Project Report

Course: Computational Biology

Slot: F2

Submitted by:

1.HARSHVARDHAN MISHRA   **19BCB0125**

2. SANGHRAJKA LAKSHIT MANISH   **19BCB0031**

3. AYUSH SAHU   **19BCB0122**

Topic: Homology Modelling and Structure Analysis of Amyloid beta protein

Literature survey:

Alzheimer's disease (AD) is a progressive neurodegenerative disease that impairs memory and cognitive judgment. It is the leading cause of dementia in late adult life and is associated with a significant social burden and increased morbidity and mortality in the elderly. Because of mixed effectiveness of medications, exercise has been considered as a treatment for pre-clinical AD, late-stage AD, and as a prevention strategy. Exercise appears to improve brain blood flow, increase hippocampal volume, and improve neurogenesis. Prospective studies indicate that physical inactivity is one of the most common preventable risk factors for developing AD and that higher physical activity levels are associated with a reduced risk of development of disease. Exercise as a treatment for AD shows improvement in cognitive function, decreased neuropsychiatric symptoms, and a slower decline in activities of daily living (ADL). Exercise has been shown to have fewer side effects and better adherence compared to medications.

Protein Information:

UniProtKB Accession Number: P05067

Name: Amyloid beta protein

Gene: APP

- Organism: Homo Sapiens (Humans)

- Sequence Length: 770

- Mass:86,943 Da

Amyloid beta peptide (Aβ) is produced through the proteolytic processing of a transmembrane protein, amyloid precursor protein (APP), by β- and γ-secretases.

Aβ accumulation in the brain is proposed to be an early toxic event in the pathogenesis of Alzheimer's disease, which is the most common form of dementia associated with plaques and tangles in the brain.

Currently, it is unclear what the physiological and pathological forms of Aβ are and by what mechanism Aβ causes dementia.

FASTA SEQUENCE:

>sp|P05067|A4_HUMAN Amyloid-beta precursor protein OS=Homo sapiens OX=9606 GN=APP PE=1 SV=3

MLPGLALLLLAAWTARALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWDSDPSG
TK TCIDTKEGILQYCQEVYPELQITNVVEANQPVTIQNWCKRGRKQCKTHPHFVIPYRCLVG
EFVSDALLVPDKCKFLHQERMDVCETHLHWHTVAKETCSEKSTNLHDYGMLLPCGIDKFR
GVEFVCCPLAEESDNVDSADAEEDDSDVWWGGADTDYADGSEDKVVEVAEEEEVAEVEEE
EADDDEDDEDGDEVEEEAEEPYEEATERTTSIATTTTTTTESVEEVVREVCSEQAETGPC
RAMISRWYFDVTEGKCAPFFYGGCGGNRNNFDTEEYCMAVCGSAMSQSLLKTTQEPLARD
PVKLPTTAASTPDAVDKYLETPGDENEHAHFQKAKERLEAKHRERMSQVMREWEEAERQA
KNLPKADKKAVIQHFQEKVESLEQEAANERQQLVETHMARVEAMLNDRRRLALENYITAL
QAVPPRPRHVFNMLKKYVRAEQKDRQHTLKHFEHVRMVDPKKAAQIRSQVMTHLRVIYER
MNQSLSLLYNVPAVAEEIQDEVDELLQKEQNYSDDVLANMISEPRISYGNDALMPSLTET
KTTVELLPVNGEFSLDDLQPWHSFGADSVPANTENEVEPVDARPAADRGLTTRPGSGLTN
IKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITL
VMLKKKQYTSIHHGVVEVDAAVTPEERHLSKMQQNGYENPTYKFFEQMQN

RESULTS OF MULTIPLE SEQUENCE ALIGNMENT USING CLUSTAL OMEGA:

