***Practicals – 8***

*-BS19B032*

*-R. Vasantha Kumar*

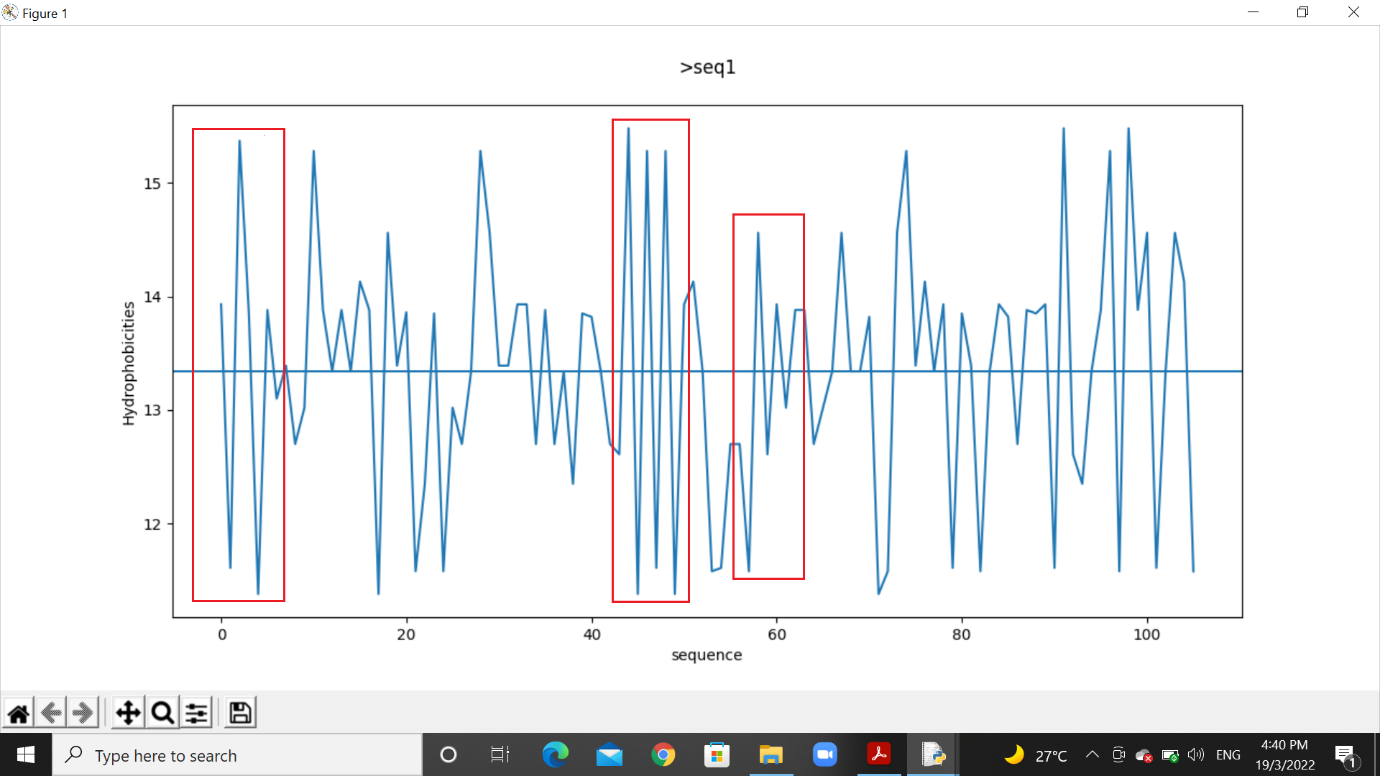
1) I wrote a code to plot hydrophobicity profile for the sequences in (Q1.fasta).

We can identify the alpha helices and beta strands from the hydrophobicity plots.

