Question 1 Answer Sheet (be sure to attach your pymol picture) – any pymol picture that shows some feature of the protein is correct!

1) What type of Residue is this? What is the distance between Alpha Carbon and Hydroxyl Oxygen?
Threonine, 2.4Angstroms

2) What is the sequence of 3ihw?
FQGPELKVGIVGNLSSGKSALVHRYLTGTYVQEESPEGGRFKKEIVVDGQSYLLLIRDEGGP
PELQFAAWVDAVVFVFSLEDEISFQTVYNYFLRLCSFRNASEVPMVLVGTQDAISAANPRV
IDDSRARKLSTDLKRCTYYETCATYGLNVERVFQDVAQKVVALR

3) How many residues does it have?
167

4) What is the name of residue 5?
GLU (Glutamic Acid, Glutamate, E, are all fine answers)

5) What are the phi, psi, and chi1 angles of residue 5?
Phi = -81.33942
Psi = 136.812456
Chi = -76.3855386
(Any number of significant figures here are fine!)

6) What are the Cartesian coordinates of residue 5?
7.728000000000000   30.71100000000000   17.28400000000000
(Note, protein manipulation should not have changed the VALUE of the N backbone atom, as changing phi and psi angles only affect things C terminal forward)

7) List three terms in the rosetta standard energy function, and what they mean
Atr = attractive van der waals (Leonard-Jones potential)
Rep = repulsive van der waals
Fa_sol = solvation
Hbond_sr_bb, hbond_lr_bb, hbond_bb_sc, hbond_sc = hydrogen bonding energy
Dslf terms = disulfide terms
Fa_dun = statistical term related to how common your side chain conformation is
P_aa_pp = term related to whether angles are in ramachandran allowed regions
Ref = propensity of that amino acid

(Any 3 of these answers)

8) What is the score of your pose (including the modifications made by changing phi and psi angles)? Now reload the pose – what's the score of the pose now? Is it higher or lower, and why?

After modifications (before reloading) = 3358.59944
Before modifications(after reloading) = -47.472117895
Making modifications to the pose (altering phi, psi, and chi angles) moved the protein away from it's lowest energy folded state. The energy after manipulations was higher (less favorable)

9) What is the score of the pose with your score function?
8275 (or 7905.5735 if didn't include intra_rep) (before reloading)
531.37 (or 162.446349if didn't include intra_rep) (after reloading)
either value is acceptable – anything close to these values is an ok answer, and some people may have included different energy terms in their score function

10) What's the most important factor (highest value) in the score of your pose?
Fa_rep (highest raw score, and weighted score by a lot)! In both the modified and unmodified pose. Also, fa_atr in the unmodified pose has the lowest (most negative) energy, showing that it balances the fa_rep in the optimized pose.

Cells have built in mechanisms to respond to foreign small molecules – including inactivation, alteration of the small molecules target, reduced uptake of the molecule, or increased influx. In general, these mechanisms serve a protective function against potentially harmful agents and are important for organism survival. In particular, many organisms contain proteins called multidrug resistance transporters, which can transport a wide range of substrates out of the cell. They are mainly ATP-binding cassette proteins, and one very important version in humans is called P-glycoprotein. P-glycoprotein is highly upregulated in cancer patients, and is thought to be responsible for well over half of chemotherapeutic failures. The protein has over 70 known substrates (many of them are chemotherapeutic) and many of the substrates are known to upregulate the protein in tumors. Understanding the structure and function of P-glycoprotein is important in improving cancer patient outcomes.

The structure of P-glycoprotein has been solved and is available from www.pdb.org

a) Download 3G61 from www.pdb.org and make a ramachandran plot to study it's secondary structure –to do this you'll need to write a pyRosetta program to extr ct all the phi and psi angles from a pose, save them to a text file (the function dump_pdb(pose, 'file.txt') will be helpful for this).

```
from rosetta import *
rosetta.init()
import matplotlib.pyplot as plt

pose = Pose()
pose_from_pdb(pose, "3G61.clean.pdb")

#for later plotting, if desired
#f = open("angles.txt", "w")
n = pose.total_residue()
phi = [0]*n
psi = [0]*n
for i in range(n):
#    f.write(str(pose.phi(i))+"\t"+str(pose.psi(i)))
    phi[i] = pose.phi(i+1)
    psi[i] = pose.psi(i+1)

plt.scatter(phi, psi)
plt.axhline(y=0)
plt.axvline(x=0)
plt.show()
helicalResidues = 0
for i in range(n):
    if (-100 < psi[i] < 20 and -120 < phi[i] < -20):
        helicalResidues+=1
print str(1.0*helicalResidues/n)
```
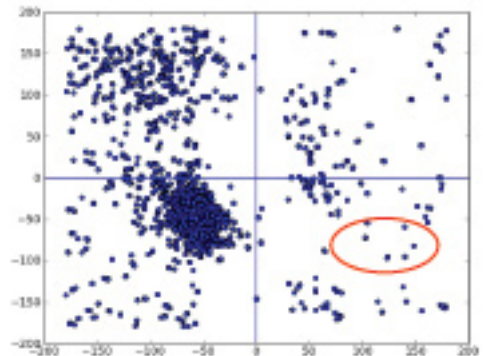
b) What secondary structural feature is the most common in this protein? What percentage of amino acids are in this secondary structure?

Protein is predominantly alpha helices. (~67% depending on what your cutoffs for what an alpha helix is were, will accept anything between ~50 and ~80)

c) Do you notice any residues (points on your ramachandran plot) that fall outside "allowed" areas of the plot? Circle one (more are ok than what we circled). What do you think this might mean?

Bad crystal structure refinement, angles are outside of traditionally allowed ramachandran regions

Chymotrypsin Inhibitor 2 is a 64-residue protein. To better understand the kinetics and equilibrium of folding of this relatively small protein, Itzhakiet. al. performed a series of mutations and studied folding equilibrium. A few of their mutations are selected here for further study. (Note, in the notation D52A represents residue D: Aspartic Acid at position 52 is mutated to an Alanine).

	ΔG_{fold}(kcal/mol)	Unfolded/Folded protein ratio
Wild Type	-32.07	3e-24
D52A	-14.1	4.55e-11
E14D	-29.9	1.17e-22
F50A	-16.2	1.31e-12
F50V	-22.2	5.2e-17

a.) Calculate the ratio of unfolded to folded protein for the wild type and each mutant (use R = 1.987 cal/K/mol and temperature = 298K)

[Folded]/[Unfolded] = K = exp(-ΔG/RT), so [Unfolded]/[Folded] = exp(ΔG/RT)

b.) Come up with a biological explanation for why each mutant had the small or large change on energy it had (you can do this by looking at the structure, or simply by knowledge of biology of the mutation made).

D52A: Aspartate may make salt bridges or hydrogen bond, but alanine can do neither, so the change is potentially drastic
E14D: mutates a glutamate to an aspartate, which amounts to a tiny steric change (one fewer carbon) but minimal change in bonding capabilities and both residues have identical charge
F50A: phenylalanines are typically towards the interior and may stabilize the core via hydrophobic interactions. Mutating to an alanine may eliminate many of these favorable hydrophobic interactions, as Alanine is much smaller.
F50V: Finally, the same mutation but to a valine preserves more of the steric footprint of phenylalanine, as well as the hydrophobicity, so the effect is not as strong as the alanine mutation, though still destabilizing.

Data taken from Itzhaki LS et al (1995) *JMB* **254**, 260-288

20.320 Analysis of Biomolecular and Cellular Systems
Fall 2012