RESULTS OF MULTIPLE SEQUENCE ALIGNMENT USING CLUSTAL OMEGA (CONT...)

```
XP_017397391.1    QAVPPRPRHVFNMLKKYVRAEQKDRQHTLKHFEHVRMVDPKKAAQIRSQVMTHLRVIYER    540
XP_032132996.1    QAVPPRPRHVFNMLKKYVRAEQKDRQHTLKHFEHVRMVDPKKAAQIRSQVMTHLRVIYER    540
XP_011781877.1    QAVPPRPRHVFNMLKKYVRAEQKDRQHTLKHFEHVRMVDPKKAAQIRSQVMTHLRVIYER    540
XP_005548940.1    QAVPPRPRHVFNMLKKYVRAEQKDRQHTLKHFEHVRMVDPKKAAQIRSQVMTHLRVIYER    540
XP_010381738.1    QAVPPRPRHVFNMLKKYVRAEQKDRQHTLKHFEHVRMVDPKKAAQIRSQVMTHLRVIYER    540
XP_030662027.1    QAVPPRPRHVFNMLKKYVRAEQKDRQHTLKHFEHVRMVDPKKAAQIRSQVMTHLRVIYER    539
XP_004062681.1    QAVPPRPRHVFNMLKKYVRAEQKDRQHTLKHFEHVRMVDPKKAAQIRSQVMTHLRVIYER    540
BAG10647.1        QAVPPRPRHVFNMLKKYVRAEQKDRQHTLKHFEHVRMVDPKKAAQIRSQVMTHLRVIYER    540
NP_001013036.1    QAVPPRPRHVFNMLKKYVRAEQKDRQHTLKHFEHVRMVDPKKAAQIRSQVMTHLRVIYER    540
                  ************************************************************

XP_017397391.1    MNQSLSLLYNVPAVAEEIQDEVDELLQKEQNYSDDVLANMISEPRISYGNDALMPSLTET    600
XP_032132996.1    MNQSLSLLYNVPAVAEEIQDEVDELLQKEQNYSDDVLANMISEPRISYGNDALMPSLTET    600
XP_011781877.1    MNQSLSLLYNVPAVAEEIQDEVDELLQKEQNYSDDVLANMISEPRISYGNDALMPSLTET    600
XP_005548940.1    MNQSLSLLYNVPAVAEEIQDEVDELLQKEQNYSDDVLANMISEPRISYGNDALMPSLTET    600
XP_010381738.1    MNQSLSLLYNVPAVAEEIQDEVDELLQKEQNYSDDVLANMISEPRISYGNDALMPSLTET    600
XP_030662027.1    MNQSLSLLYNVPAVAEEIQDEVDELLQKEQNYSDDVLANMISEPRISYGNDALMPSLTET    599
XP_004062681.1    MNQSLSLLYNVPAVAEEIQDEVDELLQKEQNYSDDVLANMISEPRISYGNDALMPSLTET    600
BAG10647.1        MNQSLSLLYNVPAVAEEIQDEVDELLQKEQNYSDDVLANMISEPRISYGNDALMPSLTET    600
NP_001013036.1    MNQSLSLLYNVPAVAEEIQDEVDELLQKEQNYSDDVLANMISEPRISYGNDALMPSLTET    600
                  ************************************************************

XP_017397391.1    KTTVELLPVNGEFSLDDLQPWHSFGADSVPANTENEVEPVDARPAADRGLTTRPGSGLTN    660
XP_032132996.1    KTTVELLPVNGEFSLDDLQPWHSFGADSVPANTENEVEPVDARPAADRGLTTRPGSGLTN    660
XP_011781877.1    KTTVELLPVNGEFSLDDLQPWHSFGADSVPANTENEVEPVDARPAADRGLTTRPGSGLTN    660
XP_005548940.1    KTTVELLPVNGEFSLDDLQPWHSFGADSVPANTENEVEPVDARPAADRGLTTRPGSGLTN    660
XP_010381738.1    KTTVELLPVNGEFSLDDLQPWHSFGADSVPANTENEVEPVDARPAADRGLTTRPGSGLTN    660
XP_030662027.1    KTTVELLPVNGEFSLDDLQPWHSFGADSVPANTENEVEPVDARPAADRGLTTRPGSGLTN    659
XP_004062681.1    KTTVELLPVNGEFSLDDLQPWHSFGADSVPANTENEVEPVDARPAADRGLTTRPGSGLTN    660
BAG10647.1        KTTVELLPVNGEFSLDDLQPWHSFGADSVPANTENEVEPVDARPAADRGLTTRPGSGLTN    660
NP_001013036.1    KTTVELLPVNGEFSLDDLQPWHSFGADSVPANTENEVEPVDARPAADRGLTTRPGSGLTN    660
                  ************************************************************

XP_017397391.1    IKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITL    720
XP_032132996.1    IKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITL    720
