

# Term Project – Part I

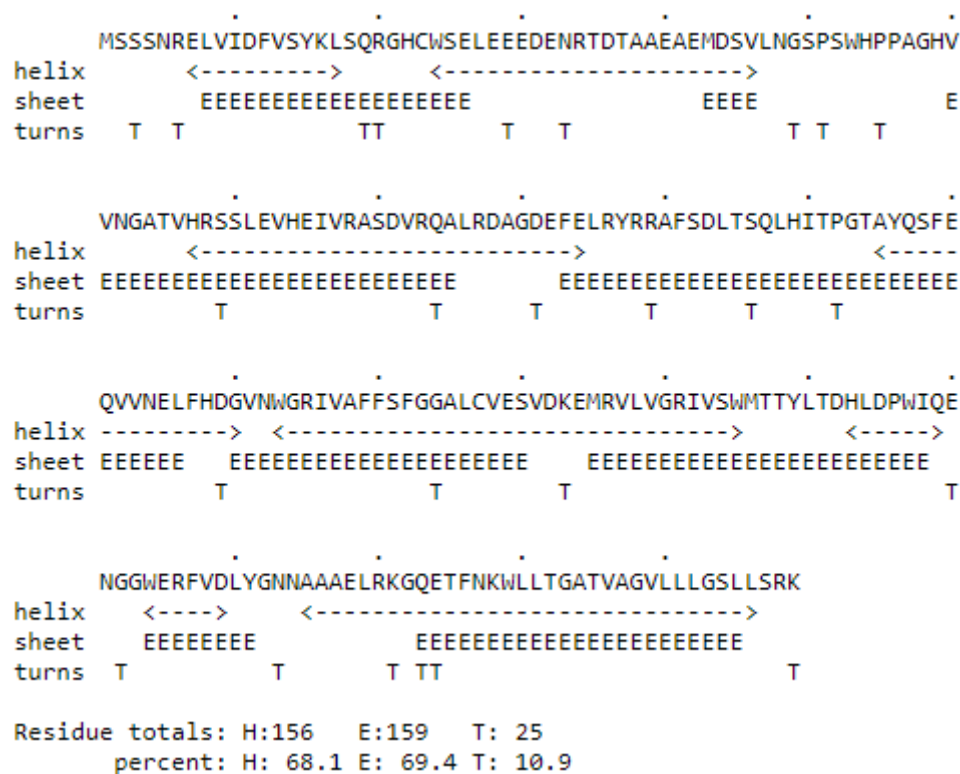
## Unidentified structured protein

Protein name is Bcl-2-like protein 1, and its gene is BCL2L1, and it is from Gallus gallus (chicken). Uniprot ID number is Q07816, B2CL1\_CHICK. FASTA sequence of the gene is given below.

```
MSSSNRELVIDFVSQKLSQRGHCWSELEEEEDENRTDTAAEAEMDSVLNGSPSWHPPAGHV  
VNGATVHRSSLEVHEIVRASDVRQALRDAGDEFELRYRRAFSDLTSQLHITPGTAYQSFE  
QVNVNELFHDGVNWGRIVAFFSFGGALCVESVDKEMRVLVGRIVSWMTTYLTDHLD  
PWIQENGWERFVDLYGNNAAEELRKQETFNKWLTTGATVAGVLLLSLRK
```

## Secondary structure predictions

**Chou-Fasman Prediction:** It predicts the secondary structure of the protein, and percentages are 68.1% for helices, 69.4% for strands, and 10.9% for turns.



**Figure 1:** Chou-Fasman prediction result for Gallus gallus

**GOR Method for prediction:** It predicts, and percentages of secondary structure are coil percentage is 10.04, helix percentage is 40.61, strand percentage is 27.94, and turn percentage is 21.39. These results are calculated by using simple python code, which is:



## Possible Functions of the protein

It is known as dominant regulator for apoptotic cell death. There are two form of the protein that are long and short forms. Displaying repressor activity for cell death belongs to long form, and promoting apoptosis is for short form. Other than those, regulation of G2 checkpoint and mitosis cytokinesis progression are responsibilities of the protein (UniProt). Additionally, BCL-2 proteins regulate neuronal activity, calcium handling, autophagy, mitochondrial dynamics, which are physiology of normal cells (Hardwick, J. M.; Soane, L., February 2013).

