**BIOC 530 Final Assignment**

**Name:**

**FoldIt Username:**

**Email:**

**Department: Student ID #:**

**Instructions: answer all questions with no more than ONE paragraph. You don’t have to provide an answer for question 5; instead your FoldIt account will be accessed for evaluation.**

**1.** One technical challenge in protein NMR is improving the signal to noise ratio. As discussed, the population of the states is described by the equation:

**(N1/N2) = e^[-(E1-E2)/kT]**

And the difference in energy levels between states is given by:

**ΔE= γ h Bo**

Given this, how can the signal to noise ratio be improved by changes in experimental setup, and why?

**2.** Use PSI-BLAST [(http://www.ncbi.nlm.nih.gov/BLAST/Blast.cg](http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi)i) to search for sequences related to the C chain of the 1KCG crystal structure (you can find its sequence on the RCSB PDB).

a. What protein is this?

b. To what class of proteins is homology revealed with PSI-BLAST?

c. Run a second iteration of PSI-Blast by clicking on the ‘Run PSI-Blast iteration 2’ button under ‘Description’ on the results page. How do the identity and ordering of results change after a second iteration of the PSI-BLAST algorithm? With what you know about PSI-BLAST from the lectures, why do you think this change occurs?

**3.** Proteins can be grouped based on structural similarity. Use the SCOP database [(http://scop.mrc-lmb.cam.ac.uk/scop/data/scop.b.html](http://scop.mrc-lmb.cam.ac.uk/scop/data/scop.b.html)) to find two SH3 domain-like folds that are not in the SH3 domain sequence family. Use PSI-BLAST to determine whether the structure similarity is detectable using sequencing based searching.

If protein A, B, and C all have similar structures, but A and B have significant sequence similarity whereas A and C are not found to be related via sequence based searching, what does this imply about the evolutionary distance between the three proteins? What information does structural analysis potentially give you that sequence based searching may not?

**4.** You have found a sequence in the Salmonella genome that codes for a protein that you think may be a good drug target. The sequence is:

**ASLTNSKIQGKYRHSFWSEIYETVLAWYIAPPTLVALINPHKGKFNVTAKGGLVEEKYVD WVISRPYIFLVLLNLLGVAAGVWRYYYGPENETLTVIVSLVWVFYNLVILGGAVAVSVESK QVRRAHRVEIAMPGAIAREDGHLFSCTVHDFSDGGLGIKINGQAQVLEGQKVNLLLKRG QQEYVFPTQVVRVTGNEVGLQLMPLTTKQHIDFVQCTFARADTWALWQDSFPEDKPLES LLDILKLGFRGYRHLAEFAPPSVKVIFRSLTALIAWIVSFIPRRPERQAAIQPSDRVMAQAQ Q**

The structure of this protein has not been solved. Submit this sequence to a fold recognition server ( you can find one at <http://toolkit.tuebingen.mpg.de/hhpred>) to search for similar structures. What protein structures (give PDB code) did you find matches to? Explain how the method you chose to find the similarity works – what information do fold recognition servers incorporate to predict the fold of a target sequence?

Does PSI-BLAST find the similarity between the target sequence and the predicted structural match? Look at the first 10 aligned residues between your chosen target and the matched PDB file. Is there much sequence similarity? Explain.

**5.** For this problem, download and install the protein folding computer game FoldIt from <http://fold.it> (it is available for PC, Mac, and Linux). Create an account (**please specify your username in your answer**) and play at least the first four ‘Intro Puzzles’.

After you’re familiar with the game, join the contest by following this [link](http://fold.it/portal/smartlogin?destination=addcontestant/31f5db8e78290cbb2bf6446030abe403). After logging in with Foldit username and password, you will be added to the contest.

**Objective**

The purpose of this design contest is to design a stable protein fold that biophysically makes sense (think about solvation, buried charge residues, etc.). You will be evaluated on the final model, as well as the number of attempts you tried designing it (wiggle, shake, recipes, etc.). Keep in mind that the Foldit score is not the best guide. For example, beware disordered poly-Arg chains or designs with a lot of polar residues in the core; these designs will have good Foldit scores, but are clearly not valuable biophysical models.

**Playing the Contest**

To play the contest, you need to open Foldit (which should already be installed), log in with Foldit username and password, and navigate to the Contest menu. (Note: If a tutorial level opens automatically when you log in, you can find the Contest menu by clicking Menu->Puzzle Menu->Main Menu->Contests)



If you have already joined the Contest, you should see a Contest named "BIOC530 Final Exam” in the Contest menu. Select this title and click Play.



**Submitting a Design**

You can save a design at any point by clicking Menu->Save Solution. There are fields to record a Name and Description of the solution (**please use your own names in the Name field and explain your design in the Description field. See screenshots on next page**).





Once you are ready to submit a saved design, you can upload it by clicking Menu->Open/Share Solutions. You should find your saved designs on the right side under “Local Solutions”—select a design and click Upload For Scientists. The shared design should now appear on the left side under “Shared Solutions."



**6.** Ab Initio Structure Prediction:



Protein A Protein B

In class, we discussed both the challenges in ab initio protein structure prediction and its utility in protein design. Above are plots for energy vs. RMSD for two computationally designed proteins. In green are small perturbations about the “native” designed structure, while in red are the ab initio structures given by folding the design from an extended chain. Based on these plots, which protein do you think would be more likely to fold up to its designed structure and why?