XP_011781877.1    IKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITL    720
XP_005548940.1    IKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITL    720
XP_010381738.1    IKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITL    720
XP_030662027.1    IKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITL    719
XP_004062681.1    IKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITL    720
BAG10647.1        IKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITL    720
NP_001013036.1    IKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITL    720
                  ************************************************************

XP_017397391.1    VMLKKKQYTSIHHGVVEVDAAVTPEERHLSKMQQNGYENPTYKFFEQMQN    770
XP_032132996.1    VMLKKKQYTSIHHGVVEVDAAVTPEERHLSKMQQNGYENPTYKFFEQMQN    770
XP_011781877.1    VMLKKKQYTSIHHGVVEVDAAVTPEERHLSKMQQNGYENPTYKFFEQMQN    770
XP_005548940.1    VMLKKKQYTSIHHGVVEVDAAVTPEERHLSKMQQNGYENPTYKFFEQMQN    770
XP_010381738.1    VMLKKKQYTSIHHGVVEVDAAVTPEERHLSKMQQNGYENPTYKFFEQMQN    770
XP_030662027.1    VMLKKKQYTSIHHGVVEVDAAVTPEERHLSKMQQNGYENPTYKFFEQMQN    769
XP_004062681.1    VMLKKKQYTSIHHGVVEVDAAVTPEERHLSKMQQNGYENPTYKFFEQMQN    770
BAG10647.1        VMLKKKQYTSIHHGVVEVDAAVTPEERHLSKMQQNGYENPTYKFFEQMQN    770
NP_001013036.1    VMLKKKQYTSIHHGVVEVDAAVTPEERHLSKMQQNGYENPTYKFFEQMQN    770
                  *************************************************
```

```
XP_017397391.1    QAVPPRPRHVFNMLKKYVRAEQKDRQHTLKHFEHVRMVDPKKAAQIRSQVMTHLRVIYER    540
XP_032132996.1    QAVPPRPRHVFNMLKKYVRAEQKDRQHTLKHFEHVRMVDPKKAAQIRSQVMTHLRVIYER    540
XP_011781877.1    QAVPPRPRHVFNMLKKYVRAEQKDRQHTLKHFEHVRMVDPKKAAQIRSQVMTHLRVIYER    540
XP_005548940.1    QAVPPRPRHVFNMLKKYVRAEQKDRQHTLKHFEHVRMVDPKKAAQIRSQVMTHLRVIYER    540
XP_010381738.1    QAVPPRPRHVFNMLKKYVRAEQKDRQHTLKHFEHVRMVDPKKAAQIRSQVMTHLRVIYER    540
XP_030662027.1    QAVPPRPRHVFNMLKKYVRAEQKDRQHTLKHFEHVRMVDPKKAAQIRSQVMTHLRVIYER    539
XP_004062681.1    QAVPPRPRHVFNMLKKYVRAEQKDRQHTLKHFEHVRMVDPKKAAQIRSQVMTHLRVIYER    540
BAG10647.1        QAVPPRPRHVFNMLKKYVRAEQKDRQHTLKHFEHVRMVDPKKAAQIRSQVMTHLRVIYER    540
NP_001013036.1    QAVPPRPRHVFNMLKKYVRAEQKDRQHTLKHFEHVRMVDPKKAAQIRSQVMTHLRVIYER    540
                  ************************************************************

XP_017397391.1    MNQSLSLLYNVPAVAEEIQDEVDELLQKEQNYSDDVLANMISEPRISYGNDALMPSLTET    600
XP_032132996.1    MNQSLSLLYNVPAVAEEIQDEVDELLQKEQNYSDDVLANMISEPRISYGNDALMPSLTET    600
XP_011781877.1    MNQSLSLLYNVPAVAEEIQDEVDELLQKEQNYSDDVLANMISEPRISYGNDALMPSLTET    600
XP_005548940.1    MNQSLSLLYNVPAVAEEIQDEVDELLQKEQNYSDDVLANMISEPRISYGNDALMPSLTET    600
XP_010381738.1    MNQSLSLLYNVPAVAEEIQDEVDELLQKEQNYSDDVLANMISEPRISYGNDALMPSLTET    600
XP_030662027.1    MNQSLSLLYNVPAVAEEIQDEVDELLQKEQNYSDDVLANMISEPRISYGNDALMPSLTET    599
