Biomolecular Structure 2005 Exam #1

This exam is to be worked on independently. You must not talk, or otherwise communicate, with *anyone* other than Professor Martin or Professor Hardy about *any* aspect of the exam. You may not communicate with your fellow students in the class, nor with any other colleagues, faculty or student. For some of the problems it is necessary to use PyMol. PyMol may be accessed from the Macintosh computers in the Chemistry Resource Center, or may be downloaded onto your own computer for free at http://pymol.sourceforge.net/ .

Due in Craig Martin's office, October 31, 12:00pm

Honesty and integrity are absolute essentials for this class. In fairness to others, dishonest behavior will be dealt with to the full extent of University regulations.

1. Looks at this Evolve.pse from the course website. Based on the structures of the two proteins shown in this file, what can you surmise happened evolutionarily to get from one protein to the other. Comment on the relationship of the two proteins in structure and in sequence and the locations of any evolutionary events. (8 points)

In Evolve.pse select Recall Scene → F1 or Recall Scene → F2 to look at the TIM barrel of 1AMK (triose phosphate isomerase) or 1BDO (alanine racemase) in similar orientations. Since both of these proteins contain TIM barrels it is likely that they evolved from a common ancestor. In alanine racemase, a gene duplication event occurred inserting an entire intact mostly b-sheet domain into the TIM barrel domain (or potentially vice versa). The addition of domains usually occurs at the end of a domain or within a large mobile loop within a domain. Insertions almost never occur in the region of the gene coding for α-helices or β-sheets because that would destabilize the overall fold of the domain. As is usually the case, insertions of the predominantly b-sheet domain in alanine racemase occurred a loop region near the end of the TIM barrel domain.

2. What forces drive a protein to fold into the native state? What assumptions do we make about the energetics of the native state? What are the possible outcomes for a protein that does not fold into the native state? (9 points)

- The forces that drive a protein to fold to the native state are the same forces that govern any other chemical reaction. The enthalpy of positive non-covalent interactions such as hydrogen bonds, van der Waals interactions and ionic bonds contribute to the overall stability of the native state. The most important contribution to the stability of the folded state is the gain entropy when water molecules that are held in a clathrate structure around the hydrophobic residues in an unfolded protein are liberated upon proper folding.

- We assume that the protein folds to the global energetic minimum.
- A protein that does not fold properly is either refolded with the aid of chaperone proteins (optimal solution), degraded (acceptable solution),

or aggregates in a way that can lead to diseases such as Alzheimer's Disease (pathological situation).

3. As part of the structural genomics initiative, Ivan recently solved the structure of a new protein, ABB1 that is 17 % homologous in amino acid sequence to it's nearest neighbor. When he solved the structure he was surprised to find that it was a TIM barrel protein. What can he conclude about the function based on this finding? (6 points)

Unfortunately the finding that ABB1 has a TIM barrel fold does not help him in determining the function of ABB1. There are 21 different superfamilies of TIM barrels that have been designated based on based on both structural and functional relatedness (kudos to Jenny Lynn Maki for pointing this out). Proteins with the same fold can evolve different function (eg. TIM barrel families). Likewise proteins with the same functions can have different folds (eg. aspartate aminotransferase and D-amino acid aminotransferase which catalyze the same reaction but have different folds).

4. Two pdb files from NMR structures 1HKS.pdb and 1CFD.pdb are available on the course website of from the Protein Data Bank. Based on the information in the pdb itself, which structure do you think is the better structure? Why? (12 points)

	1CFD	1HKS
Total number of distance constraints	1855	1135
Number of amino acid residues	148	105
constraints/residue	12.5	10.8
Intraresidue restraints	383+158=541	378
Interresidue sequential restraints	392+68=324	158
Medium range restraints	355	190
Long range restraints	443	377
med & long range restraints/residue	5.4	5.4
Torsion angle restraints	400	65
hydrogen bond restraints	45	32
RMSD with respect to mean structure	most atoms less than 0.5A	not reported

Based on our lecture from Prof. Joanna Swain the quality of an NMR structure is based on: 1) The number and type of NOEs used, 2) number of torsion angle restraints, 3) number of H-bond restraints, 4) maximum restraint violations/residue, 4) number of deviations from idealized geometry 5) the precision of the structures and 7) the percent of residues falling into the allowed torsional regions.

From the table above, we note that 1CFD has a greater number of constraints/residue. On the other hand, both 1CFD and 1HKS have the same number of long and medium range restraints per residue. Long and medium range restraints are some of the most important factors in the quality of a structure, so on that basis, the two structures seem to be of similar quality. The authors of 1HKS do not report overall RMSD for their structure, so it is difficult to compare the two structures on this basis without further analysis. Based on the information provided in the PDB header these two structures seem to be of similar quality with 1CFD perhaps being of slightly higher quality.

