BIO213-Introduction to Quantitative Biology
Assignment 2
Kesar Shrivastava, 2019051

Question 1) (a)
Code submitted as .py file as well as .ipynb file.

Question 2) (b) This is the output obtained:

```
THHHHHHHHHHHSSSSSSSSSSSSSSHHHHHHHHHSSSSHHHHHHHHHHHHTTHHHHHHHHH
HSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSTTTTHSSSSSSSSSSSSSSSSSSSSTTSS
SSSSSSSSTTTHHHHHHTTTTTTTTT
```

Question 2) (a)

This is the output using the STRIDE webserver:

```
SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDTVYCPRHVICTAEDML
TTTT     HHHHHH EEEEEETTEEEEEEEETTEEEEGGGG  HHHHH

NPNYEDLLIRKSNHSFLVQAGNVQLRVIGHSMQNCLLRLKVDTSNPKTPK
   HHHHHHH  GGG EEEETTEEE EEEEEEETTEEEEE   TTTT

YKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNHTIKGSFLNGSCGSVGF
      TTTEEEEEEEEETTEEEEEEEEETTTT B    TTTTTTTEE
```

The result obtained in question one is as follows:

```
SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDTVYCPRHVICTAEDML
THHHHHHHHHHHSSSSSSSSSSSSSSHHHHHHHHHSSSSHHHHHHHHHH

NPNYEDLLIRKSNHSFLVQAGNVQLRVIGHSMQNCLLRLKVDTSNPKTPK
HTTHHHHHHHHHSSSSSSSSSSSSSSSSSSSSSSSSSSSTTTTHS

YKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNHTIKGSFLNGSCGSVGF
SSSSSSSSSSSSSSSSSSTTSSSSSSSSSTTTHHHHHHTTTTTTTTT
```

Our algorithm identifies beta strand, alpha helix and turns but the STRIDE webserver has identified beta strand, alpha helix, turn, coil, 310helix, and bridge too. In the webserver's output E refers to strand, B to bridge, H to helix, T to turn, G to 310Helix and the blank space

for coil. The regions of difference are given below in the output. The regions differing in both in terms of assigned secondary structure is denoted by -

```
SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDTVYCPRHVICTAEDML
T--------HHH----SSSSSS--SS---------SSSS------HHHHH

NPNYEDLLIRKSNHSFLVQAGNVQLRVIGHSMQNCLLRLKVDTSNPKTPK
---HHHHHHH------SSSS--SSS-SSSSSSS--SSSSSS---TTTT-

YKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNHTIKGSFLNGSCGSVGF
---------SSSSSSSSS----SSSSSSSS--TT-------TTTTTT---
```

Question 2) (b)

The differences in the assignment are observed because the STRIDE method uses an empirically derived hydrogen bond energy and phipsi torsion angle criteria to assign secondary structure while the Chou-Fasman method uses the propensity of the residues which is the percentage of the residue in that conformation to the percentage of all residues in the same conformation as the parameter to assign the secondary structure.
Next to hydrogen bonds STRIDE also includes backbone geometry in the form of dihedral angle propensities. It aims to provide secondary structure assignments that are more consistent with the assignments performed by experimentalists who determined the protein structure.
Moreover, it assigns other secondary structures as well while our algorithm assigns alpha helix, beta strand and turns.
Since they follow quite dissimilar method of assigning protein structure and the chou-fasman is completely dependent upon the propensities of amino acid residues we observe the differences.

(Read about STRIDE from Wikipedia and the information available at the STRIDE webserver: http://webclu.bio.wzw.tum.de/stride/stride.pdf)