XP_004062681.1    MNQSLSLLYNVPAVAEEIQDEVDELLQKEQNYSDDVLANMISEPRISYGNDALMPSLTET    600
BAG10647.1        MNQSLSLLYNVPAVAEEIQDEVDELLQKEQNYSDDVLANMISEPRISYGNDALMPSLTET    600
NP_001013036.1    MNQSLSLLYNVPAVAEEIQDEVDELLQKEQNYSDDVLANMISEPRISYGNDALMPSLTET    600
                  ************************************************************

XP_017397391.1    KTTVELLPVNGEFSLDDLQPWHSFGADSVPANTENEVEPVDARPAADRGLTTRPGSGLTN    660
XP_032132996.1    KTTVELLPVNGEFSLDDLQPWHSFGADSVPANTENEVEPVDARPAADRGLTTRPGSGLTN    660
XP_011781877.1    KTTVELLPVNGEFSLDDLQPWHSFGADSVPANTENEVEPVDARPAADRGLTTRPGSGLTN    660
XP_005548940.1    KTTVELLPVNGEFSLDDLQPWHSFGADSVPANTENEVEPVDARPAADRGLTTRPGSGLTN    660
XP_010381738.1    KTTVELLPVNGEFSLDDLQPWHSFGADSVPANTENEVEPVDARPAADRGLTTRPGSGLTN    660
XP_030662027.1    KTTVELLPVNGEFSLDDLQPWHSFGADSVPANTENEVEPVDARPAADRGLTTRPGSGLTN    659
XP_004062681.1    KTTVELLPVNGEFSLDDLQPWHSFGADSVPANTENEVEPVDARPAADRGLTTRPGSGLTN    660
BAG10647.1        KTTVELLPVNGEFSLDDLQPWHSFGADSVPANTENEVEPVDARPAADRGLTTRPGSGLTN    660
NP_001013036.1    KTTVELLPVNGEFSLDDLQPWHSFGADSVPANTENEVEPVDARPAADRGLTTRPGSGLTN    660
                  ************************************************************

XP_017397391.1    IKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITL    720
XP_032132996.1    IKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITL    720
XP_011781877.1    IKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITL    720
XP_005548940.1    IKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITL    720
XP_010381738.1    IKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITL    720
XP_030662027.1    IKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITL    719
XP_004062681.1    IKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITL    720
BAG10647.1        IKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITL    720
NP_001013036.1    IKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITL    720
                  ************************************************************

XP_017397391.1    VMLKKKQYTSIHHGVVEVDAAVTPEERHLSKMQQNGYENPTYKFFEQMQN    770
XP_032132996.1    VMLKKKQYTSIHHGVVEVDAAVTPEERHLSKMQQNGYENPTYKFFEQMQN    770
XP_011781877.1    VMLKKKQYTSIHHGVVEVDAAVTPEERHLSKMQQNGYENPTYKFFEQMQN    770
XP_005548940.1    VMLKKKQYTSIHHGVVEVDAAVTPEERHLSKMQQNGYENPTYKFFEQMQN    770
XP_010381738.1    VMLKKKQYTSIHHGVVEVDAAVTPEERHLSKMQQNGYENPTYKFFEQMQN    770
XP_030662027.1    VMLKKKQYTSIHHGVVEVDAAVTPEERHLSKMQQNGYENPTYKFFEQMQN    769
XP_004062681.1    VMLKKKQYTSIHHGVVEVDAAVTPEERHLSKMQQNGYENPTYKFFEQMQN    770
BAG10647.1        VMLKKKQYTSIHHGVVEVDAAVTPEERHLSKMQQNGYENPTYKFFEQMQN    770
NP_001013036.1    VMLKKKQYTSIHHGVVEVDAAVTPEERHLSKMQQNGYENPTYKFFEQMQN    770
                  *************************************************
```