5. On the course website you can download a file called ElectronDensity.pse. Please open this file in PyMol. For the electron density for one stretch of peptide from an intact protein is shown in this file. The backbone has been modeled in for you. This density is derived from a portion of the sequence:

YAWYQYERVDPRGVRYYWLYGRDLAPE

a. What stretch of the peptide shown above fits into this electron density? (5 points)

RGVRYYWLYG—See electondensityanswer.pse on the course web site.

b. What type of secondary structure is this stretch adopting? Please describe how you came to this conclusion. (5 points)

This segment comprises a strand—See electondensityanswer.pse on the course web site. Click on Scene→ Recall → F2.

c. Based on the amino acids present in this peptide, where do you surmise this peptide is found in the intact protein? (5 points)

Since this peptide has both some polar and some non-polar residues, it is likely that this segment is near the surface of the protein where polar groups can be exposed and non-polar groups are buried. This is true. See electondensityanswer.pse on the course web site. Click on Scene→ Recall → F3.

6. Jane recently isolated a new species of bacteria from a hydrothermal vent in the bottom of the ocean. Jane then isolated a gene (DMD1) for converting methane to diamond. Explain in some detail each of the steps you would undertake for determining the crystal structure of the protein coded for by DMD1. Please note considerations and assumptions you made for this course of action. Your answer should be no more than one page single-spaced. (12 points)

1. Clone the gene into an expression vector for over-expression or grow the thermophilic bacterium in your lab in order to purify the DMD1 protein from natural abundance.

2. Purify DMD1 protein to homogeneity.

3. Grow crystals of DMD1 using hanging drop vapor diffusion method. This works by slowly concentrating DMD1 protein in the presence of a precipitant until crystals form. The crystals are expected to be 100 mm on a side to be useful for crystal diffraction experiments.

4. Mount your crystal in a small loop and freeze at liquid nitrogen temperature.

5. Pass x-rays through your DMD1 crystal. X-rays that interact with the crystal lattice at the Bragg conditions are diffracted.

6. Collect data based on the symmetry of space group present in the crystal.

7. Solve the phase problem by:

a. molecular replacement (MR) method if a structure of a protein with a highly homologous sequence is available OR

b. by multiple isomorphous replacement (MIR) Note: if you said you would purify DMD1 from natural abundance and are not using (MR) this is the only option OR

c. by using Multiple Anomalous Dispersion (MAD) with protein that has incorporated some highly scattering atom like selenium in selino-methionine.

8. Calculate an electron density map from structure factors which are calculated from the measured intensities and the phases obtained in 7.

9. Build a model into the electron density map.

10. Refine your structure. Be sure to set aside 5-10% of your data with no refinement so that you can calculate R_{free} which is an unbiased measure of the goodness of fit of your model to the measured data.

7. Two pdb files from x-ray crystal structures 1SHT.pdb and 1GQF.pdb are available on the course website of from the Protein Data Bank. Look at the text of these PDB files.

a. What is your estimate of strength of these structures? Which is the better structure? What factors form the basis of your estimation? (4 points)

1SHT is the better structure. It was solved at better resolution (1.8 A) and has less overall error in the structure (R value = 0.15). In both cases the R and R_{free} are within 3% of each other with is a great sign meaning that neither of the structures are over refined.

	1SHT	1GQF
Highest Resolution	1.81	2.90
R value	0.150	0.268
R_{free} value	0.186	0.285

b. Which regions of which structures have the most imprecision in the coordinates? Why? (4 points)

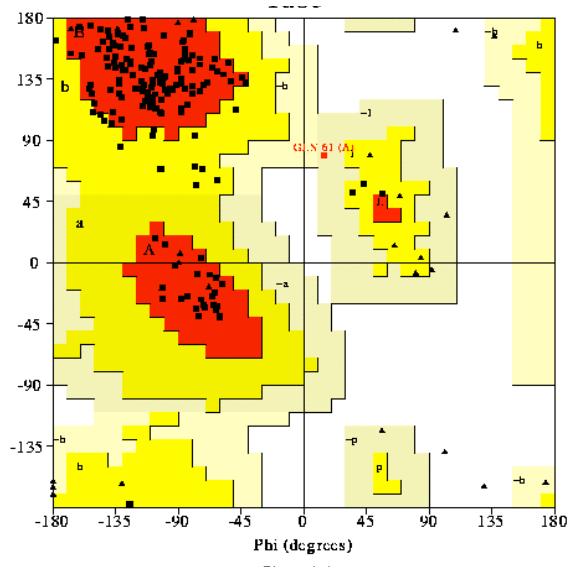
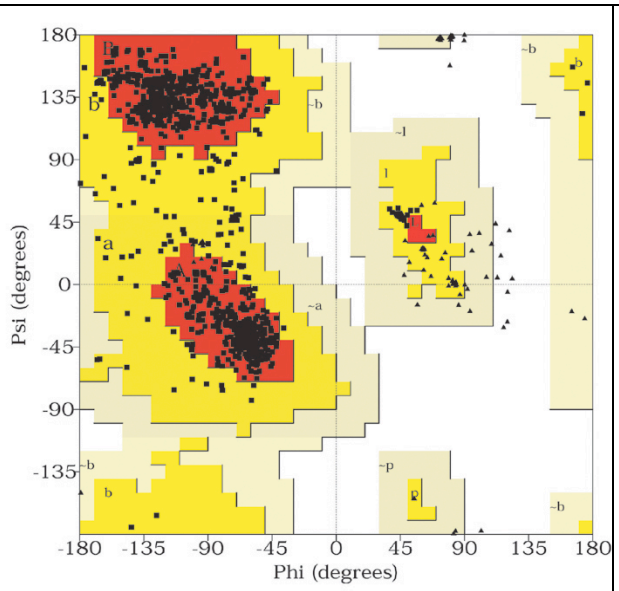
The regions of 1GCF with the highest B-factors have the most imprecision. B-factors are an estimate of the motion present in an atom. Atoms that have the most motion are least accurately modeled.

