Problem Set #3 – Chemistry 391

Due September 20th in class.

Modeling: Part I

Luedtke et al. discuss the folding and assembly of GCN4, a dimeric "leucine zipper" protein. The following exercise will provide some background to the protein and also illuminate some issues from the paper.

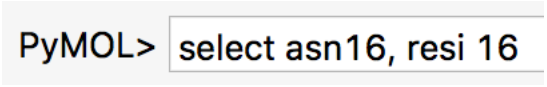
Download **gcn4.pse** from the Chem391 Moodle page, and open in *PyMOL*.

The initial image shows ribbon "cartoons" of two helices (colored yellow and pink) that comprise the dimeric protein. In addition I have highlighted the N-terminal acetyl groups (not natural, but useful synthetically) in light blue to orient the helices.

The following will help you answer the relevant question in the problem set.

Q1a. What is the heptad repeat in the helix? To get a visual select **sticks** under the **S** button next to **aface**. The residues of the yellow chain (chain A) that interact with chain B will be shown. To ID their sequence positions, select **residue** under the **L** button next to **aface**.

Q1b. Why is OK to have Asn16 as part of the heptad repeat? In the command box type:



And then select **sticks** under the **S** button next to Asn16. Looking at the residues from both chains, explain why they might be allowed to persist in an otherwise non-polar, buried position. To measure the distance between relevant atoms use the **measurement** tool under the **wizard** menu (top of screen) and left click on the relevant atoms. You can **delete all measurements** when **done**.

Q1c. Why does proline disrupt secondary structure? Select **everything** under the **H** button next to **all.** Then select **sticks** under **S** next to **achain**. You might want to label the residues of **aface** again. To change Leu19 (the sequence number in our model) select the **mutagenesis** tool under the **wizard** menu and then select **PRO** by clicking on the menu headed "**No Mutation"** on the right panel. You will see some warning images projected over the model. Take note and click **apply** to seal the deal with proline. Look at the impact of having a side chain bond to the backbone nitrogen. Why isn't proline tolerated well in α-helices?

Problem Set #3 – Chem 391 Name_____
Due in class on 9/20/18

1. GCN4 possesses an amphiphilic helix with the sequence:
RMKQLEDKVEELLSKNYHLENEVARLKKLVG

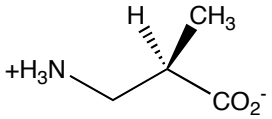
a. The sequence contains a heptad repeat of non-polar residues that generates an amphiphilic helix, with the exception of Asn16 (the number 16 refers to the position in the sequence). Underline the residues that contribute to a non-polar face of the helix and circle Asn16 in the sequence above.

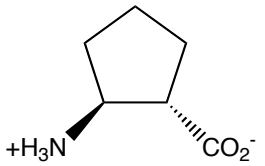
b. Examine the position of Asn16 and the interactions it makes. Why might it be tolerated on the otherwise non-polar face of the helix? Words and/or a drawing can be used.

c. The L20P mutant converts Leu20 to proline to disrupt protein folding. Why does proline generally disrupt the structure of α-helices? (In our PyMOL model, the numbering is Leu19).

d. In this study, the structure of GCN4 is altered so that the N-terminal residues are CCGGQLED. The "CC" is for labeling purposes. Why would the authors use GG instead of the native residues at that position? (Google "glycine Ramachandran plot" if you're unsure and compare to the alanine plot from last week).

2. The secondary structure of polypeptides composed of β-amino acids, rather than the standard α-amino acids, have been explored. Two possible building blocks are shown. The compound on the left does not lead to stable conformations, but these folks were successful in creating stable secondary structure from polymers of the cyclic compound on the right. Given the apparent analogy of these two residues to alanine (on the left) and proline (on the right), this is perhaps surprising. Perhaps not. Explain the results given your understanding of secondary structure stabilization.





4. Collagen comprises about 25% of the protein in the human body (see http://pdb101.rcsb.org/motm/4 for background). It is a fibrous protein comprised of a repeating sequence that is chiefly Gly-Pro-Hpr, where Hpr is hydroxyproline, a modified form of proline with a hydroxyl group. Collagen exists as an extended "triple helix" of three interwoven strands interacting via vdW forces and H-bonds. The following questions rely on the PyMOL session file **collagen.pse**, available on the 391 assignments web page.

a. A three residue triplet (Gly-Pro-Hyp) on chain A is highlighted in pink (**zoom** in on the object "**triplet**"). Identify two hydrogen bonds made by the triplet to two other strands in the triple helix and note here the donor and acceptor atoms (e.g. donor = Gly NH, chain A/acceptor = Gly C=O, chain B).

b. Why is glycine present in most triplets? Zoom in on the "triplet" object. Mutate Gly18 of strand A (it's pink) to Ala using the Mutagenesis wizard. Explain why other amino acids are not suited to this position.

c. Why is hydroxyproline present in many triplets? The Raines lab at Wisconsin found that Hyp stabilizes collagen by about 0.6 kcal/mol per residue over proline (Holmgren et al., 1999; available via link on assignments page). Please look over that paper and provide a quick pictoral and/or written summary of why they believe that to be the case and how they support their case with fluoroproline. Figure 6 may be most useful to you.