## Homologous Proteins to our protein Bcl-2-like protein 1

According to BLAST results from both Uniprot, there are four homologous proteins from *Mus musculus*, *Homo sapiens*, and *Rattus norvegicus*. Their sequences are not identical, and for both *Mus Musculus* and *Homo sapiens*, 3D structure is predicted and available in PDB.



**Figure 4:** Homolog proteins of Bcl-2-like protein 1 from *Gallus gallus*

### Bcl-2-like protein 1 from *Mus musculus* (PDB ID: 1PQ1)

FASTA sequence:

```
MSQSNRELVVDFLSYKLSQKGYSWSQFSDVEENRTEAPEETEARETPSAINGNPSWHLADSPAVNGATGHSSSLDAREVIPMAAVKQAL
REAGDEFELRYRRAFSDLTSQLHITPGTAYQSFEQVVNELFRDGVNWGRIVAFSFGGALCVESVDKEMQVLVSRIASWMATYLNHLEP
WIQENGWDTFVDLYGNNAAESRKGQERFNRWFLTGMTVAGVLLGSLFSRK
```

### Bcl-2-like protein 1 from *Homo sapiens* (PDB ID: 1G5J)

FASTA sequence:

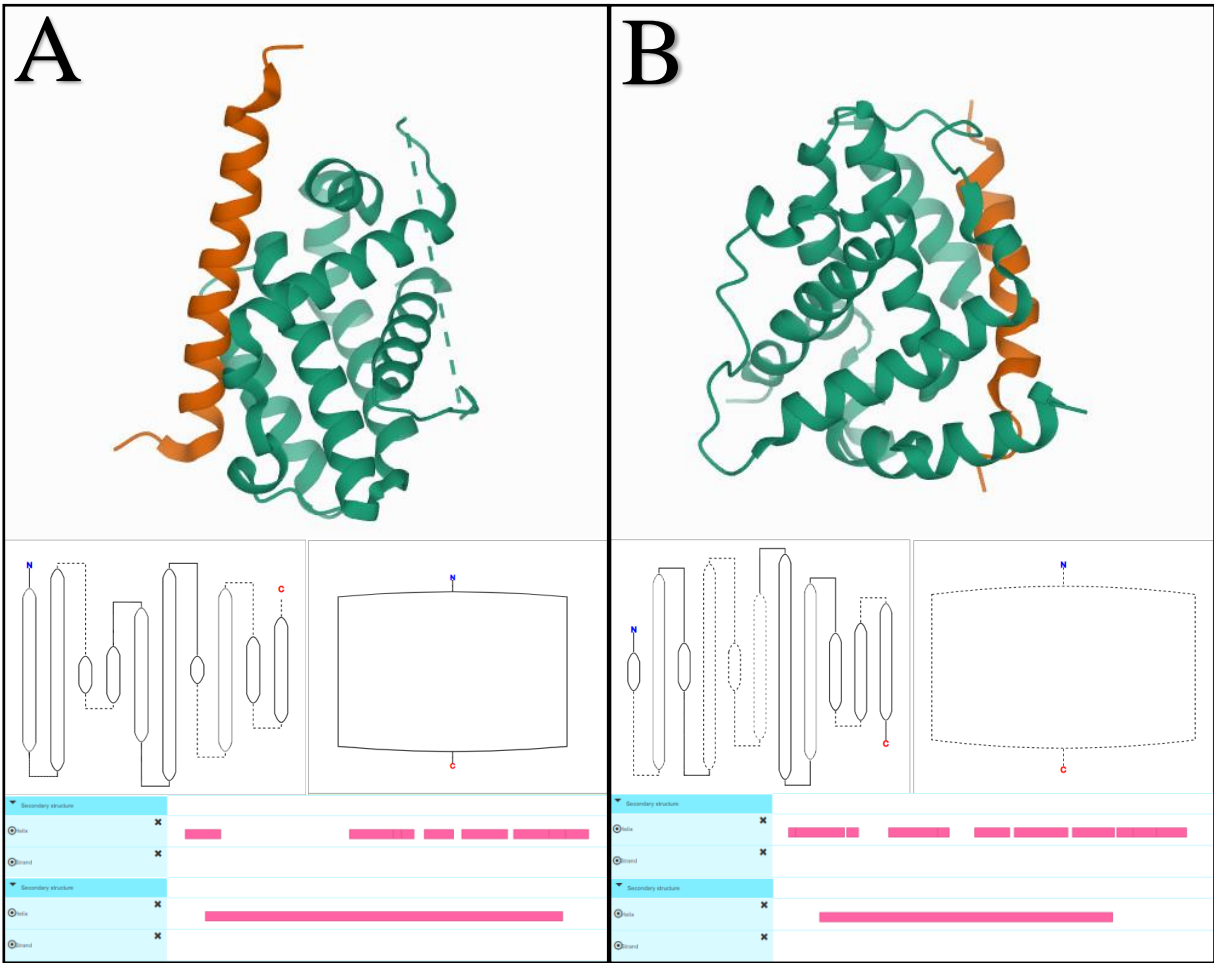
```
MSQSNRELVVDFLSYKLSQKGYSWSQFSDVEENRTEAPEGTESEMETPSAINGNPSWHLADSPAVNGATGHSSSLDAREVIPMAAVKQAL
REAGDEFELRYRRAFSDLTSQLHITPGTAYQSFEQVVNELFRDGVNWGRIVAFSFGGALCVESVDKEMQVLVSRIAAWMATYLNHLEP
WIQENGWDTFVELYGNNAAESRKGQERFNRWFLTGMTVAGVLLGSLFSRK
```