c. Are there any regions of either PDB that you would completely ignore? (4 points)

Yes. I would completely ignore the atoms in 1GQF that have an occupancy of 0 (eg. Thr 288 to Tyr 320). This means that these atoms are not visible at all in the structure.

8. Look at the two Ramachandran plots below. What can you conclude are the differences between proteins #1 and #2. What types of structures are the two proteins? Circle any residues below that you are concerned about and state why you are concerned. (6 points)

Protein #1 is smaller and contains a high percentage of β-sheet residues. Protein #2 is larger and contains a mix of α-helical and β-sheet regions. In general, one should be suspicious of any residue for which the phi/psi angles do not fall in the most favorably allowed regions (red) or additionally allowed regions (yellow). For residues in the generously allowed regions (tan) or disallowed regions (white) there ought to be extenuating circumstances. In general these residues should be revisited and corrected during the NMR or crystallographic refinement process. In these Ramachandran plots P symbolizes proline, triangles symbolize glycine and squares symbolize all other residues. In both structures only glycines and proline appear in the disallowed or generously allowed regions so there are few residues that cause great alarm. In protein #1 1GLN 61 (red square) is noted by the analysis program because of being in the generously allowed region. This is the residue that causes the greatest alarm.

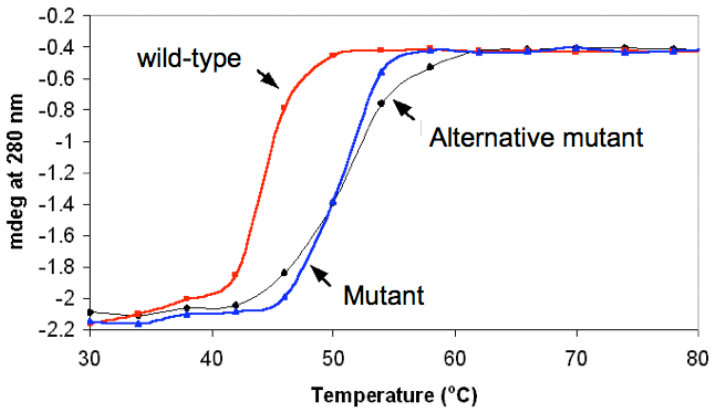
Ramachandran Plot Protein #1	Ramachandran Plot Protein #2
	

9. John has developed an algorithm for repacking protein cores. He used his computational algorithm to redesign the core of ubiquitin. He then introduced those mutations into the core of ubiquitin protein. In his new variant the side chains are more optimally packed than in the wild-type protein.

   a. What effect do you predict optimal side-chain packing will have on the thermal stability of ubiquitin? Draw the thermal denaturation curves you expect for wild-type and mutant ubiquitin with the optimally repacked core. (8 points)

In real experiments, repacking the core to optimize side-chain packing resulted in a variant of ubiquitin with increased thermal stability.

The expected melting curve would look like:



b. What effect will his mutations have on the ΔG of unfolding? (4 points)

$DG_{unfolding}$ for ubiquitin under standard conditions is positive (non-spontaneous). For more stably folded variant, $DG_{unfolding}$ would become more positive (more non-spontaneous).

c. What do you predict will be the effect on the function of the protein? Why? (4 points)

Tight packing of the core could be expected to rigidify the protein and in general would have negative effects on protein function.

d. Why did ubiquitin not evolve to have the thermal stability that of John's designed variant? (4 points)

Because evolution selects for protein function and in mesophiles, high thermostability is not required for function. In addition overly stable proteins may be resistant to regulation by degradation and might accumulate erroneously. Thus it may actually be unfavorable or pathological to have higher thermostability.