5.44\text{Å/turn}}{3.69 \text{ residues/turn}} = 1.47\text{Å/res} \\
1.47\text{Å/res} \times 100 \text{ res} = 147\text{Å}
\]

ii) (2 pts) Folded into a collagen triple helix?

As stated on page 285 of B&T, collagen has a rise per residue of 2.9Å along the helix axis. Prof. King gave a more precise value of 2.86Å in class. Therefore:

\[
\text{Rise/residue} = 2.86\text{Å/res or 2.9Å/res}
\]
2.86Å/res * 100 res = 286Å  OR  2.9Å/res * 100 res = 290Å
If your interpretation was that the entire triple helix was composed of just 100 residues, credit was given for 290/3 = 96.7Å or 286/3 = 95.3Å.

iii) (2 pts) Folded into a β-strand?

\[(3.5 \, \text{Å}/\text{residues}) \times 100 \, \text{res} = 350\, \text{Å}\]

c) (5 pts) What conformation would you predict for the repeated sequence \((\text{Ala-Pro-Ser})_n\) in solution?
   i) An α-helix?
   ii) Collagen-like triple helix?
   iii) Coiled-coil?
   iv) Other?

The answer is iv) other. This sequence would likely exist as random coil or a series of loops/turns. Due to abundant prolines, this sequence would not form an α-helix, which also rules out a coiled-coil. The triplet repeat is reminiscent of the collagen repeat, but a collagen-like triple helix critically relies on the presence of glycines every third residue (in at least the majority of the repeats) because there is no room to accommodate a side chain within the helix. Partial credit was given for choice ii) provided you explained your reasoning.

d) (5 pts) Which of the above conformations is the following sequence most likely to take in solution, under physiological conditions?

RMKQLEDKVEELSKNYHLENEVARLKKLVGER

Coiled Coil. The above sequence is the amino acid sequence of GCN4, as written in the O’Shea paper handed out for homework. A subsequence of this one is also given in Figure 3.2 of B&T. The point of this exercise, however, is not to test your memory but to give you practice looking through amino acid sequences and recognizing patterns. This sequence matches the heptad repeat pattern, with the bolded leucines taking up the ‘d’ position of the heptad repeat.

2) (20 pts)
   a) (5) Which amino acids are found most frequently at the ends of helical conformations in globular proteins?
b) (5) Do the residue types differ between the N-termini and the C-termini?
These two questions can be considered together, and came primarily from the Richardson/Richardson and Presta/Rose readings. The most important point is the importance of hydrogen bonding patterns in determining the ends of helices. In general, polar amino acids are favored at the ends of helices because their side chains H-bond with the backbone carbonyls and amide hydrogens, which terminates the secondary structure. Charged residues (E, D, H, K, R) tend to be asymmetrically distributed, with negative charges near the N-term and positive charges near the C-term, countering the helical dipole. Pro and Gly are often found at the caps – proline because of its backbone geometry and the inability of its amide to H-bond, and Gly (especially at the C-terminal cap) because its flexibility and lack of a side chain allows it to satisfy two successive carbonyls at the helix terminus (one with each amide flanking Gly).

c) (5) How does the composition of the α-helix capping residues compare to the residues found at the helix/helix docking sites in globular proteins?
Helix/helix docking sites generally do not involve the ends of helices, whose composition was explained above (Gly, Pro, and polar residues). Docking sites are enriched for hydrophobic residues (A, I, L, V, F), and generally do not contain Pro or Gly since these residues are not favored in helices.

d) (5) What was Marquesee and Baldwin’s strategy for designing amino acid sequences that would fold into soluble monomeric α-helices in aqueous solution?
They wanted to understand intrahelical electrostatic interactions that stabilized helicity. Starting with poly-alanine, they inserted charged residues at regular spacings of either three or four residues. Thus, in a helix (3.7 residues/turn), each positively charged residue could ion pair with a negatively charged residue three or four amino acids away. The pairs were positioned along the helix so as to disrupt large hydrophobic surfaces that could cause association of the chains. The formal charges at the ends were blocked to avoid complications from their charges and also to lessen the helix dipole. They analyzed helicity with circular dichroism while varying temperature, pH, and peptide concentration. They ultimately determined that salt bridges separated by 4 residues
were helix stabilizing, with additional stability gained if the dipole was also stabilized.