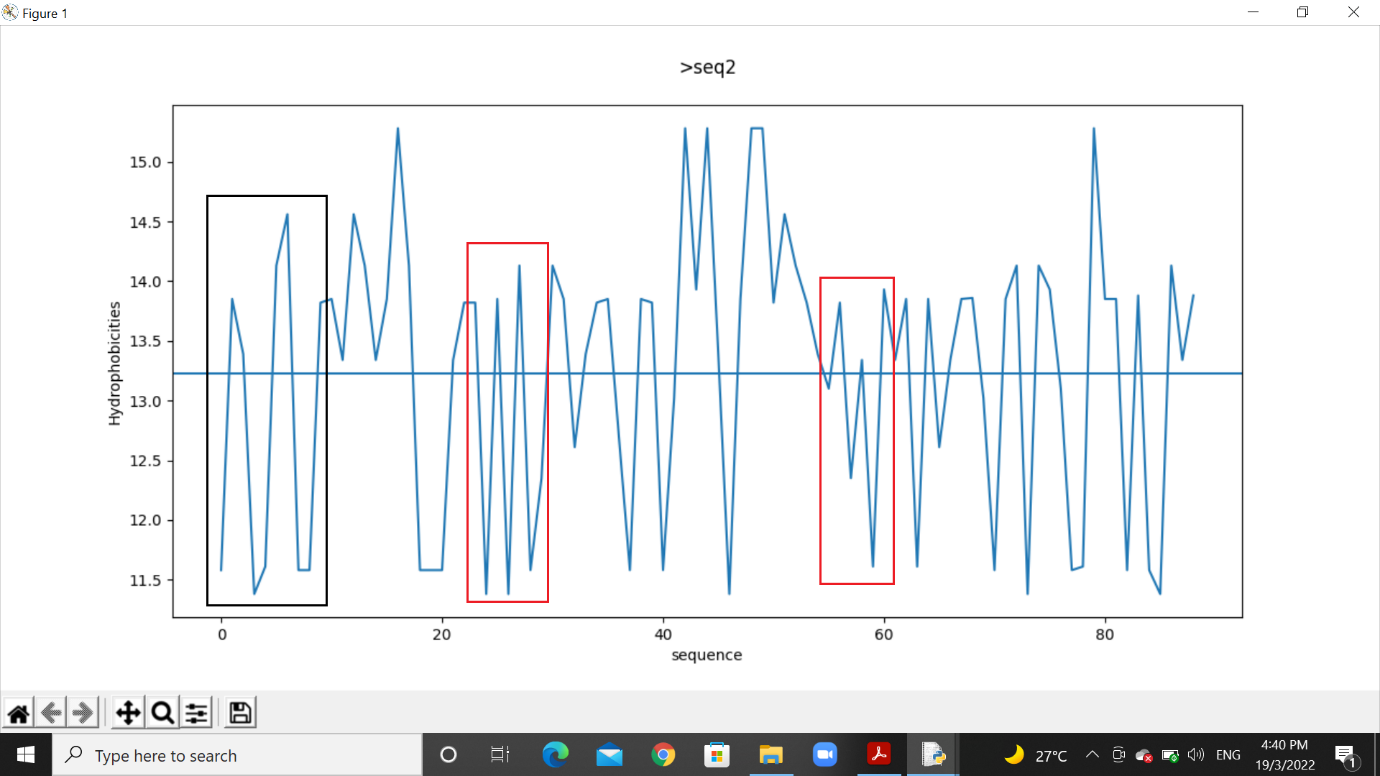
When a pair of amino acids has low hydrophobicity, is followed by a pair of amino acids with high hydrophobicity, it can form alpha helices. I took the verification length as 8 amino acids, i.e.., 4 pairs.

Similarly, when an amino acid with low hydrophobicity is followed by an amino acid with high hydrophobicity, it can form beta strands. I took the verification length as 6 amino acids.

