

# Bioinformatics Tools Assignment

Name: Laiba Suleman

Program: BS 7<sup>th</sup> (M)

Reg: 04032213027

## DNA Sequence Used

```
CAAACCTGCACATGAATTGCACGTTGACCTCTAGTAATCGGTAGCCGAGGACTTGGC  
AAGATTAAAAGGGCTTCTGGTAGTGTGAGCGACTATTGTGCTCATCATGTGCAA  
TGT CCT ATT CTC ATT GT GAAG CC ACC CA AGG AAA ACC CCA AA AGC AG CT A ATT AAC  
AAGGATATATCATGTGCCAGTCTATGAACAAATGTCATTACTGATGATGAATAAAC  
ACAAGTTCATTGAGGATTATAAAACTTATGTATATATTAAATTATATTCAATAAAAT  
GGAGGTTGTATCTGTTGTCTCAATAGCTAGCCCAGTCCACAGGCAATAGGTTTTGG  
GTAAAAGCCAGTCGTACAACACTAGTGAATCTGAATATTGACCCTCATTCTCTGG  
TAACGCAAGTTGCTGCAGCTTATGTGGCTCTGTATTGGTTAATAATATAACAA  
TTC AATTAGAGATATCTAGCACTCACCTCGATGGAACAAA
```

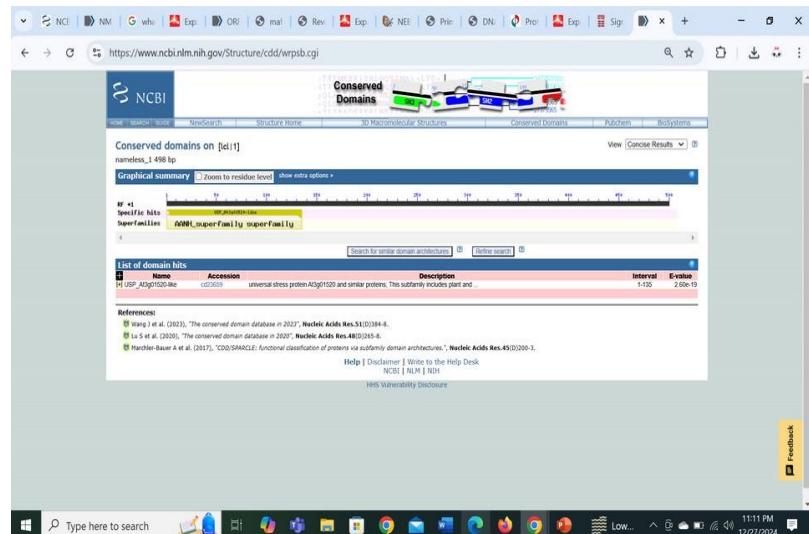
## BLAST

BLAST results show similar sequences on 12 subjects and a graphic summary of the top hits.

The screenshot shows a web-based peptide sequence analysis tool. At the top, there's a navigation bar with various icons. Below it is a search bar with the URL <https://pepdraw.com>. The main interface has a dark header with the text "PEP DRAW" and "A tool to draw peptide primary structure and calculate theoretical properties". Below the header, there's a text input field labeled "Enter sequence here" containing the sequence: CAAACCTGCACATGAATTGCACGTTGACCTCTAGTAATCGGTAGCCGAGGACTTGGC... . To the right of the sequence input is a "Draw Peptide" button and a "Help" link. On the left, there's a "Peptide Structure" section with a "Peptide Keypad" and two tabs: "Natural Amino Acids" and "Special Amino Acids". The "Natural Amino Acids" tab is selected, showing buttons for Alanine (A), Arginine (R), Asparagine (N), Aspartic acid (D), Cysteine (C), Glutamine (Q), Glutamic acid (E), Histidine (H), Isoleucine (I), Leucine (L), Lysine (K), Methionine (M), Phenylalanine (F), Proline (P), Serine (S), Threonine (T), Tryptophan (W), Tyrosine (Y), and Valine (V). The "Special Amino Acids" tab shows buttons for Proline (Pro), pTyr (pT), pTyr (pY), and Allys (Alk). Below these are buttons for N-terminal: NH<sub>2</sub>, C-terminal: COOH, and pH: 7.0. To the right of the keypad, there's a "Peptide properties" section with fields for "Sequence" (containing the entered sequence), "Length" (80), "Mass" (4937.7500), "Isoelectric point (pI)" (4.73), and "Net charge" (-1). It also lists hydrophobicity (-36.87 Kcal \* mol<sup>-1</sup>), extinction coefficient<sup>a</sup> (875 M<sup>-1</sup> \* cm<sup>-1</sup>), and extinction coefficient<sup>b</sup> (0 M<sup>-1</sup> \* cm<sup>-1</sup>). At the bottom, there's a "Peptide Structure" visualization showing the sequence as a zigzag line, and a taskbar at the very bottom of the screen.