### Functions of those proteins

According to Hardwick and Soane (2013), they are found nearby the junctions of chromosomes 14 ad 18, and they were available for promoting the cell survival to cell proliferation in follicular lymphoma patients' tumour cells. Prevention of cell apoptosis leads to increase at total number of cell. As a result, cancer occurs, and it is assumed that Bcl-2 like proteins prevent cell death and regulate the cell for survival.

# Secondary Structures of template proteins

Proteins from Mus musculus and Homo sapiens are given below, respectively.



**Figure 5:** Protein from Mus musculus, and its primary structure map and distribution in terms of helix and strand for both chain A and chain B of the protein (A), and protein from Homo sapiens (B).

## Prediction of those template proteins with programs

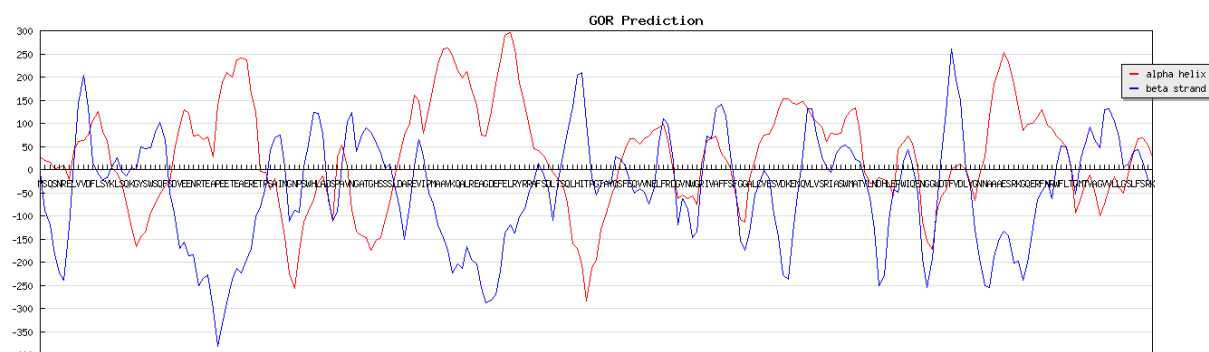
### Chou-Fasman Method



**Figure 6:** Chou-Fasman prediction results of Mus musculus and Homo sapiens, respectively

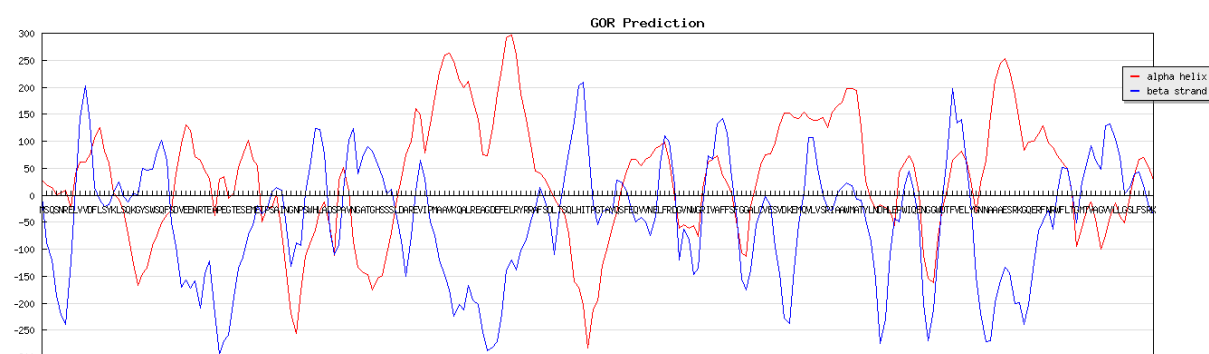
For *Mus musculus*, helices are 67%, strands are 61.4%, and 12% is predicted for turns while for *Homo sapiens*, helices are 64.4%, strands are 61.4%, and 12 percent is for turns.

## GOR Method



**Figure 7:** GOR prediction results of *Mus musculus*

In the results, helix percentage is 45.49, strands percentage is 23.17, turn is 19.74% and coil is 11.58% for *Mus musculus*.