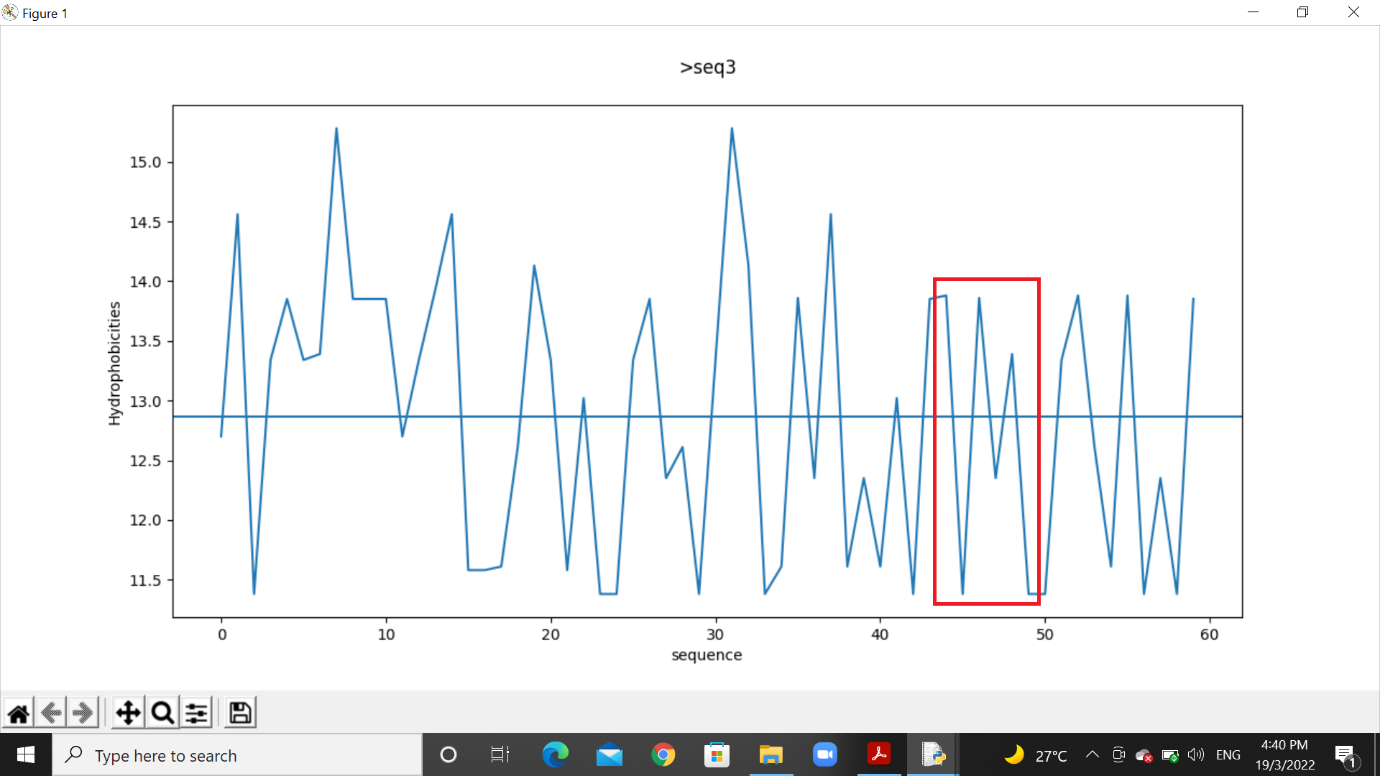
Hydrophobic plot of seq 1:



Hydrophobic plot of seq 2:



Hydrophobic plot for seq 3:



Here, the red rectangle indicates a beta strand segment, and, black rectangle indicates alpha helix segment.

