Iqb assignment -2

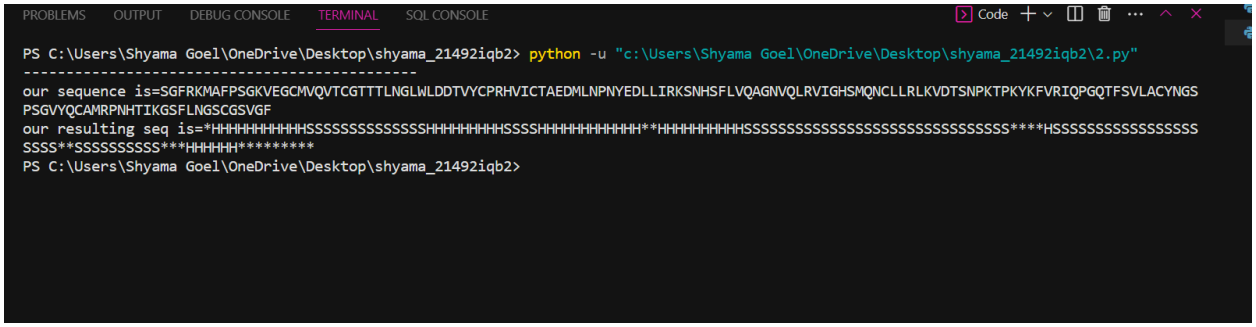
Submitted by- Shyama Goel

2021492

1) The resulting output is as follows->

*HHHHHHHHHHHHSSSSSSSSSSSSSSSHHHHHHHHHHSSSSHHHHHHHHHHHHH**HHHHHHHHHHSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS****HSSSSSSSSSSSSSSSSSSSSS**SSSSSS SSSS***HHHHHH*********

A well commentedcode is submitted it in .py file already.



2)

1)Prediction approach->Category I methods, such as Chou and Fasman and GOR, use amino acid sequence information for secondary structure prediction, while Category II methods, such

as DSSP, P-curve, and Stride, rely on 3D structural data. The former relies on empirical rules and statistical analysis of amino acid sequence, while the latter uses geometric and hydrogen-bonding features of protein structure.

2)Performance: Category II methods are slower and more computationally intensive, making them better suited for predicting secondary structure for specific protein structures, while Category I methods are relatively quick and can predict secondary structure for huge datasets.

3)Accuracy: When predicting secondary structure, Category II approaches are typically more accurate than Category I methods. This is due to the fact that Category II approaches account for the polypeptide backbone's true three-dimensional configuration.

4)Input data:Category II methods need the three-dimensional structure of the protein, which can be obtained by X-ray crystallography or NMR spectroscopy, while Category I methods just need the amino acid sequence of the protein as input data.

5)Format of output: Category I methods usually offer a score for the likelihood of each secondary structure type at each place in the protein sequence, but Category II methods offer a thorough output that includes the precise secondary structure assignments for each residue.

Chou and Fasman	GOR method
1 The Chou-Fasman approach determines the possibility that a specific amino acid will be found in an alpha-helix, beta-sheet, or random coil structure by statistically analysing known protein structures.	1) It is an information theory-based method for the prediction of secondary structures in proteins.
2)To create predictions using the Chou-Fasman approach, at least 20 amino acids are needed.	2) GOR method can make predictions with as few as seven amino acids.
3) The Chou-Fasman method was developed using a small dataset of 12 protein structures.	3)Using a bigger dataset of 426 protein structures, the GOR technique was created. Because of the larger training sample, the GOR technique may be more generalizable.
4)The Chou-Fasman method is more specific for predicting alpha-helices	4)GOR method is more specific for predicting beta- Sheets.

5)It is less accurate as compared to gor	5)It is more accurate as compared to the chou fasman method.

b)

DSSP	P-curve	Stride
1)DSSP employs a set of principles based on the hydrogen-bonding patterns and backbone dihedral angles of protein structures	1) P-curve predicts secondary structure from primary sequence data using a machine learning method.	1)Stride predicts the secondary structure elements by combining geometry, hydrogen-bonding patterns, and amino acid sequence data.
2)DSSP is standalone software packages that can be downloaded and installed on local machines.	2)P-curve is a web-based server that offers online prediction services	2)Stride is standalone software packages that can be downloaded and installed on local machines.
3)DSSP can also calculate other parameters such as solvent accessibility, hydrogen bond patterns, and amino acid burial.	3)It is based on distribution of p values in dataset.	3)It is also used in the calculation of other parameters such as solvent acesssibility and hydrogen bonding patterns.
4)It is used for studying protein structures at atomic level	4)It is used for analying large datasets.	4)It is also useful for studying proteins structures at atomic level which can have further applications.