1. GCN4 possesses an amphiphilic helix with the sequence:

RMKQLEDKVEELLSKNYHLENEVARLKKLVG

.1234567123456712345671234567 1.heptad!

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.

Removing the polar amide sidechain of Asn from water and placing it in a hydrophobic region could strip it of H-bonds, which would be enthalpically costly to folding. However, the cross-chain H-bond that forms between Asn residues compensates, at least in part, for the loss of H-bond to water.

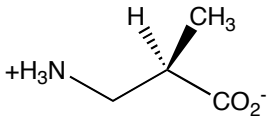
c. The L20P mutant converts Leu20 to proline to disrupt protein folding. Why does proline generally disrupt the structure of α-helices? (In our PyMOL model, the numbering is Leu19).

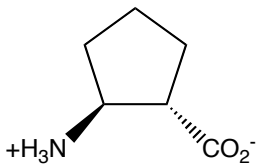
Helices maintain a series of internal H-bonds in which the C=O of the nth residue H-bonds to the HN of the n+4th. Proline removes the NH group and replaces it with N-CH_2. The methylene group can't H-bond with the C=O and, worse, it is sterically bulky and will clash with a C=O group.

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).

Glycine is more flexible than Ala (no substitution). Two of them may be needed to provide the needed conformational flexibility to interact with FlAsH/ReAsH in a product way.

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.





Although the five-membered ring of proline impedes H-bonding in α-amino acids, the five-membered ring shown above would result in peptides that have primary amides, with N-H groups that can act as donors. Although the alanine-analog looks like a reasonable amino acid, it suffers from having 3 rotatable bonds along the backbone, where α-amino acids only have two (ϕ and ψ). As a result, the proline analog is a suitable modification. The ring restricts the conformation flexibility of the backbone of the β-amino acid, making the adoption of secondary structure less entropically costly.

3. 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).

Pro19 C=O•••HN Gly81/C chain
Gly18 N-H•••O=C Pro48/B chain

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.

The methyl group collides with C=O groups on residues 78 & 79 on the C chain

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.

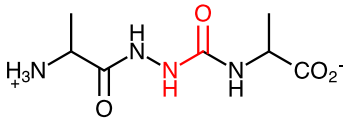
The hypothesis, supported by the result with F-Pro is that the hydroxyl (or fluoro) group provides a dipole that is anti-parallel to the dipole created by the C=O group on the same residue.

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).

Could H-bond to C=O on residues 66, 67 on C chain

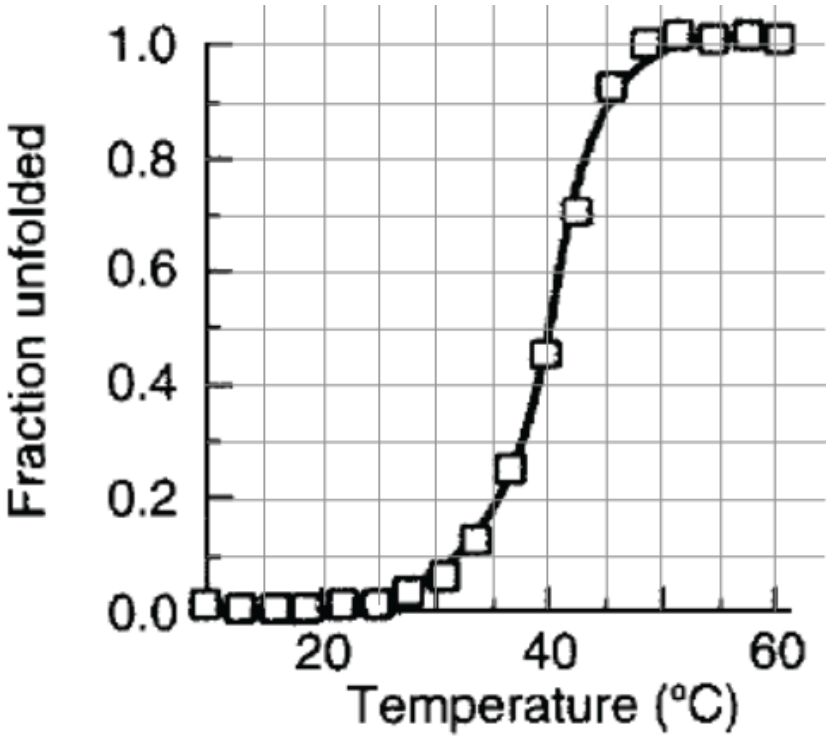
ii. Draw a tripeptide Ala-azaGly-Ala. Provide an entropic argument for collagen stabilization.



