From DNA Blueprint to Protein Properties: Insights Through Bioinformatics Tools
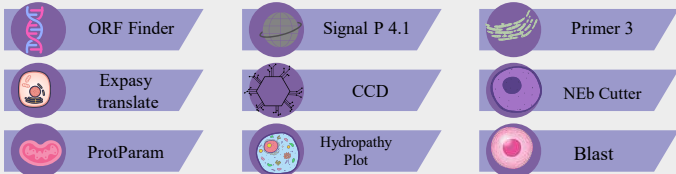
Presented By : Laiba Suleman
BS Biochemistry 7th M
04032213027
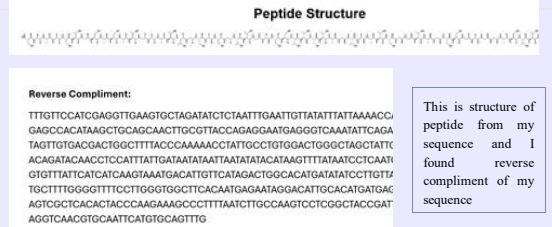
QUAID-I-AZAM UNIVERSITY ISLAMABAD

DNA Sequence

CAAACTGCACATGAATTGCACGTTGACCTTCTAGTAATCGGTAGCCGAGGACTTGGC
AAGATTAAAAGGGCTTTCTTGGGTAGTGTGAGCGACTATTGTGCTCATCATGTGCAA
TGTCCTATTCTCATTGTGAAGCCACCCAAGGAAAACCCCAAAAGCAGCTAATTAAC
AAGGATATATCATGTGCCAGTCTATGAACAATGTCATTTACTTGATGATGAATAAAC
ACAAGTTCATTGAGGATTATAAAACTTATGTATATATTAATTATATTATCAATAAAT
GGAGGTTGTATCTGTTGTCTCAATAGCTAGCCCAGTCCACAGGCAATAGGTTTTTGG
GTAAAAGCCAGTCGTCACAACTAGTGCAATCTGAATATTTGACCCTCATTCCTCTGG
TAACGCAAGTTGCTGCAGCTTATGTGGCTCTGTATTTGGTTTTAATAAATATAACAA
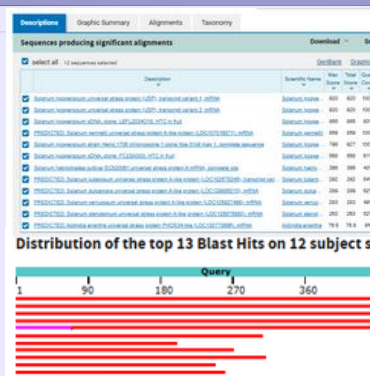TTCAAATTAGAGATATCTAGCACTTCAACCTCGATGGAACAAA

Abstract

This study aimed to characterize a given DNA sequence using a suite of 14 bioinformatics tools. We employed BLAST, Expasy, and NCBI resources to analyze the sequence for potential protein coding regions, predict protein properties, and identify conserved domains. The analysis revealed four potential open reading frames (ORFs) and predicted a protein with a theoretical isoelectric point (pI) of 8.70 and a molecular weight (Mw) of 17424.08 Da. The hydropathy plot provided insights into the protein's potential interaction with water. These findings offer a preliminary characterization of the protein encoded by the DNA sequence.

Conclusion

Analysis revealed four ORFs, the largest spanning 182-391 nucleotides, and six potential protein translations. The predicted protein has a pI of 8.70, MW of 17424.08 Da, and a potential signal peptide. Hydropathy analysis suggests its interaction with water. The DNA sequence is cuttable by 19 restriction enzymes, aiding cloning. These findings provide a preliminary characterization of the encoded protein and guide further research.

Methodology

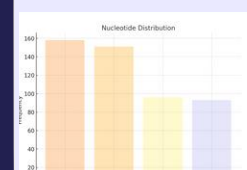- ORF Finder
- Signal P 4.1
- Primer 3
- Expasy translate
- CCD
- NEb Cutter
- ProtParam
- Hydropathy Plot
- Blast

PEP DRAW & Reverse compliment

Peptide Structure

Reverse Compliment:

TTTGTTCCATCGAGGTTGAAGTGCTAGATATCTCTAATTTGAATTGTTATATTTATTAAAACCA
GAGCCACATAAGCTGCAGCAACTTGCGTTACCAGAGGAATGAGGGTCAAATATTCAGA
TAGTTGTGACGACTGGCTTTACCCAAAAACCTATTGCCTGTGGACTGGGCTAGCTATTG
ACAGATACAACCTCCATTTATTGATAATAATTAATATATACATAAGTTTATAATCCTCAAT
GTGTTTATTCATCATCAAGTAAATGACATTGTTCATAGACTGGCACATGATATATCCTTGTTA
TGCTTTTGGGGTTTTCCTTGGGTGGCTTCACAATGAGAATAGGACATTGCACATGATGAC
AGTCGCTCACACTACCCAAGAAAGCCCTTTTAATCTTGCCAAGTCCTCGGCTACCGAT
AGGTCAACGTGCAATTCATGTGCAGTTTG

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

BLAST & Hydropathy plot

Distribution of the top 13 Blast Hits on 12 subject sequences

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

Hydropathy
Hopp & Woods

Top is hydrophilic
Bottom is hydrophobic
Color codes
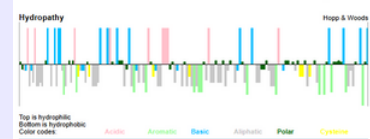
The hydropathy plot, which shows the hydrophobicity of amino acids on protein sequence

ORF FINDER

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

Nucleotide Distribution

DNA Stats

SIGNAL P4.1

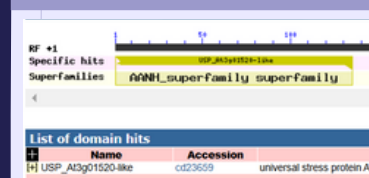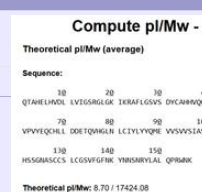SignalP-4.1 prediction (euk networks): nameless:

- C-score
- S-score
- Y-score

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 Domains

RF +1
Specific hits
Superfamilies   AANH_superfamily superfamily

List of domain hits
Name	Accession
USP_At3g01520-like	cd23659	universal stress protein At3

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

Expasy compute PI/MW

Compute pI/Mw -

Theoretical pI/Mw (average):

Sequence:

```
        10         20         30         40
QTAHELHVDL LVIGRGLGK IKRAFLGSVS DYCAHHVQCI

        70         80         90        100
VPVYEQCHLL DDETQVHGLN LCIYLYYQHE VVSWVSIASI

       130        140        150
HSSGNASCCS LCGSVFGFNK YNNSARYLAL QPR6HK
```

Theoretical pI/Mw: 8.70 / 17424.08

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

PRIMER 3

DNA sequence: clear
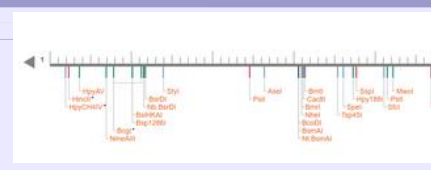
RNA sequence:

transcription   reverse transcription

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 TOOL

5'3' Frame 1
5'3' Frame 2
5'3' Frame 3
3'5' Frame 1
3'5' Frame 2

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

NEB Cutter

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.