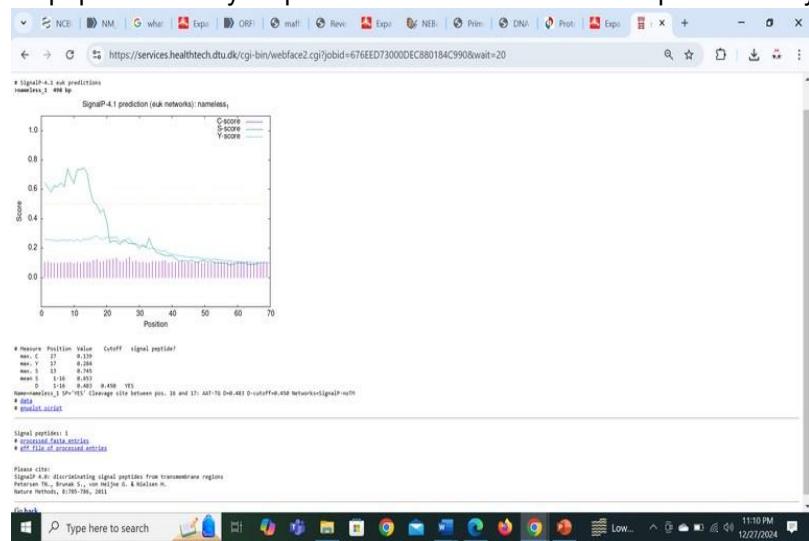
## Hydropathy Plot

The hydropathy plot, which shows the hydrophobicity of amino acids on protein sequence.



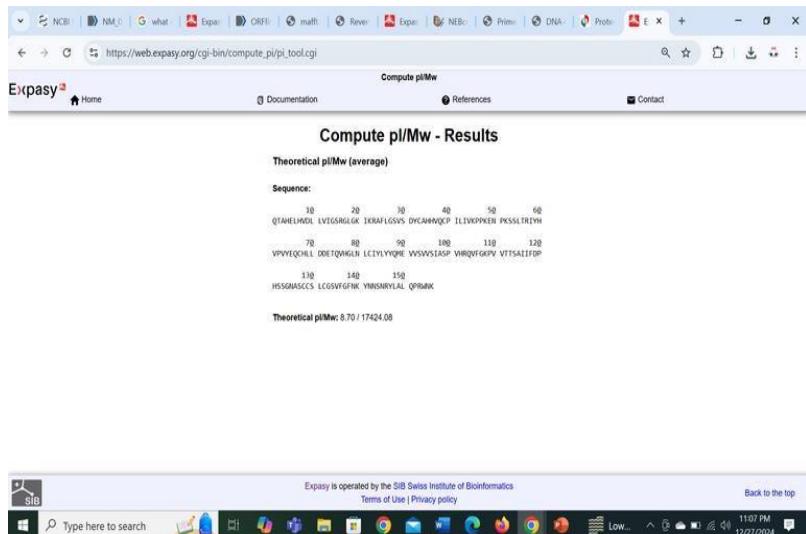
## PEP DRAW & Reverse Complement

This is structure of peptide from my sequence and I found reverse compliment of my sequence.



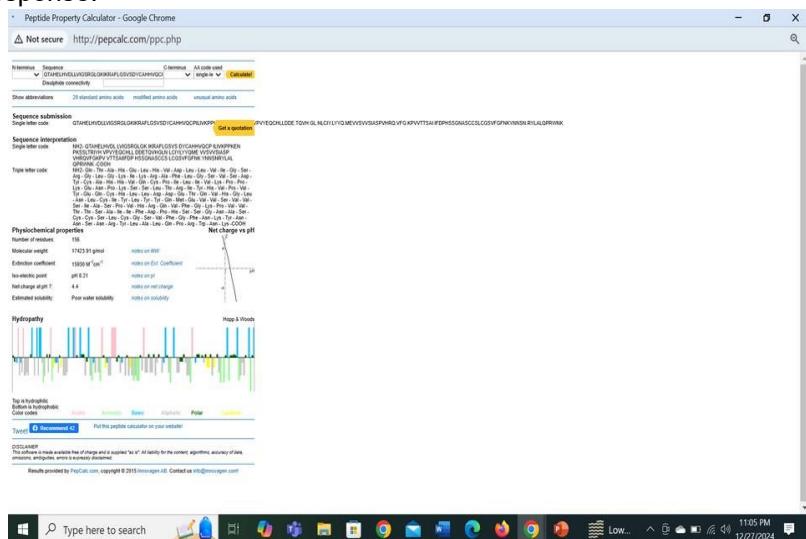
## ORF Finder & Signal P 4.1

The analysis shows four potential open reading frames, with the largest spanning 182-391 nucleotides, and predicts a signal peptide with a cleavage site between positions 16 and 17.



