**Question1:**

Performing BLASTp through NCBI using default parameters

A screenshot of a computer

Description automatically generated

Fig1: BLASTP for query sequence

This sequence has 378 amino acids.

a

**A screenshot of a computer

Description automatically generated**

Fig 1: blast hits of the accession id’s

The sequence Q4R804.1, can possibly be considered as a true hit as the query cover shows 98% and its percentage identity are 69.00% the accession length is 368 amino acids as we compare it with the query sequence Q8K2X3.2 which has an accession length of 378 amino acids.

The sequence Q9LMK5.1, is not a true hit because the query cover shows 41% compared to the other sequences as the results show an accession length of 160 amino acids and percentage identity is 33.54%.

Graphic summary A screenshot of a computer

Description automatically generated

Fig2: blast hits of the summary

b.

Here in the graphic summary, we can conclude that the sequence Q4R804 and query sequence Q8K2X3 in the figure above show shared ancestry with highest similarity indicating a true hit. However, sequence Q9LMK5 shows very less similarity levels and it is certainly not enough to be considered as shared ancestry.

A screenshot of a computer

Description automatically generated

Fig3: conserved top blast hits

A screenshot of a computer

Description automatically generated

Fig4: Domains on Q4R804

A screenshot of a computer

Description automatically generated

Fig5: domains on Q9LMK5

The query sequence Q8K2X3 and sequence Q4R804 are orthologs because they both share the same functional domains that are STN1\_2 and hOBFC1\_like, this indicates that they share the same function however in different species. Also query sequence and Q9LMK5 share only one functional domain hOBFC1\_like. All three sequences are share the function of supressing cdc thirteen in different species and are thus orthologs.

2.

A screenshot of a computer

Description automatically generated

Fig6: BLASTn hits of the unknown sequence

A screenshot of a computer

Description automatically generated

Fig7: annotation results of the top blast hits

In the annotation results we summarise the vase tunicate Ciona intestinalis produces "cionin" (X81378.1). The intron II-associated gene is 57 base pairs long. The annotation describes the gene's DNA and Ciona intestinalis parent organism. The report for the gene is unpublished. The gene is located at position 317 in the cionin cDNA with accession number X69130. The annotation emphasises that the gene comes from adult Ciona intestinalis stomach tissue and is developing. This detailed annotation interprets the gene's properties, helping us understand its significance in the organism's genome and evolutionary background.

1. No corresponding protein sequence is available in the uniprot database.

A screenshot of a computer

Description automatically generated

Fig8: BLASTx results of the known sequence

1. In blastn results the first top hit was found to be with 100% query cover as well as the % identity in conclusion, the BLASTn search found the "cionin" gene in Ciona intestinalis using a nucleotide sequence query, revealing its characteristics and evolutionary context. While blastn was successful in giving similar nucleotide sequences, blastx is used to find the corresponding aminoacid sequence of the query in a protein sequence database, which was unavailable in this case.

(Refer figure 7 & 8)

3. a. Despite a percentage identity of only 41.24%, the top hit has query coverage of 97% and thus homology modelling is possible and acceptable. The alignment coverage imply that the two sequences might have a similar structure.

Statistics:

Query sequence

Per. identity: 42.42%

e-value: 6e-16e

query cover: 100%

acc. Length: 378 aa

template sequence:

per. Identity: 41.24%

e-value: 2e-16

query cover: 97%

acc. Length: 368aa

Given that homology modelling assumes sequences with significant similarity share a similar structure and function, the query sequence's strong coverage with the template sequence suggests that the modelled structure may accurately represent the query sequence's 3D structure.

Based on the template sequence's structure, homology modelling may accurately anticipate the query sequence's structure. The sequence identity and structural similarities suggest function conservation and a robust foundation for modelling.