3) (20 pts) Carbonic anhydrase is an enzyme lacking both disulfide bonds and heme groups. The following experimental observations were made:

a. Upon equilibrium denaturation with urea, the protein showed the following changes in tryptophan fluorescence and circular dichroism at 222 nm with varying concentrations of urea.

![Graph showing changes in fluorescence and circular dichroism with urea concentration.]

b. 6M urea-denatured proteins were diluted out of denaturant with buffer. The native tryptophan fluorescence was monitored with time following dilution.
i) (10 pts) Provide a model folding pathway to explain these results and describe how these pieces of evidence support your pathway.

The folding pathway we were seeking is \([U] \rightarrow [I] \rightarrow N\), where the unfolded protein proceeds through a partially folded intermediate before reaching the native state. The existence of a partially folded intermediate is supported by the non-coincidence of the CD and fluorescence signals in (a). This folding intermediate appears to retain its secondary structure to a greater degree than it retains the buried nature of its tryptophans, as seen at intermediate concentrations of denaturant. The non-coincidence of signals rules out an alternative folding pathway, namely the 2 state folding model of \([U] \rightarrow N\). Note that this equilibrium experiment does not give kinetic information, it merely gives a generalized structure to the intermediate.

The kinetic curve in (b) shows two kinetic phases. The first phase, occurring before the kink at approximately 10 seconds, could represent a fast \([U] \rightarrow [I]\) transition in the folding pathway, while the second slow phase could represent a slower \([I] \rightarrow N\) transition. Alternatively, this data is consistent with the presence of two populations of chains in the unfolded state. An example of such a phenomena can be found with proline isomerization, where one species representing the native isomer folds more rapidly to the native state and the other more slowly, since it must first convert to the native isomer.

ii) (10 pts) If you wished to estimate the fraction of the peptide bonds that are hydrogen bonded in a native-like conformation at any point along the folding pathway, which of the two signals would be more informative? Explain.
The CD signal would be the more informative signal, as opposed to the fluorescence signal. As has been described in class, the fluorescence signal gives information about the local environment primarily around tryptophan residues. Hence, this signal is often interpreted as how buried the tryptophan residues are along the folding pathway. This does not provide any information about the hydrogen bonding of the protein’s peptide bonds. The CD signal, however, is an average signal reflecting the secondary structure of the protein. Since secondary structure is intimately dependent on the hydrogen bonding state of the protein, this would be an appropriate signal for estimating the fraction of peptide bonds that are hydrogen bonded. In essence, the fraction of peptide bonds hydrogen bonded would be directly proportional to the amount of native-like secondary structure realized at any point on the folding pathway.

4) (20 pts).

   a) (5 pts). Collagens from different species of vertebrates have varying numbers of proline and hydroxyproline residues in their sequence. How do the thermostabilities of these collagens vary with their composition of proline and hydroxyproline?

   b) (15 pts). Mammalian procollagen molecules can be isolated so as to maintain their C-terminal registration peptides in disulfide bonded form (for example from fibroblasts in tissue culture). After urea denaturation (without disulfide bond reduction), dilution to physiological buffers results in the recovery of native-like collagen conformers, as opposed to gelatin. The yields are good but the rates are much slower than found for the globular proteins discussed in class. Propose a folding model that accounts for these observations and describe how it accounts for the following:

   i. Collagen is recovered, rather than gelatin.
ii. Folding rates are much slower than for globular or coiled coil proteins

i) The presence of the disulfide bonded, C-terminal registration region is sufficient to maintain collagen chains in the correct registry with each other. This correct registry or alignment allows for the triple-helices to form while avoiding incorrect pairings or alignments that lead to the gelatin state.

ii) Folding rates might be slower than for globular or coiled coil proteins for a number of reasons. One reason is the requirement for nucleation to initiate triple helix formation, generating a lag time in forming the triple helix. Triple helix propagation is dependent on the formation of an unstable structural nucleus that is kinetically difficult to form.

A second factor slowing folding is that denaturation allows for proline isomerization to occur. In order to properly fold back to its native conformation, all of the non-native isomers would have to undergo rearrangement without new non-native isomers forming. Given the large number of prolines and hydroxyprolines in the collagen sequence, this could be a very long process that accounts for the slow folding rate.