

Part 1 - Calculating Amino Acid Frequencies [40 pts.]

3- Most common amino acid is Alanine (A) while the least common amino acid is Aspartic Acid (D) as shown in figure 1.

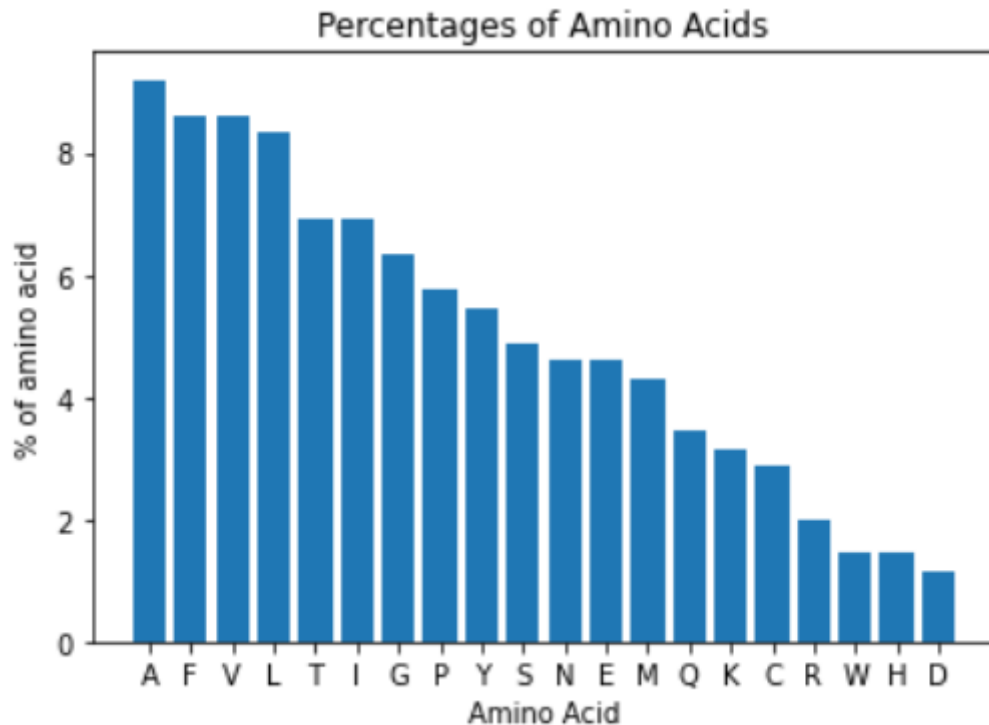


Figure 1: Amino acids vs percentages of Amino Acids bar graph. Taken from the python file.

4- In the Uniprot, most common amino acid is L (Leu) with 9.86% while the least common amino acid is C (Cys) with 1.24% for the complete database as shown in the figure 2. In the other hand, for the rhodopsin protein, most common amino acid is A (Ala) with 9.2%, while least common amino acid is D (Asp) with %1.15. We can see that in the Uniprot second most common amino acid is Ala with 9.23% which is very close to the percentage of Ala in rhodopsin's protein sequence. For overall distribution, although there is a lot differences between the frequencies of some proteins, rhodopsin's distribution is quite similar to the general amino acid distribution in Uniprot.

Classification of the amino acids by their frequency in Uniprot

Leu, Ala, Gly, Val, Ser, Glu, Arg, Thr, Ile, Asp, Pro, Lys, Phe, Asn, Gln, Tyr, Met, His, Trp, Cys

Classification of the amino acids by their frequency within Rhodopsin

Ala, Phe, Val, Leu, Thr, Ile, Gly, Pro, Tyr, Ser, Asn, Glu, Met, Gln, Lys, Cys, Arg, Trp, His, Asp

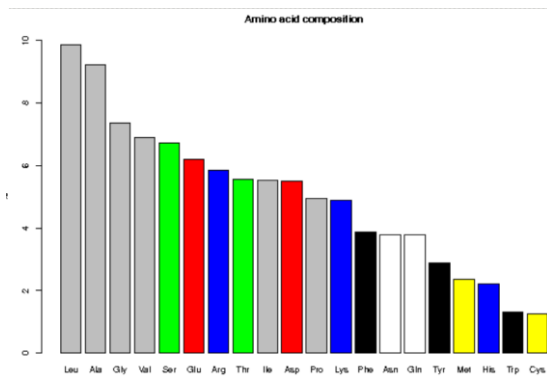


Figure 2: Amino acids vs percentages of Amino Acids bar graph. Taken from <https://www.ebi.ac.uk/uniprot/TrEMBLstats>

Part 2 - Plotting an Hydropathy Plot [60 pts.]

2- When we set window size from 5 to 20, fluctuations of the scores get rare although the overall hydropathy score doesn't change as in the figures 3 and 4. Instead of random fluctuations, we can clearly see the increases and decreases in score. Furthermore, it becomes easy to see the regions where the hydropathy score is higher. With that information, we can comment about the hydrophobicity of that region.

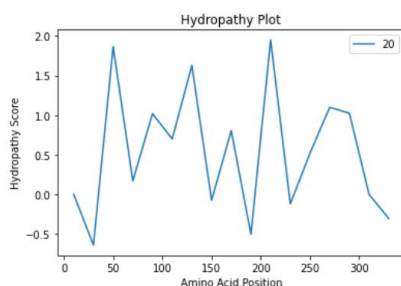


Figure 3: Hydropathy plot with window size 20

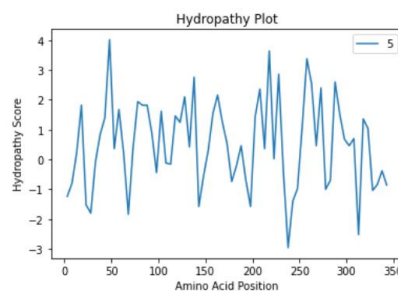


Figure 4: Hydropathy plot with window size 5

3- Transmembrane helices start from positions 37 to 61, 74 to 96, 111 to 133, 153 to 173, 203 to 224, 253 to 274, 285 to 309 as shown in figure 5. When we look at the plot graph in figure 3, these intervals are where the hydropathy score got higher. We know that high hydropathy score points to high hydrophobicity. Thus, we can conclude that the membrane is hydrophobic. As the hydrophobic tails of phospholipids left inside the phospholipid bilayer of the membrane, transmembrane parts of the protein should be hydrophobic regions of that protein.

Feature key	Position(s)
Topological domain ¹	1 – 36
Transmembrane ¹	37 – 61
Topological domain ¹	62 – 73
Transmembrane ¹	74 – 96
Topological domain ¹	97 – 110
Transmembrane ¹	111 – 133
Topological domain ¹	134 – 152
Transmembrane ¹	153 – 173
Topological domain ¹	174 – 202
Transmembrane ¹	203 – 224
Topological domain ¹	225 – 252
Transmembrane ¹	253 – 274
Topological domain ¹	275 – 284
Transmembrane ¹	285 – 309
Topological domain ¹	310 – 348

Figure 5: Topology of Rhodopsin. Taken from Uniprot.