2) I found the amphipathicity of alpha helices and beta strands using these formulas:

Amp. of helix = (∑hi+j)/n

Amp. of strands = (∑hi+j)/n

Where,

hi+j = hydrophobicity of amino acid residue

n = number of residues in helix or strand

Seq 1:

Strand 1 = (13.93+11.61+15.37+13.85+11.38+13.88)/6

= 13.33

Strand 2 = (12.61+15.48+11.38+15.28+11.61+15.28)/6

= 13.60

Strand 3 = (11.58+14.56+12.61+13.93+13.02+13.88)/6

= 13.26

Seq 2:

Helix 1 = (13.85+13.39+11.38+11.61+14.13+14.56+11.58+11.58)/8

= 12.76

Strand 1 = (13.82+11.38+13.85+11.38+14.13+11.58)/6

= 12.69

Strand 2 = (13.10+13.82+12.35+13.34+11.61+13.93)/6

= 13.025

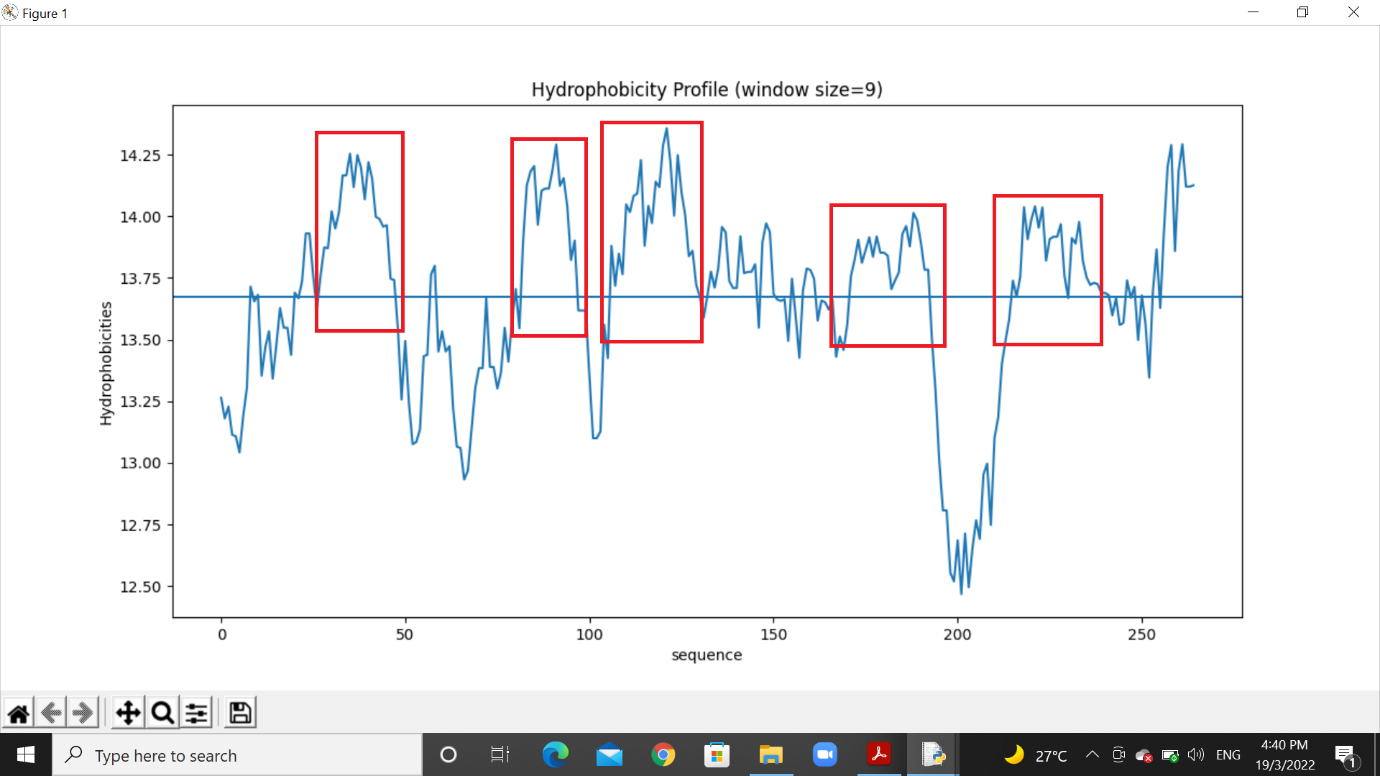
Seq 3:

