README

- ## IMPORTANT FUNCTIONS

  ### sequencegenerator(cnt):
  This function generates random sequences with 'L', "R' and 'F'. Here, 'cnt' specifies the number of sequences to be generated

  ### legalitycheck(text):
  This function checks of the sequences are legal or not. That is, it checks each sequence for collisions. The function takes the array of sequences as input and returns the number of collisions.

  ### getScore(matr,matr_angle,matr_visited):
  This function takes 3 matrices as input"
    a. matr: stores the path of hydrophobic-hydrophilic sequence with respect to a particular LFR sequence.
    b. Matr_angle: stores the angles with respect to the LFR sequence
    c. Matr_visited: stores 1 for all the visited nodes

  The function returns the score of the LFR sequence. Score is the number of non-covalent H-H bonds.

  ### rmvIllegal(ind,text):
  This function removes illegal sequences from the array of LFR sequences

  ### crossoverAndSelection():
  This function creates the new set of sequences with child sequences created by doing crossover on the initial sequences.

  ### mutation():
  This function mutates each direction in the sequences with one of the remaining directions with mutation probability p.

- ## INPUT
  1) ip: input file name which contains the amino acid sequence
  2) n1: number of LFR sequences to be generated
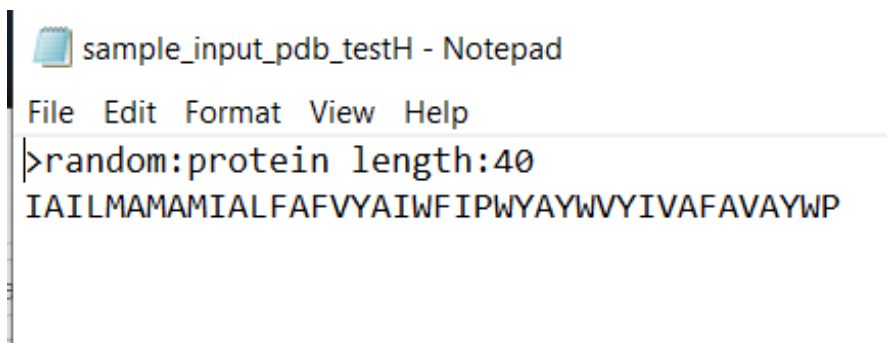 3) n2: number of generations
  4) p: mutation probablitiy

- **OUTPUT**

  4 graphs
  1) Maximum Score graph: plots maximum scores with respect to generation
  2) Average Score graph: plots average scores with respect to generations
  3) Maximum Score trend: plots the trend of max scores over a set of generations
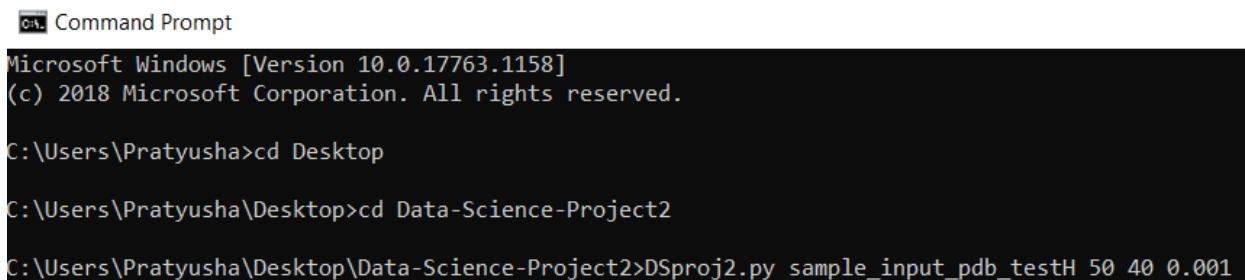 4) Average Score trend: plots the trend of avg scores over a set of generations

- **SAMPLE RUN**

Input file 'sample_input_pdb_testH' contains:

sample_input_pdb_testH - Notepad

File   Edit   Format   View   Help

>random:protein length:40
IAILMAMAMIALFAFVYAIWFIPWYAYWVYIVAFAVAYWP

Command line input:

- Input file: sample_input_pdb_testH.txt
- Number of sequences: 50
- Number of generations: 40
- Mutation Probability: 0.001

Command Prompt

```
Microsoft Windows [Version 10.0.17763.1158]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\Pratyusha>cd Desktop

C:\Users\Pratyusha\Desktop>cd Data-Science-Project2

C:\Users\Pratyusha\Desktop\Data-Science-Project2>DSproj2.py sample_input_pdb_testH 50 40 0.001
```

Output Graphs:

Maximum scores for generations

Average scores for generations

Trend for Max Scores

Trend for average scores