## Conserved Domain (CDD)

The analysis identified a conserved domain, HUP A, in the input sequence, suggesting a potential role in stress response.



DNA Stats

The analysis shows that the 498 bp linear sequence has 38% GC content and 62% AT content and can be cut by 19 different restriction enzymes, including both frequent and rare cutters.

**Transcription and Translation Tool**

Converts sequences from DNA to RNA and from this to protein.

**DNA (deoxyribonucleic acid)**

- permanent copy of the genetic information
- uses "T" rather than "U"
- without 2' OH group
- more stable than RNA
- lower error rate during replication than RNA

**RNA (ribonucleic acid)**

- messenger RNA (mRNA) is a temporary copy of the sequence of the gene that codes for the protein

**Protein**

- consists of amino acids linked by amide bonds ("peptide bonds")
- most enzymes and many structural components in cells are proteins

DNA sequence: clear  
RNA sequence: clear  
protein sequence: clear

Help with amino acid abbreviations

## Expasy Translate Tool

The Expasy Translate tool translated the nucleotide sequence into six possible protein sequences based on different reading frames.

**Graphical View**

Display: Circular, Linear, Alternative

Cleavage: 1 cutters, 2 cutters, 3 cutters, 1' Cuts 1 strand, 1' Extension, 3' Extension

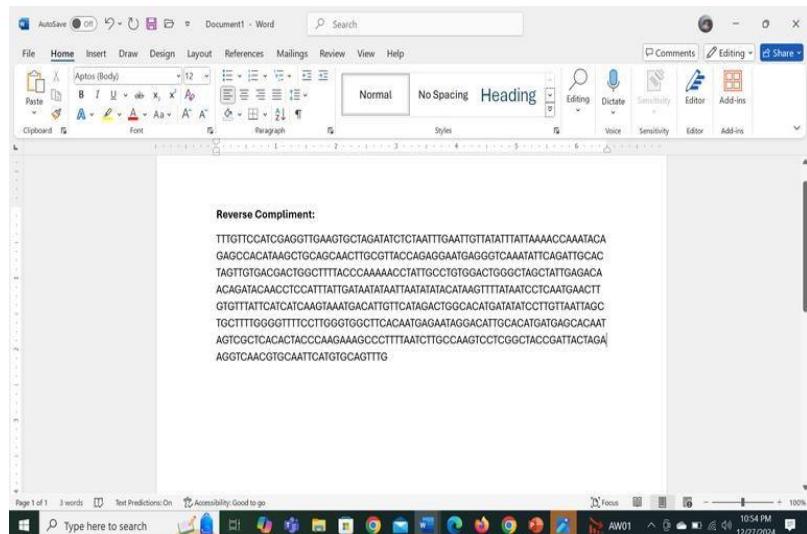
Supplier: NEB, Other supplier, Not commercially available

Site: Affected by CpG methylation, Affected by other methylation, Ambiguous site

Results For: Enzymes, ORFs, Sequence Information: 498 bp, Linear, GC = 38%, AT = 62%

## Primer 3 & Transcription/Translation

It shows primer pairs, pair 1 is good pair and below I use Transcription and translation tool to analyze Amino Acid sequence from DNA Sequence.



## Expasy Compute pi/Mw (ProtParam)

The Expasy Compute pi/Mw tool calculated the (pi) of the protein : 8.70 and (Mw) to be 17424.08 Da.

Label	Strand	Frame	Start	Stop	Length (nt/aa)
ORF1	+	2	182	310	129/42
ORF3	-	1	390	304	87/28
ORF2	+	3	114	197	84/27
ORF4	-	1	108	31	78/25

## Additional Bioinformatics Tools

### STRING Tool

The STRING tool is used to predict protein–protein interaction networks, helping to understand functional associations between proteins derived from the analyzed sequence.

## **CELLO Tool**

The CELLO tool predicts the subcellular localization of proteins based on their amino acid composition, providing insight into where the translated protein may function inside the cell.

## **RNAfold Tool**

The RNAfold tool predicts the secondary structure of RNA sequences using minimum free energy models, helping to understand RNA stability and folding patterns derived from the DNA sequence.