Strand 1 = (13.88+11.38+13.86+12.35+13.39+11.38)/6

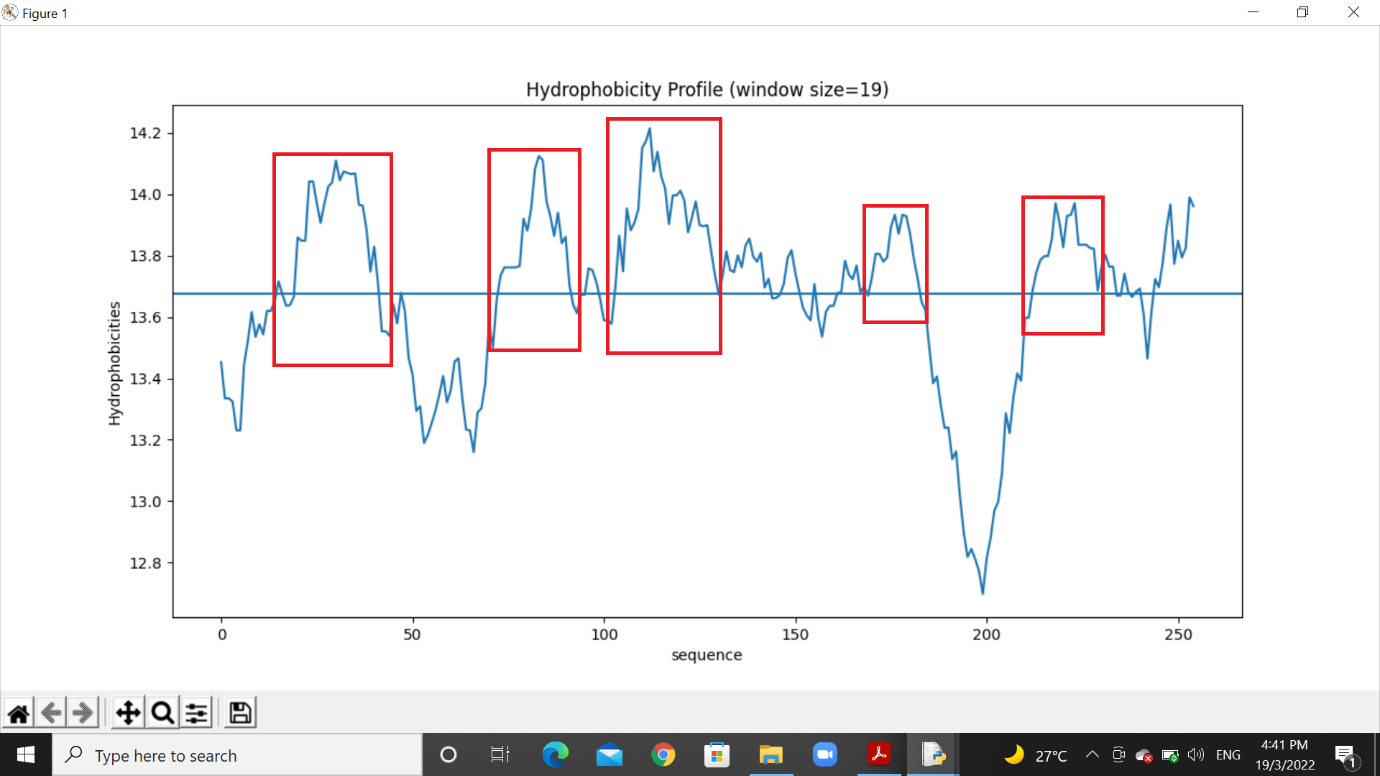
= 12.70

3) I wrote a code to plot hydrophobicity profile for the sequence (Q2.fasta) with window lengths 9 and 19. I have attached it in submission. (Q3\_profile)

Plot for window size = 9:



Plot for window size = 19:

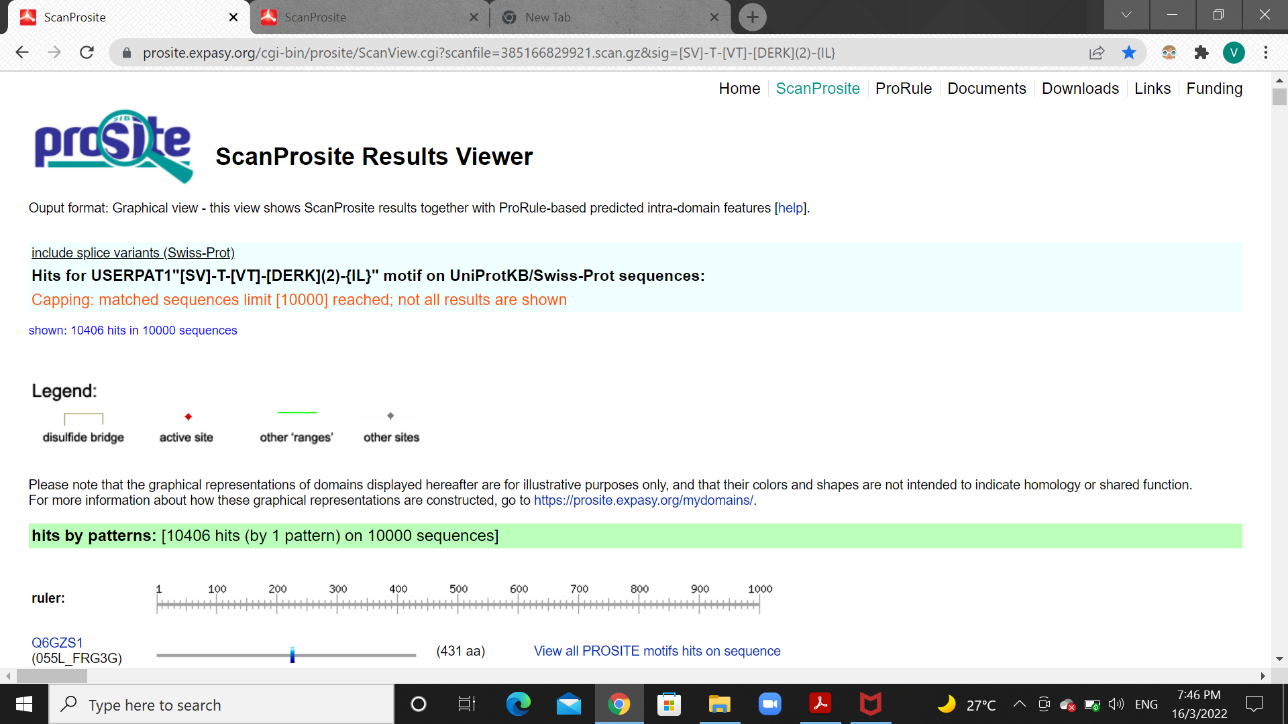


The regions inside the red rectangle are transmembrane segments, which are highly hydrophobic in nature. There are a total of 5 transmembrane segments in the given query sequence.

4) Using the ScanProsite tool, I found the number of matches for each sequence in UniProt-KB(including Swiss-Prot and isoforms).

