ASSIGNMENT-2

Question-1:
In this question we need to predict the regions where α-helix and β-sheets can be found.For this we used Chou-Fassman's algorithm for predicting the regions.

For α-helix: We found the nucleation sites by finding 6 consecutive residues where p(H)>1 for at least 4 residues. The we extended the structure leftwards and rightwards to get the maximum length by considering 3 residues from the original 6 and adding a new one

Output Sequence of the code:
_HHHHHHHHHHSSSSSSSSSSSSSSHHHHHHHHHSSSSHHHHHHHHHHHH__HHHHHHHHHHSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS____HSSSSSSSSSSSSSSSSSSSS__SSSSSSSSSS___HHHHHH_________

H → α-helix
S → β-sheets

(screenshot added in the folder)

Question-2:
In this question we calculated the predicted structure using stride as given in the question. Then we calculated the difference between the output of stride and the output that we got in the above question.

Note: In stride β-sheets are denoted by E and in the first part we took β-sheets as S.

Finally we calculated the % similarity between the output given by stride and the output that we generated. The % similarity is calculated by:

$$\%\,similarity = \left(\frac{similarity\,count}{total\,count} \right) \times 100$$

The output of the code is:

Screenshot added to folder

Reasons for dissimilarity:

1. The Chou Fasman Method is a traditional method. Since it is traditional method the data available was very limited hence the pridications were not very accurate
2. The Stride method showed some empty spaces where the Chou Fasman method showed α-helix or β-sheets due to overprediction done by this method. Hence this causes dissimilarity.
3. Chou Fasman method doesn't take into account the hydrophobicity profile of the amino acid which is essential for prediction
4. The Chou Fasman method doesn't consider protein folding while predicting the structure. In secondary structures protein folding plays an important role and hence there is a vast dissimilarity between the two results.
5. We haven't considered other parameters like turns and coils in the Chou Fasman method while the output of the stride contains those parameters as well hence there is a dissimilarity.

Keshav Gambhir
2019249