**Figure 8:** GOR prediction results of *Homo sapiens*

In the results, helix percentage is 45.49, strands percentage is 21.88, turn is 21.03%, and coil is 11.58% for *Homo sapiens*.

## Neural Network Method

In the results, for *Mus musculus*, helix percentage is 52.36, strands percentage is 6.44, and coil is 41.20%, while for *Homo sapiens*, percentages for helix, strands, and coil are 49.36, 8.58, and 42.06, respectively.

Alpha helix (Hh) :	122 is 52.36%	Alpha helix (Hh) :	115 is 49.36%
3 <sub>10</sub> helix (Gg) :	0 is 0.00%	3 <sub>10</sub> helix (Gg) :	0 is 0.00%
Pi helix (Ii) :	0 is 0.00%	Pi helix (Ii) :	0 is 0.00%
Beta bridge (Bb) :	0 is 0.00%	Beta bridge (Bb) :	0 is 0.00%
Extended strand (Ee) :	15 is 6.44%	Extended strand (Ee) :	20 is 8.58%
Beta turn (Tt) :	0 is 0.00%	Beta turn (Tt) :	0 is 0.00%
Bend region (Ss) :	0 is 0.00%	Bend region (Ss) :	0 is 0.00%
Random coil (Cc) :	96 is 41.20%	Random coil (Cc) :	98 is 42.06%
Ambiguous states (?) :	0 is 0.00%	Ambiguous states (?) :	0 is 0.00%
Other states :	0 is 0.00%	Other states :	0 is 0.00%

**Figure 9:** Neural Network prediction results of *Mus musculus*, and *Homo sapiens*, respectively

At the end, comparison of methods' results and actual percentages are in the table below.