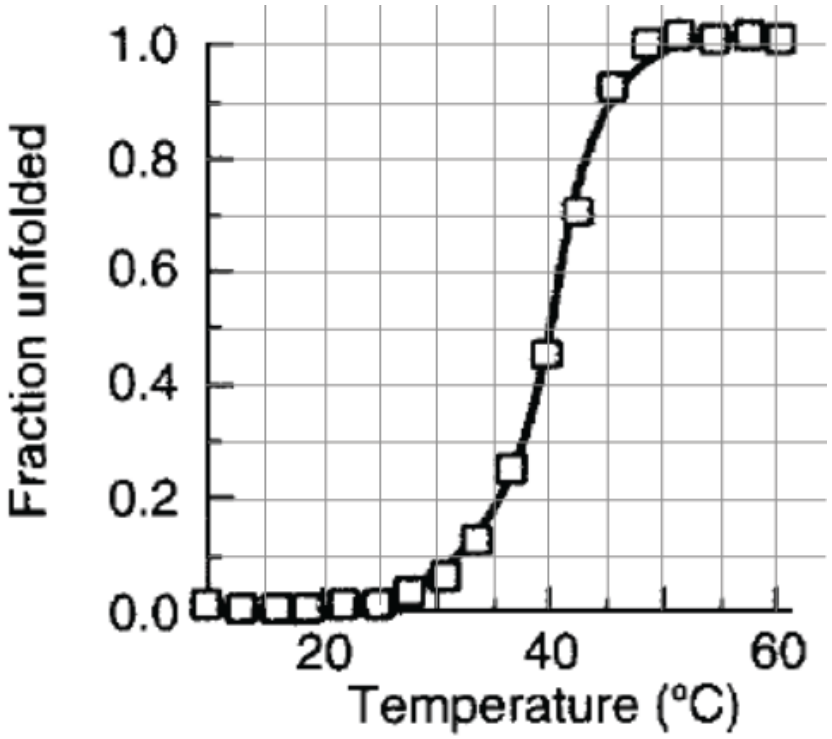
d. A recent paper from the Chenoweth lab (Zhang et al., 2015) suggests that inclusion of a novel substitute for glycine (azaglycine) can increase the stability of collagen by 1-3 kcal/mol per substitution. To explore why, zoom in on the "azatrip" object. The new nitrogen of azaglycine would be where the magenta atom is.

i. Provide an enthalpic argument for collagen stabilization by azaglycine (drawing may help).

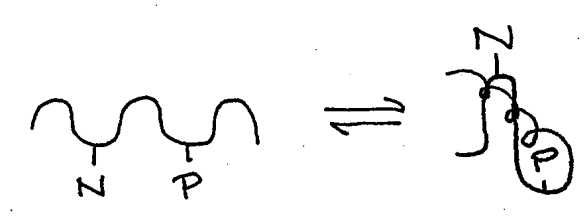
ii. Draw a tripeptide Ala-azaGly-Ala. Provide an entropic argument for collagen stabilization.

5. In the Raines paper (question 4), a thermal melting curve is presented for collagen containing a novel proline mimic (Figure 3e, modified below). Use van't Hoff analysis to calculate ΔH° and ΔS° for the melting of this novel collagen triple helix. Fill in the table below (making clear how you got K_{unfold} at one T value). Also attach a plot of ln(K) vs. 1/T (in K^{-1}) that gave you the necessary information to get ΔH° and ΔS°.

T (K)	K_{unfold}
306	
310	
313	
315	
319	



6. <u>In one or two short sentences</u>, (referencing the following diagram) and using terms like "enthalpy" and "entropy" explain...



a. Why it **is** thermodynamically unfavorable to locate a polar residue (P) in the core of a protein...

b. ...and why it **is not** thermodynamically unfavorable to locate a non-polar residue (N) on the surface of a protein.

Paper of the week – for discussion

Luedtke et al. (2007) Surveying polypeptide and protein domain conformation and association with FlAsH and ReAsH. *Nat. Chem. Biol.* **3**, 779-784

1. What is the hypothesis/motivating question/technical goal of this paper?

2. What is going on in the "biarsenical dye" labeling? What are the structural requirements for a simple tag on a single hexapeptide peptide and what is the strategy for transferring that structure to aPP and GCN4?

3. Consider the data in Figure 1. K_{app} is and equilibrium constant, and helpfully, the concentration of protein necessary to bind 50% of the tag.

a. Why did they select those positive and negative controls and why do the results support that the controls are working?

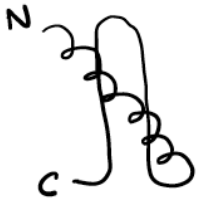
b. Refer to question 1c in the regular PS for why proline is chosen to disrupt protein structure.

i. Do the data in Figure 1c indicate success?

ii. How do the results in Figure 3 further substantiate that choice?

iii. How about Figure 4? How should one interpret those plots?

4. eGFP is a green fluorescent protein. Let's say it looks like the image at the right (not even close, but good for what I need). Sketch how the eGFP-aPP and eGFP-GCN4 fusions are intended to function to bind ReAsH.



5. Focus on the results related to GCN4 in Figure 5a and 5d. How do the gel results and the cell images support (or not support) the use of this methodology in assessing in vivo protein folding?

6. Now, the final points.

a. Why is this fluorescent technique capable of yielding superior structural resolution?

b. How might this technique facilitate screening of small molecules that stabilize specific protein folds