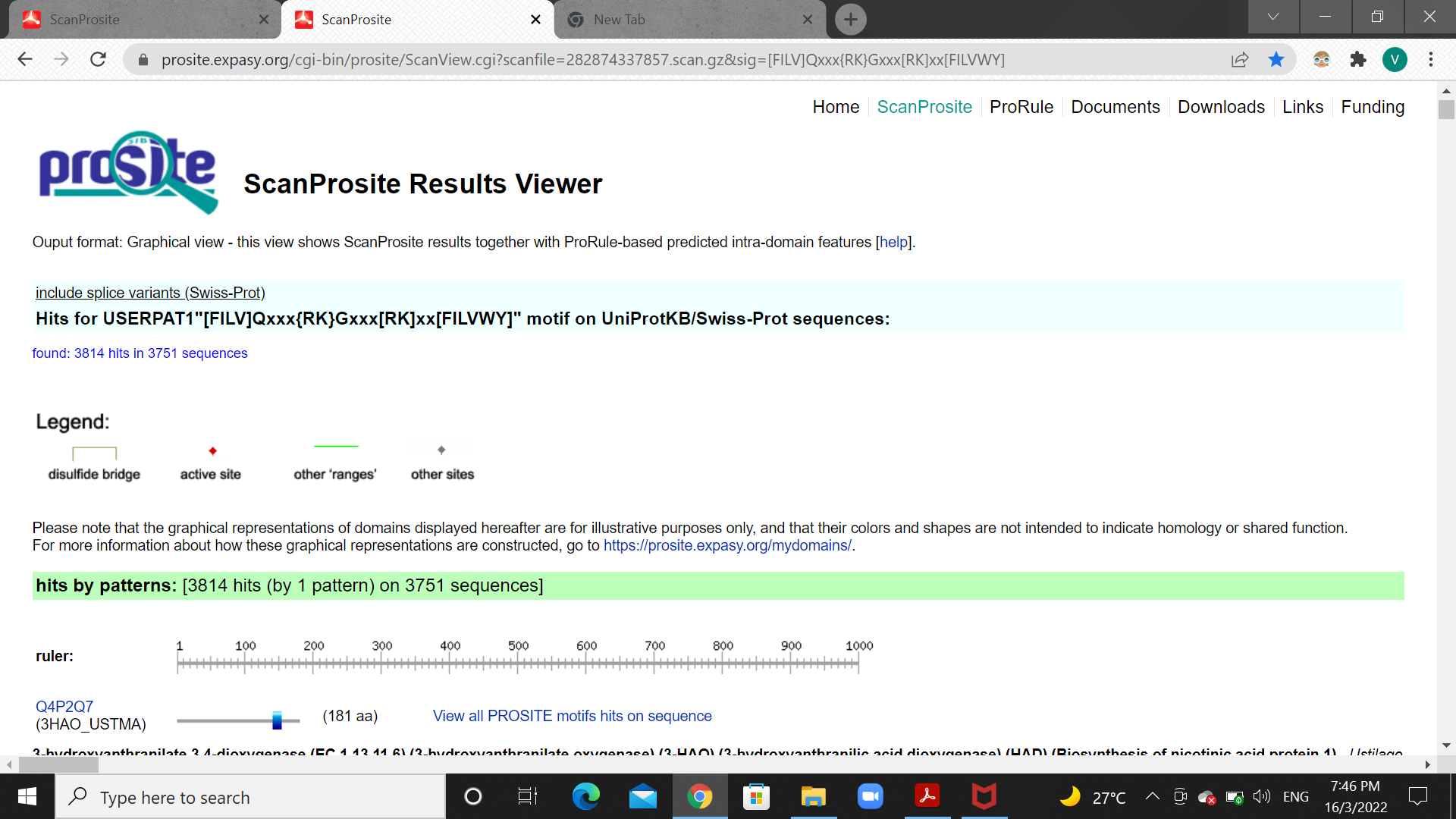
Pattern (a): [SV]-T-[VT]-[DERK](2)-{IL}

The number of matches found was 10,406 in a total of 10,000 sequences.



Pattern (b): [FILV]Qxxx{RK}Gxxx[RK]xx[FILVWY]