	<b>Chou-Fasman</b>	<b>GOR</b>	<b>Neural Network</b>	<b>PDBsum</b>
<b>Mus musculus</b>	67% Helices	45.49% Helices	52.36% Helices	73.3% Helices
	12% Turns	19.74% Turns	41.20% Coils	26.7% others
	61.4% Strands	23.17% Strands	6.44% Strands	0.0% Strands
<b>Homo sapiens</b>	64.4% Helices	45.49% Helices	49.36% Helices	69.5% Helices
	12% Turns	21.88% Turns	42.06% Coils	30.5% others
	61.4% Strands	21.03% Strands	8.58% Strands	0.0% Strands

**Figure 10:** Table of results belonging to all prediction methods and PDBsum secondary structure percentages

From the table, it is clear that the most successful method is Neural Network method, which provides the closest percentages with respect to PDBsum results. Taking the sum of differences is the core of the calculation, which is:

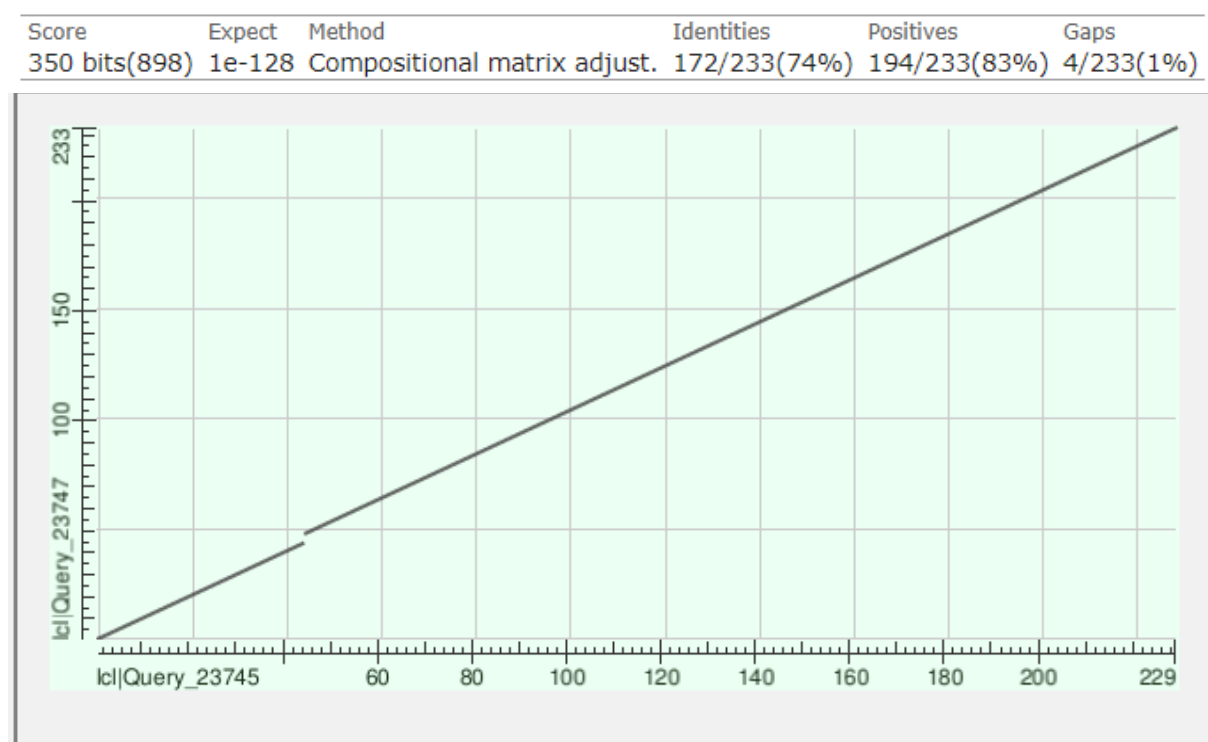
$$((73.3-67)+(26.7-12)+(64.4-0))/3=28.46 \text{ for chou-fasman}$$