Phylogram:

Phylogram is a phylogenetic tree whose branch lengths are proportional to how many character changes have been inferred along the branches

Phylogram

Branch length: ● Cladogram ○ Real

XP_017397391.1 0.000649351
XP_032132996.1 0.000649351
XP_011781877.1 0.000974026
XP_005548940.1 0.000649351
XP_010381738.1 0.000649351
XP_030662027.1 0.00243823
XP_004062681.1 0.000974026
BAG10647.1 0.000649351
NP_001013036.1 0.000649351
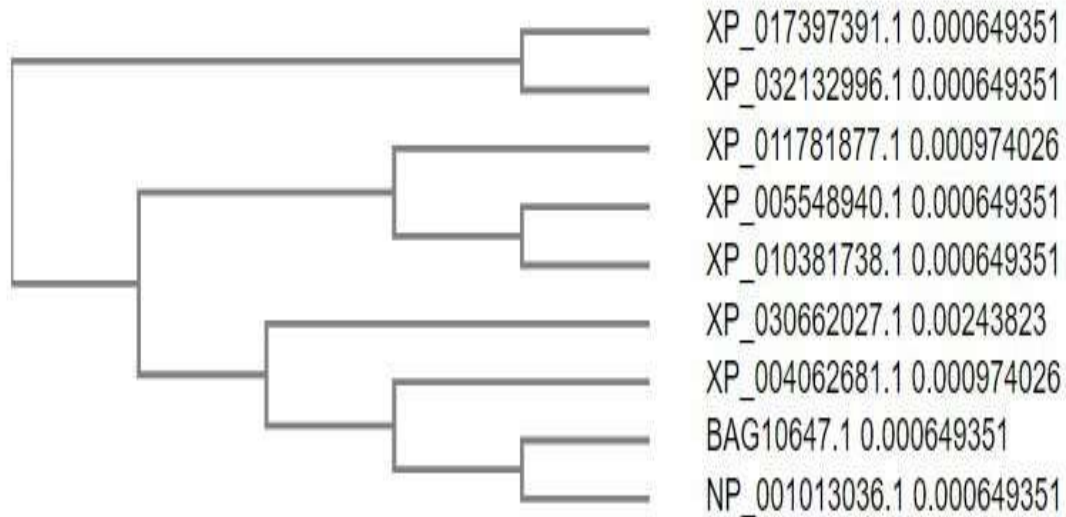
Phylogenetic Tree:

It is a diagram that represents evolutionary relationships among organisms. The pattern of branching in a phylogenetic tree reflects how species or other groups evolved from a series of common ancestors.

Phylogenetic Tree

This is a Neighbour-joining tree without distance corrections.

