

BIN 515 – Structural Bioinformatics

Assignment 3

Due date: May 5, 2020

Late policy: For each assignment, 20 points deduction will be applied for one day late, and 10 points additional deduction for each extra day.

Use your preferred programming language to implement the required methods to solve the problems below. Please, submit a complete report with the program codes. Discuss your results in your report.

Part 1: (50 pts) Secondary Structure Prediction: Select a single chain human protein from PDB with at least 150 and at max 250 residues. Use the advanced search option at PDB to select the protein. Use the sequence of the selected protein to predict the secondary structure of each amino acid using the Chou-Fasman algorithm. Your task is first to implement the Chou-Fasman algorithm. Then, find the real secondary structure assignments of each amino acid in the selected protein. Calculate the accuracy of the CF algorithm on your case study.

Part 2: (50 pts) Interface Extraction: Select a two chain human heterodimer from PDB with at least 150 residues. Use the advanced search option at PDB to select the protein complex.

Proteins interact through their interfaces. There are several approaches to find interface regions from 3-dimensional coordinates of a protein complex, such as calculating the accessible surface area (ASA) of the residues or calculating the atomic distances. In this assignment, you will use the ASA calculation method. If the difference between the ASA of a residue in monomeric state and complex state is greater than a threshold (usually 1 \AA^2) that residue is labeled to be an interface residue.

1. Use the getArea server (<http://curie.utmb.edu/getarea.html>) to find residue level ASA values for the monomer state of chain A and chain B and the ASA value of the protein complex.
2. Write a program to extract interface residues based on the cutoff value given above.
3. Visualize the complex in new cartoon representation and represent the interface residues in VdW using VMD. Color each chain of the protein and interface region in different colors.
4. Calculate the residue type propensity of the interface region. The formula of interface propensity is present in the lecture notes. Additionally, which other features would you obtain about the interface region from the available databases? Discuss it in your report.