## Problem Set #4 – Solutions

1. Does the NMR model of the switch mutant describe an alpha helix or $3_{10}$ helix? Explain briefly. A drawing would be permitted.

The carbonyl oxygen of Leu11 makes an H-bond to the amide nitrogen on the backbone of Trp 14. That is an n → n+3 H-bond, which means this is a $3_{10}$ helix.

2. On the NMR data from Cordes et al. (2003). See Table I.

a. Was enough data collected to prepare a high-resolution model for this 53-residue protein? Explain briefly.

There are 816 distance restraints and 31 dihedral restrints, giving 847 total for 53 residues. That's a 16:1 ratio, which is certainly very good.

b. Do the calculated models fit the data that were collected? What data are you using to reach that conclusion?

Pretty well. The RMSD from NOE distance restraints is 0.04 Å (much less than a bond length) and from angles is 0.09° (amazingly little).

c. Is there good structural agreement between the models? Again, indicate the relevant data.

Surprisingly, not as good as I might have expected given the above. They see 2.0 Å RMSDs between the 13 best models, but if they just look at residues 7-51, the numbers are better, with 0.56 Å RMSD for the backbone atoms – which is pretty good, though not excellent.

d. Do the models have appropriate stereochemistry? That is, do they possess structural parameters consistent with generally understood principles of protein structure? Once again, what entries in the table support your answer?

They did well on the Ramachandran plot with 88.9% in the most favored region and the other 11.1% in allowed regions. No outliers is great. The RMSDs for bonds is 0.003 Å (wow) and 0.51°, which is also quite good.

**Modeling: Part II**

Not all models are created equal. Skepticism should reign. Consider the crystallographic model of a $Zn^{2+}$ containing enzyme with a bound tripeptide of leucine. Download the file **LLL.pse** from the Moodle page. The paper that reports this model focuses on the conformation of the bound peptide. To view it more carefully, select **zoom** under the **A** button next to peptide.

**3**. There are four "obvious" stereochemical problems with the conformation of the tripeptide modeled here. Your mission, should you choose to accept it is to:

a. Find two inappropriate dihedral angles in the backbone (there are three that I found).
Worst is non-planar peptide between 302 and 303
Eclipsing carbonyl carbons between residue 301 and 302 (bad φ angle on residue 2)
Eclipsing carbonyl carbons between residue 302 and 3033 (bad φ angle on residue 3)
One could also argue that eclipsing amine and carbonyl oxygen on first residue is problematic but interactions with metal could cause that to happen.

b. Find one bond angle in the backbone that is clearly not appropriate for the hybridization state (I found two).
The worst is 154.3° for the C-N-CA bond angle between residue 1 and 2. 120° expected.
Second worst is CA-C-N angle that precedes the earlier one: 107 vs. 120 expected.

If you aren't having luck with your eyeballs, you can type in the **get_angle** command. The LLL tripeptide is numbered 301-303. Thus, to get the bond angle formed by N, CA and C of the first leucine, type "**get_angle 301/N,301/CA,301/C".** The value of the bond angle will appear above the command line. Also, if you'd like to see whether the electron density calculated from the x-ray data support this odd conformation, click on the **density** button and it will show the density specific to the tripeptide.

4. In a ground-breaking paper, Kleywegt and Jones (*Structure*, **3**, 535-540, 1995) published two refinements of models using crystallographic data from the cellular retinoic acid binding protein. In one case, they reversed the protein sequence (yes, they did something so wild no one in their right mind would ever repeat it) and placed the resulting unnatural peptide chain in the electron density map. In separate trial, they took the correct sequence and placed it in the map and refined it. The following statistics resulted (I'm not specifying which refinement effort led to model X or Y). Questions next page.

| **Model** | **X** | **Y** |
|---|---|---|
| Resolution range (Å) | 8.0–3.0 | 8.0–2.9 |
| R-factor | 0.214 | 0.251 |
| Rmsd bond lengths (Å) | 0.009 | 0.009 |
| Rmsd bond angles (°) | 2.1 | 1.6 |
| Ramachandran plot, most favored areas (%) | 42.7 | 81.6 |
| Additional allowed areas (%) | 36.3 | 16.0 |
| Generously allowed areas (%) | 12.1 | 1.6 |
| Disallowed areas (%) | 8.9 | 0.8 |

4. Kleywegt question.

a. Which model is in better agreement with the data? Why?

Model X is in closer agreement – the R-factor measures the deviation of calculated amplitudes (taken from the model) from the observed amplitudes (the actual data). Lower values indicate better fit.

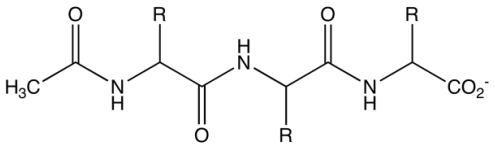
b. Which model has better stereochemical and conformational parameters? Explain your answer explicitly.

In two ways, model Y is superior. It has a lower root mean square deviation in bond angles (using a library of model compounds as a reference) and shows better agreement with the allowed regions of the Ramachandran

c. Which is the wrong structure? Explain why you think so.

X is wrong. The tip-off is the lousy fit to the Ramachandran plot (wholly unacceptable statistics). Essentially, the model could only fit the data by completely ignoring the permitted values for backbone dihedrals.

