

## BCB410 - Homework 3: Protein Structure Prediction

Due: Wednesday December 2nd

1. Assume a colleague wanted to estimate the structural propensities for short sequences of amino acids. Specifically they want to build a large table of all sequences of 7 amino acids. For each 7-mer they want to record the most likely conformation as observed in experimentally solved protein structures. Their hope is that given a query protein sequence they could use the table to look up the most likely conformation of each 7-residue window and then use this information to generate an initial estimate of the query protein's structure.
  - (a) How many 7-mers are there?
  - (b) Approximately how many 7-mers are present in the current PDB database? (Please state any assumptions you make in estimating this number)
  - (c) Given your answers to (a) and (b), is there any hope that our colleague's table will be useful? If not, propose a work-around.
2. In the last assignment, you learned that the mystery cancer protein isolated by your collaborators was in fact a fusion protein, consisting of two domains. One of the functional domains is a kinase which has now been fully sequenced. In this question, you will attempt to perform tertiary structure prediction for the kinase domain of Protein X (given in "protein.fasta"). To do so please perform the following:
  - (a) Go to [http://en.wikipedia.org/wiki/Protein\\_structure\\_prediction\\_software](http://en.wikipedia.org/wiki/Protein_structure_prediction_software) and locate the list of protein structure prediction programs. You will use two programs from this list to predict the structure of Protein X. The first program you will use will be ESyPred3D and the second program will be your choice (although you should not use any programs from the 'secondary structure prediction' nor 'transmembrane helix and signal peptide prediction' lists). The ESyPred3D server is very easy to use, simply enter your email address (to be notified when the run is complete), copy/paste the fasta sequence of Protein X into the large text box, and **most importantly, be sure to enter 1PKG in the 'Use this PDB template'**. You will also want to make sure that 'Summary of results' and 'PDB format' are selected under the output section. (Note: You may need to exclude the comment line of protein.fasta when you copy/paste into the ESyPred3D sequence box). Depending on the server load, the ESyPred3D server will take between 10 minutes and several hours to complete your job (so submit this right away). Document your experience using both the ESyPred3D server and the second server you have chosen to use. Your description should include (but not be limited to): the name and category of program you used, a few sentence description of the method used by the program, the algorithm parameters, how long the run took, how easy it was to get a result. Use screenshots to help document your experience. (The text of your write-up for part (a) should not exceed one page of typed text (not including figures)).
  - (b) You should have obtained a pdb file from the ESyPred3D server. Using PyMOL or any other molecular visualization program, open and visualize either the ESyPred3D prediction or the prediction from the other software you selected. (Note that the newer versions of PyMOL require registration and licensing, but older versions are available freely at: <http://delsci.com/re1/099/>). It may be useful to visualize the protein in a 'cartoon' view or backbone only. Paste a screenshot of the PyMOL view of the structure as part of the assignment. Now open the pdb file 1T46 (available on the course website). The 1T46 structure is the experimentally determined structure of Protein X (in reality, it's not exactly protein X, it's just very close). Do the two structures look similar? Are they the same size? Are there noticeable differences?

A useful list of molecular visualization programs can be found at:

[http://en.wikipedia.org/wiki/List\\_of\\_molecular\\_graphics\\_systems](http://en.wikipedia.org/wiki/List_of_molecular_graphics_systems).