A screenshot of a survey

Description automatically generated

Fig9: BLASTp results of the target sequence

b

A close-up of a computer code

Description automatically generated

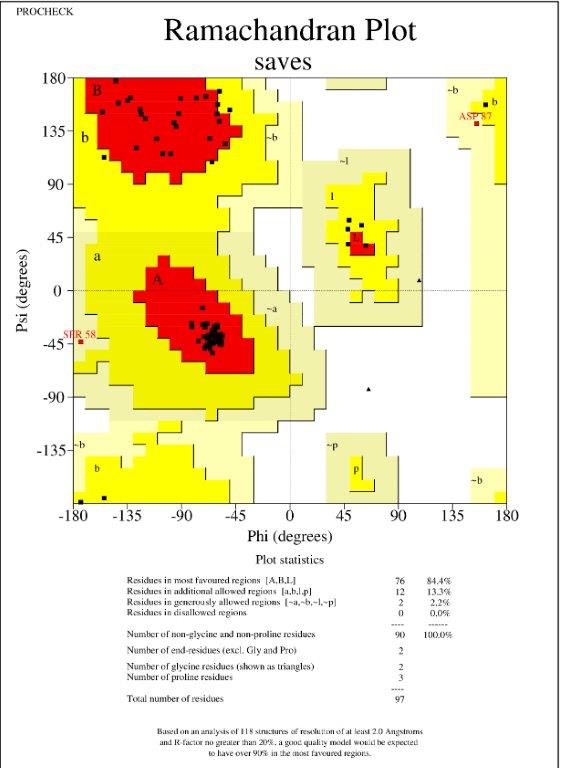
Fig10: alignment file

c

A computer screen shot of a program

Description automatically generated

Fig11: python file

 A diagram of a graph

Description automatically generated with medium confidence

Fig12: Ramachandran plot for model 1. Fig13: Ramachandran plot for model 2.

1. To assess the dependability of two Ramachandran plots with residue rates of 84% and 87%, consider their backgrounds and features. Residue rates show the Ramachandran plot's amino acid residue. Higher numbers are more consistent with molecular principles and protein structures; therefore, the plot with 87% is more feasible. With more residues as the residues are in the allowed region and they represent proteins psi and phi angles which are realistic, a Ramachandran plot is more accurate, but sample size, structural context, outliers, and protein-specific properties must also be considered (Hollingsworth & Karplus, 2010).

e

A graph of a green line

Description automatically generated with medium confidence A graph of a graph of a graph

Description automatically generated with medium confidence

Fig14: pdb model 1

Fig15: pdb model 2

Local model quality plot for both models indicate that they align well with knowledge-based energy expectations and have similar graphs exhibiting favourable that is, (negative values) for local energy at different amino acid positions

A screenshot of a graph

Description automatically generatedA screen shot of a graph

Description automatically generated

Fig15: ProSA 1st model Fig16: ProSA 2nd model after realignment

The results interpret:

1st model z-score: -6.8

2nd model z-score: - 7.27

According to Z-scores, both models have energy profiles that match well-folded protein structures. The second model's lower Z-score after realignment shows structural and energy stability improvements. This supports the idea that lower Z-scores imply better protein structure.

1. 1. A structure with a good Ramachandran plot may not score highly by ProSA. The Ramachandran plot evaluates amino acid by measuring their phi and psi angles. but the ProSA score compares the protein structure's energy to a hypothetical energy distribution. A structure with a strong Ramachandran plot may have energy-related difficulties or local interactions that lower its ProSA score.

No, a structure with a high ProSA score (showing low energy and stability) may not have a good Ramachandran plot. The ProSA score assesses the protein structure's energy profile, including non-local and global interactions. However, the Ramachandran plot concentrates on amino acid phi and psi angles. A structure with a strong ProSA score may have unfavourable dihedral angles, resulting in a poor Ramachandran plot (Hollingsworth & Karplus, 2010).

# References

Hollingsworth, S. & Karplus, P., 2010. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins.. *Biomol Concepts.,* 1(3-4), pp. 271-283.