Lab - Protein-Protein Interfaces

Many biological functions involve the formation of protein-protein complexes. A complex is a structure consisting of two or more interacting proteins. The interface of interacting proteins is defined as the set of amino acids of the two proteins within a certain threshold distance. The characterization of protein-protein interfaces is important because it will enable the prediction of protein interactions providing insight into their function.

This project is about the determination of protein amino acids which are at the interface of two interacting proteins. In the previous project you have computed the Euclidean distance of two atoms of a protein. In this project, given two interacting chains A and B of a protein complex, you compute the distance between every amino acid of chain A and every amino acid of chain B. ***If the distance of two amino acids is less than a certain threshold, both amino acids are registered as the interfaced amino acids.*** The threshold is generally selected arbitrarily by trial and error.

The project consists of two parts.

Part 1. (80 points)

Write a program that computes protein-protein interfaces.

In a pdb file the interacting proteins in a complex are listed as separate chains. Your program will consist of the following steps carried out sequentially.

1. It takes in input the name of a protein, the names of two chains (for instance A and B), and a threshold value
2. It reads the corresponding PDB file and extracts the C_alpha atoms of the amino acids of the two chains
3. It computes the distances Dist_Calpha of all pairs of C_alpha atoms, with one C_alpha in chain A and the other in chain B.
4. For each chain, it outputs the interface amino acids, i.e. the amino acids with distance less than the input threshold from an amino acid of the other chain

This is an exercise on subroutine in perl. Thus, in your program **you are required** to use subroutines. Use one subroutine to read in the pdb file and to output the coordinates of the C_alpha atoms. Use another subroutine that takes as input parameters the coordinates of two points (C_alpha atoms) and returns their distance. You may use subroutines for other functions as well.

Part2. (20 points)

Visualization. Use JMOL or any other tool to visualize the the two chains and their interface. You may do so by displaying the entire protein and then selecting a set of amino acids (those at the interface) and color them in a different color or display them in a different style. If you use JMOL you select the amino acids by using a console command, as explained at:

http://people.virginia.edu/~dta4n/biochem503/console.html

Implementation and Submission.

Submit a file with name "your_last_name_interface.pl".

Test your program on protein 1atp, chains E and I, or any other protein of your choice. Use a threshold of 6, or 7 or choose your own threshold. The input arguments should be provided on the command line, for instance:
perl your_last_name.pl 1atp.pdb E I 7
For output, you will report each pair of interacting amino acid from two different chains in the following format:
<CHAIN_ID>:<AA CODE>(<AA NUMBER>) interacts with <CHAIN_ID>:<AA CODE>(<AA NUMBER>) as for instance in the line below
A:LYS(255) interacts with B:LEU(353) (This is an example not a real correct part of the output for 1atp)