Introduction to Bioinfromatics

Task - 2

Name: Priyank Patel
UIN: 00985578
Email: ppatel@cs.odu.edu

Extraction of SLS Patterns from Protein Structures

Project Requirements:

In this task we have to consider multiple .pdb file as an input. We have to extract the SLS from each of them and align those SLS in a way that their CA1 and CA2 will align on same vector.

My approach:

I have created two versions off the program,
The functionality of the program is completely same but both are following different calculation for rotation angles. Due to which they are generating slightly different results. (file names: a2_v1.pl, a2_v2.pl)

I have created a program in "perl" language to parse the PDB file to get the required data to serve the task and process the data.
I have processed the files using the approach below,

1. Ask user for the input file names separated by comma.
2. Also whether user wants to align SLSs identified or no.
3. The third user input I asked for is the size of the desired loop in a SLS.
4. If the use do not want to align the SLSs then the program will simply open the .pdb file one by one.
5. Parse all the location information about Sheets and helixes from each of the files.
6. Sort this information to make the searching easy.
7. Compute interval length between all the pairs of Sheets
8. Compare this computed length with the desired length of the loop.
9. If the computed length and the desired lengths are equal then check if the helix is present in that interval or not.
10. If helix is not present then extract the Atom coordinates information of that SLS and write it to the file, else go to the next structure.

The program if it finds any suitable SLS in the file will write a file with the following file name convention.
NameOfProtein_startPositionOfLoop_SizeOfLoop.pdb

11. If the user wants to align all the SLSs then the program will open the 1st .pdb file.
12. Open the steps 5, 6, 7, 8.
13. After that the program will identify CA1 and CA2 of the SLS and consider them as base CA1 and CA2 to align all the other SLSs accordingly.
14. The program will open rest of the .pdb files and performs the same steps 5, 6, 7, 8 and 13 on it.
15. After that the program will make all the CA1 coordinates for all the new structures similar to base CA1 and translate the structures.
16. The program will calculate the vector between the base CA1 and CA2 as well as the vector between CA1 and CA2 of all other incoming structures.
17. The program will calculate the rotation angles and from that rotation matrix from all the base vectors and other vectors.

18. It rotate the new structures based on the base structure rotation matrix to make all the CA1 and CA2 coordinates in the same direction.

Running the program:

To run the program, follow the steps below,

1. Put all PDB structure files in a folder with .pl file.
2. Run the perl file with the command “perl –w filename.pl”.
3. The program will ask you to give the PDb file names separated by comma.
4. The program will ask you whether you want to align the SLSs or not.
5. Also it will ask for the length of the loop you want.
6. The result without alignment will print all the sheet and loops present in the file in sorted order and creates a file for each SLS found.
7. The results with alignment will generate a file named “aligned_result.pdb”, which can be opened in any 3D structure visualization tool.

Test:

The program is tested for files input,

1NVS.pdb,1AOP.pdb

And the results generated successfully.