$$((73.3-45.49)+(26.7-19.74)+(23.17-0))/3=19.31 \text{ for GOR}$$

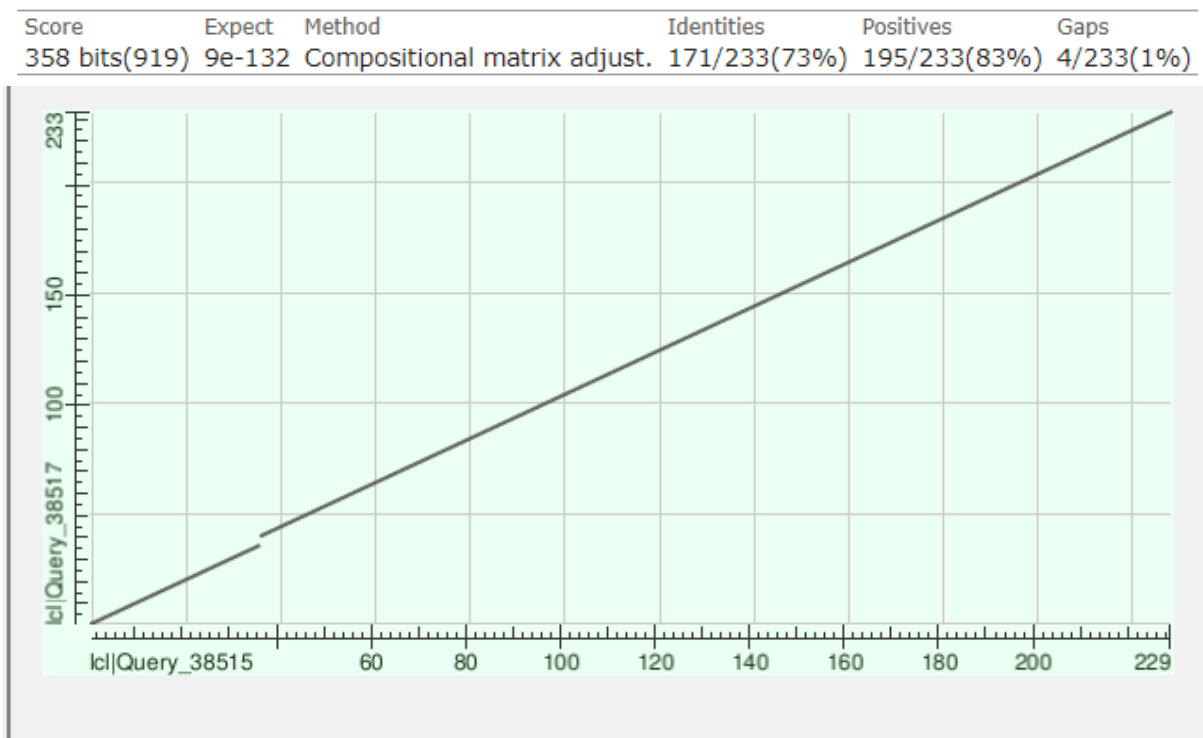
$$((73.3-52.36)+(41.20-26.7)+(6.44-0))/3=13.96 \text{ for Neural Network, which is the least.}$$

## Pairwise alignment of the protein sequence with template sequences

Alignment of Mus musculus and Homo sapiens with the query sequence, which is Gallus gallus, and their scores and dot plots are given below, respectively.



**Figure 11:** Pairwise alignment results and dot plot of Gallus gallus and Mus musculus obtained from NCBI



**Figure 12:** Pairwise alignment results and dot plot of Gallus gallus and Homo sapiens obtained from NCBI

Sequence identities are 73.82% for Mus musculus, and 73.39% for Homo sapiens, which means aligning the exactly matched characters within two sequences. In the case of positive score, which means similarity that is about the aligned residues, 83% for both sequences are obtained. Additionally, their coverages are 100%, and E-value is less than 0.01. With all obtained results in terms of E-value, identity, similarity, and coverage, the matches for both two protein are significant.

## References

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3552500/#:~:text=BCL%2D2%20family%20proteins%20are,as%20the%20intrinsic%20apoptosis%20pathway.>

<https://www.differencebetween.com/difference-between-similarity-and-identity-in-sequence-alignment/>