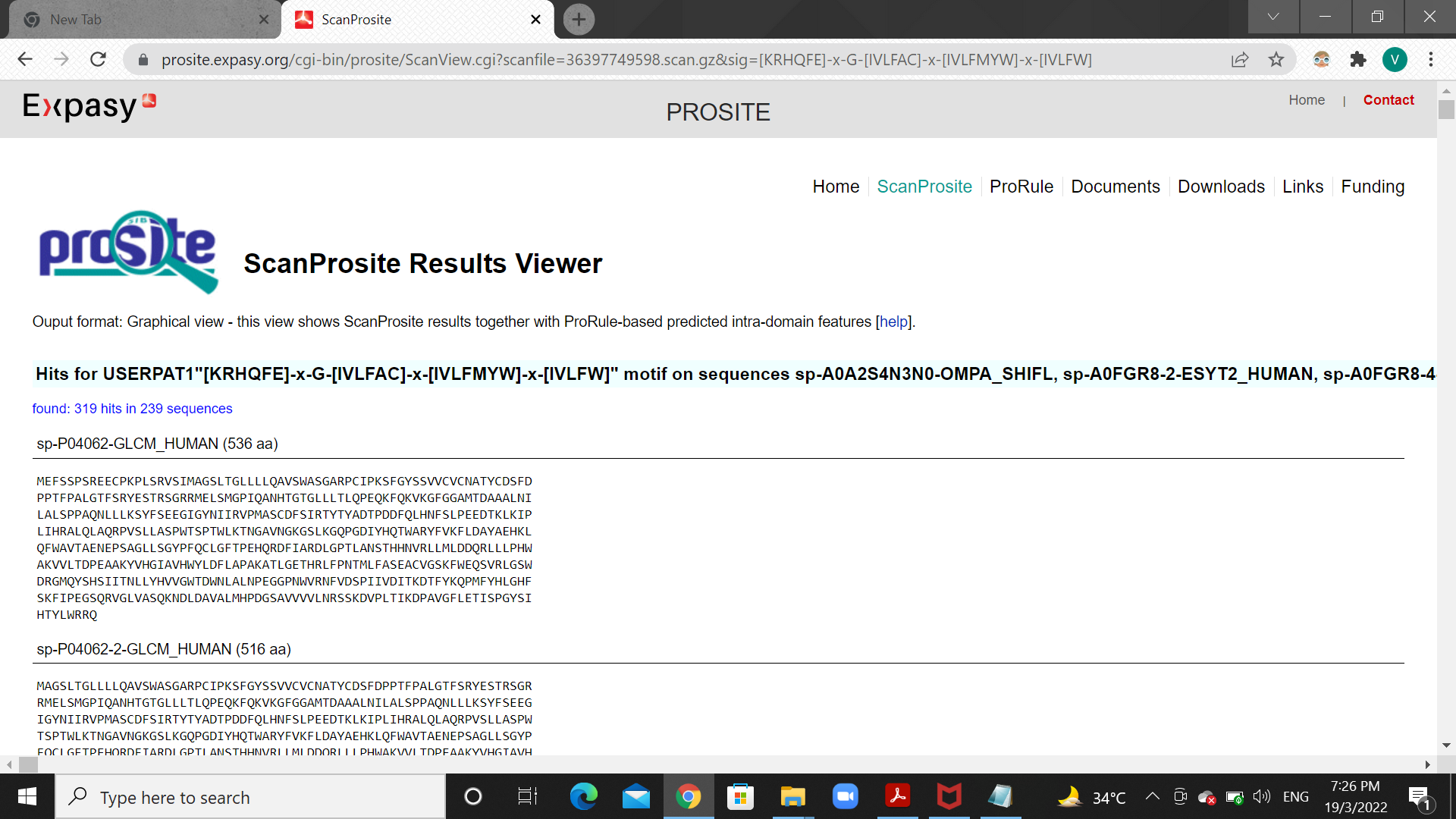
The number of matches found was 3,814 in a total of 3,751 sequences.



5) I wrote a program to identify the patterns, given in Q4, to the sequence database (Q4.fasta). I have attached the codes in submission. (Q5\_(a) & Q5\_(b))

Using the program, I listed the matches along with the sequence header and location of the matches in the sequence. I have also attached the match files in submission. (Q5\_(a)\_matches & Q5\_(b)\_matches)

6) I collected beta barrel membrane protein sequences from Uniprot-KB(Swiss-Prot). I have attached the sequences as txt file in submission (Q6\_betabarrel\_sequences). Then I used the ScanPrositeTool to match these sequences with the given pattern: [K,R,H,Q,F,E]-x-G-[I,V,L,F,A,C]-x-[ I,V,L,F,M,Y,W]-x-[ I,V,L,F,W]



The total number of matches is 319, in a total of 239 sequences.