5. Solve and NMR problem! Consider an N-acetylated tripeptide that contains a alanine, glycine and serine residue, but in unknown sequence. On the next page are two mock NMR spectra. Spectrum (A) is a 2D COSY spectrum for the tripeptide and spectrum (B) is a 2D NOESY spectrum. For simplicity I have made the peaks that are in the COSY spectrum circles and those that are unique to the NOESY spectrum squares.




a. Fill in the following table by identifying the chemical shifts of each set of protons belonging to each type of residue (you only need to use the COSY spectrum for this). Note that the "α $^1$H" for the acetyl group is the methyl group.

| | Chemical shifts (ppm) | | |
|---|---|---|---|
| | Amide $^1$H | α $^1$H | β $^1$H |
| Ala | **8.0** | **4.5** | **1.8** |
| Gly | **7.2** | **5.5** | **-** |
| Ser | **8.3** | **4.8** | **4.0** |
| Acetyl | **-** | **2.1** | **-** |

b. Use the NOESY spectrum to determine the sequence of the tripeptide. State your logic briefly. If you are curious which cross-peaks you would expect to see (< 5 Å), you can download a simple model of a tripeptide (one that has the right bond and dihedral angles!): **tripeptide.pse**

More logic.

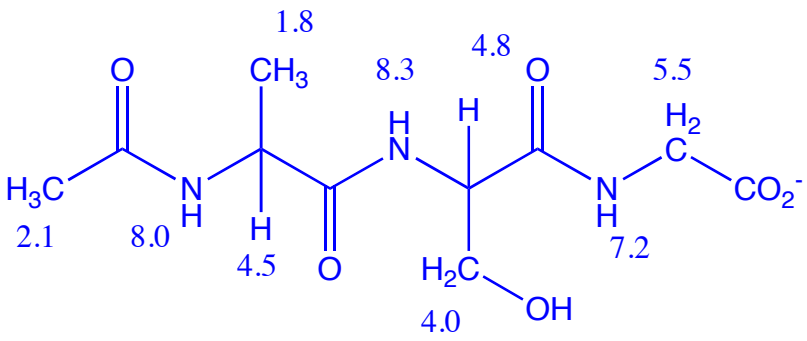
i. We know acetyl group is at N-terminus.
NOESY x-peak between 2.1 ppm and 8.0 ppm identifies N-terminal residue (Ala)!

ii. NOESY x-peak between 8.0 ppm and 8.3 ppm indicates 2$^{nd}$ residue (Ser)

iii. Process of elimination indicates that Gly is 3$^{rd}$ residue.

Note that pattern of x-peaks confirms this generally, but that's all you really needed:

**N-acetyl-AlaSerGly (structure on next page)**


1.8
4.8
O
CH3
8.3
O
5.5
H
H
H2
N
C
H3C
N
H
N
H
CO2-
2.1
8.0
H
4.5
O
H2C
7.2
4.0
OH

Reading of the week:
Cordes et al. (1999) Evolution of a Protein Fold in Vitro. *Science* **284**, 325-327.

1. What is the guiding question/hypothesis of this work?

"it is plausible to evolve a new protein fold from an existing fold by the accumulation of simple substitution mutations"

2. Write the initial and mutated sequence spanning residues 9-14. How can these sequences be described as amphiphilic?

Initial: QFNLRW Note alternating residues are hydrophobic and will be on one face of the strand.

Switch: QFLNRW Note that residues are 3-4 residues apart in sequence. Could occupy one face of a helix.

"In wild-type Arc, the sequence from 9 to 14 has the alternating pattern PHPHPH (where P is polar and H is apolar), which in a b strand leads to partitioning of apolar and polar residues on opposite faces. Switching the positions of Asn11 and Leu12 changes this pattern to PHHPPH."

3. Look at Figure 1.
a. What valuable information do panels A and B provide? Specify what each panel reports on.

The slightly larger dip in the far UV CD (A) suggests more helix content, but the huge disruption in the near UV (B) indicates major changes in environment of aromatic AA's. Since structure is relatively concerned (A), but changes are afoot (B), that's good news.

b. Similarly what is the import of panels C and D?

C is telling us that the Trp residue is seeing a major difference in its environment, but again thermal stability is roughly constant. Change, but stable.

4. In Figure 2, left, what is being plotted on the y-axis and how does it support the arguments made by the authors?

Differences in chemical shift of the alpha carbons for residues 9-14 indicate a change in chemical environment. Since that is where residue swap was made, it's nice to see localized changes indicative of 2° structure alteration that doesn't have global impact on protein structure.

5. Figure 2, right, is a NOESY spectrum of switch Arc.

a. Note the regions of the spectrum plotted on the y-axis and the x-axis. What atom types contribute resonances in each region?

y-axis is aliphatic protons (side chain of Leu 11 of big interest) and x-axis is the aromatic region, showing off resonances for F10 and W14.

b. Residues 10, 11, 14, 21 and 45 contribute to this spectrum (the atom types are given as Greek letters – the ID's can be seen in my amino acid handout of several weeks ago). Indicate the cross-peaks that are particularly relevant to this study and how they support the claims of the authors.

The x-peaks between 10 and 11, and between 11 and 14 could not happen in a strand, where residue 11 would be on the opposite face from 10 and 14, but are consistent with a helix.

6. Footnote 9 includes much of relevance to the NMR work reported here. Originally 28 models of the switch mutant were produced, but only 14 of those were employed in structural analysis of the protein. Provide two reasons, with brief explanations, for the exclusion of the other 14.

a. Nine of the structures had NOE violations greater than 0.3 Å suggesting a significant disagreement between model and data.

b. Five of the nine had problems in structure, with dihedrals in the disallowed region of the Ramachandran plot. Not good looking models.

7. They comment on features associated with the β strand of Arc that make it amenable to structural variation. Note two of them, with a brief explanation.

a. It appears that strand is pretty dynamic to begin with and that region is capable of accepting a variety of peptide conformaitons.

b. Because it is at one end of the polypeptide chain, it is more amenable to mutation than perhaps an internal secondary structure element would be.

8. Summarize the driving force behind the structural change observed in this work.

Hydrophobic effect!