Biol 591 Introduction to Bioinformatics (Fall 2003)
Problem Set 7: Protein threading and modules

Modules

P7.1: Sometimes lower case indicates doubt in sequence quality. Change FastA_module so that sequences read in are not always upper case.

P7.2: Make a module consisting of three subroutines:

Left (**string**, **n**): returns the **n** left-most characters of **string**
Right(**string**, **n**): returns the **n** left-most characters of **string**
Mid(**string**, **m**, **n**): returns a region from within **string** from coordinate **m** to **n**

Be sure to use as a model one of the modules you already have and to carefully copy all portions that are necessary for the module to work.

P7.3 It is often useful to calculate the molecular weight of a protein from its sequence. For example, suppose you identify by gel electrophoresis an interesting protein. The gel gives you the approximate molecular weight. You want to find the gene that encodes it. You can go through a file containing all gene products of an organism whose genome has been sequenced and for each one calculate the molecular weight. From this, you can construct a list of candidate genes for your protein.

For starters, modify AA_module so that it facilitates the calculation of molecular weights of amino acid sequences. To do this:

a. Add to the amino acid data (at the end of module) a column containing the molecular weights of each amino acid.

b. Define a hash that will contain the molecular weight of an amino acid, keyed to the one-letter code of the amino acid (e.g. $AA_mol_weight{"L"} = molec weight of leucine).

c. Modify Read_AA_info so that it reads in the molecular weights and assigns them to the hash described above.

d. Make a function that takes as an argument an amino acid and returns its molecular weight.

e. Test these changes by using Protein_mol_weight.pl (available from the unit web page).

Protein visualization and threading

P7.4. Which specific amino acids of UDP-glucose dehydrogenase from *Mesorhizobium loti* do you consider prime candidates for site-specific mutagenesis, in the hopes of obtaining highly soluble enzyme in *E. coli*? Why do you think so? [This is a summary of SQ1-SQ18 from the notes for November 17/19. Of course all study questions are implicitly deemed part of this problem set, but I thought I'd make this one explicit.]

P7.5. ThreadProtein.pl may mislead by retaining deleted amino acids in the displayed protein. Fix this by modifying the program to ignore deleted amino acids.

P7.6. Sickle cell anemia is a genetically determined trait, the result of mutation in the beta chain of hemoglobin: glu 6 val. The disease state arises when hemoglobin precipitates within the red blood cell, leading to a sickle-shaped cell. Bring up hemoglobin in Protein explorer and investigate why this mutation might lead to precipitation. You might also look for a structure for hemoglobin from a person with sickle cell anemia to see what precipitated hemoglobin looks like (note the facility on the front page of Protein Explorer to find PDB files).

P7.7. Make your own molecule! Modify a pdb file, tossing out all the atoms and replacing them with your own. Provide coordinates such that Protein Explorer will display something interesting. Your choice,, but here are some suggestions:

a. Boxane: Put an atom at the eight corners of a cube and put a big atom inside the cube

b. Christmas tree: Build a tree from the bottom up with concentric circles of glycine (a small amino acid), colored green. Then put a brown pole of alanines down the middle, and a bright red arginine on top.

c. A smiley face: Circle, semicircle, two dots.