A new amide bond makes backbone more rigid, so it will lose less entropy on triple helix formation

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 (°C)	K_{unfold}
33	0.14
37	0.35
40	0.75
42	2.33
46	13.3

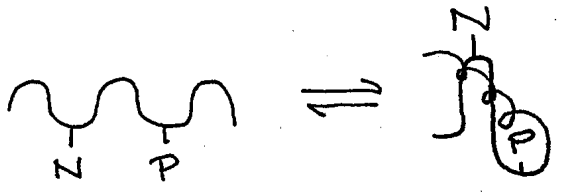


At 33°C
$K_{unold} = (f_{unfold})/(f_{fold})$
$K_{unfold} = 0.12/0.88 = 0.14$

Plot of ln(K) vs. 1/T gives
slope = -34400 K^{-1}
intercept = 110

ΔH° = +68 kcal/mol
ΔS° = 220 cal/molK
(see next page for details)

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



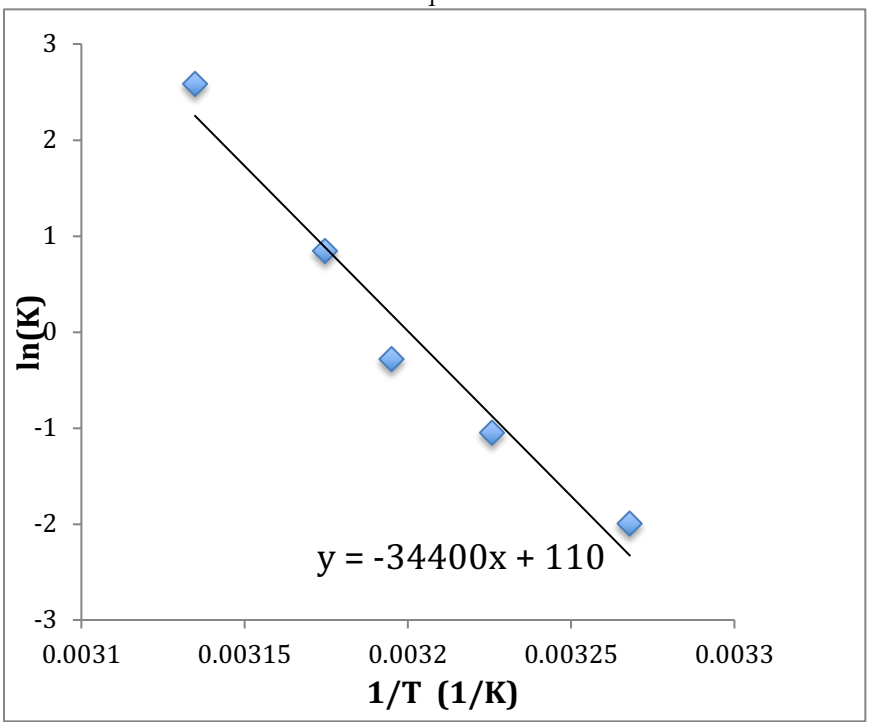
a. Why it **is** thermodynamically unfavorable to locate a polar residue (P) in the core of a protein...

The polar residue likely makes H-bonding and other strong electrostatic interactions with aqueous solvent, but those would be lost in the non-polar environment of the protein core ($\Delta H > 0$).

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

The non-polar residue is sees no change in its environment, so no change in interactions, so no change in free energy. Zero is not unfavorable (though one can imagine ways to do better than zero).

5. Plot of data from table in question 5



Note that ΔH = -slope * R = (-34400 K^{-1})(0.001987 kcal/molK) = 67 kcal/mol
ΔS = intercept * R = 110(1.987 cal/molK) = 220 cal/molK

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?

To us, this sensitivity suggested that the tetracysteine motif could function as a reporter of protein conformation or protein-protein binding if displayed in a bipartite mode, with two Cys-Cys pairs located distal in primary sequence (or on separate protein strands) but proximal in the folded or assembled state. If so, then bulky FPs used in cellular fluorescence assays and FRET-based sensors could be replaced by just two Cys-Cys pairs that, when assembled into close proximity (about 7 A °), bind to FlAsH or ReAsH and form a fluorescent complex.

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?

Hexapeptide: CCPGCC (the intervening PG sequence appears important for binding and fluorescence). Instead of separating the CC dipeptides with PG, the authors plan to insert a larger stretch of folded peptide (aPP or Zip4) or put CC residues at the N-terminus of two subunits of a dimeric protein (Jun 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?

Positive has CCPGCC sequence and activates fluorescence at a low peptide cont. The negative control is negative indeed (8 prolines between CC groups) and shows no measurable activation of fluorescence.

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?

Proline containing peptides have higher Kapp's than the native sequences, indicating that the bind FlAsH, ReAsH less well.

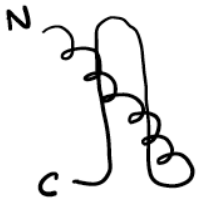
ii. How do the results in Figure 3 further substantiate that choice?

CD shows a distinct loss in secondary structure for those peptides containing Proline.

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

Generally, proline-disrupted peptides produce less fluorescence than the native proteins and in most cases appear not to saturate the dye at sufficiently low concentrations.

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?
5a – we see that the gel containing eGFP constructs all fluoresce with 488 nm excitation, indicating that eGFP is functioning as suspected. Only lanes containing GCN4 fusions excite at 532 nm once mixed with ReAsh, though the strength of the band with GCN4^{L20P} is surprising (though they say it's actually weak.)

5d – These images are more impressive. The green of eGFP shows up consistently, but the red of ReAsH only shows up when a properly folded version of GCN4 (no L20P mutant) is present.

6. Now, the final points.
a. Why is this fluorescent technique capable of yielding superior structural resolution?

The requirements of FlAsH and ReAsH are that the cysteines need to be close enough to bind the As atoms, but not too close, putting a strong structural constraint on the ability of the protein to react and fluoresce. FRET gives a signal over a broader range, making results less clear.

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

Add CCGG ends to protein.
Unstabilized protein should give little reactivity wish arsenical dyes.
Screen many small molecules – look for one that reacts with protein to give fluorescence.