Branch length: ● Cladogram ○ Real

XP_017397391.1 0.00037
XP_032132996.1 0.00093
XP_011781877.1 0.0013
XP_010381738.1 0
XP_005548940.1 0.00013
XP_030662027.1 0
NP_001013036.1 0
XP_004062681.1 0.0013
BAG10647.1 0.0013
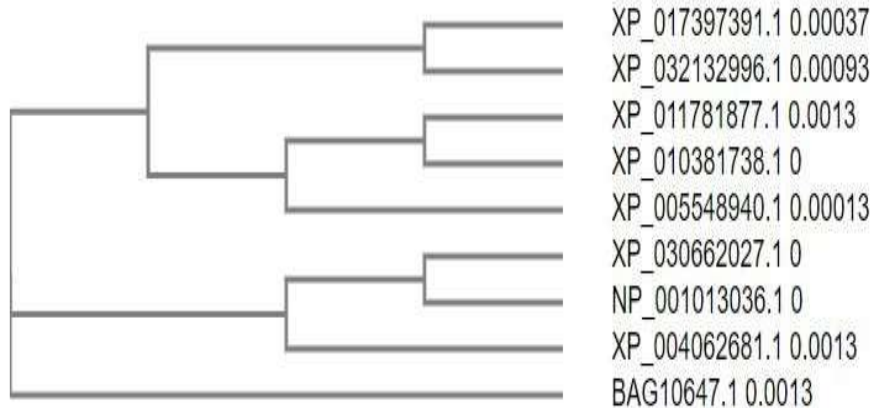
Implementation:

Structure Prediction: (Model 1):
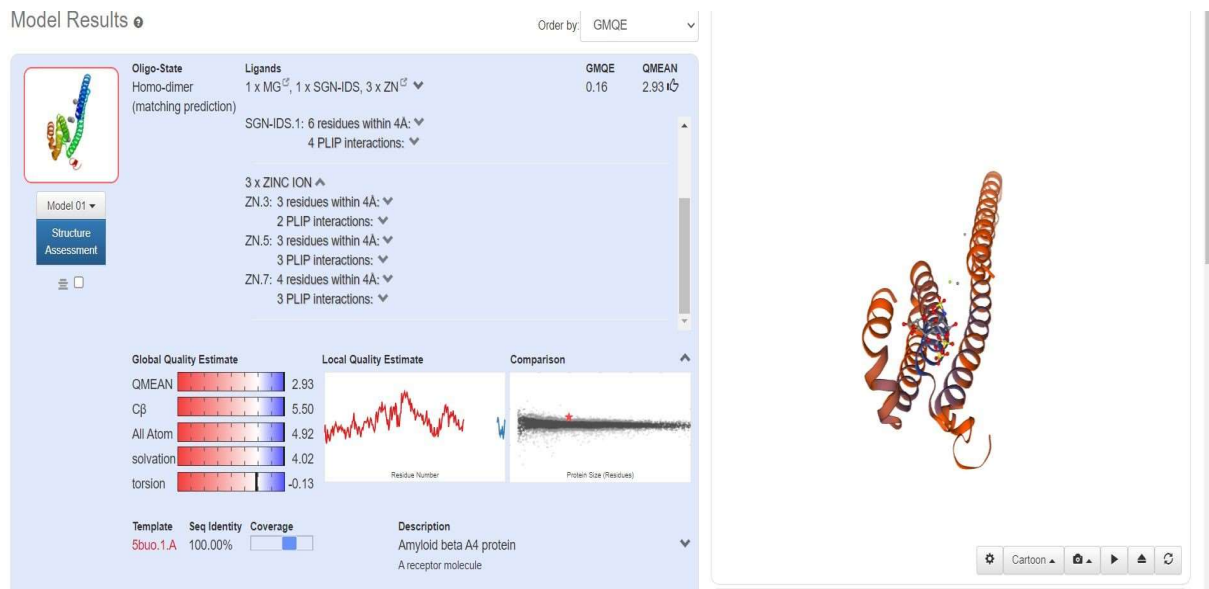
Ramachandran Plots: Here Ramachandran favored are 96.93%
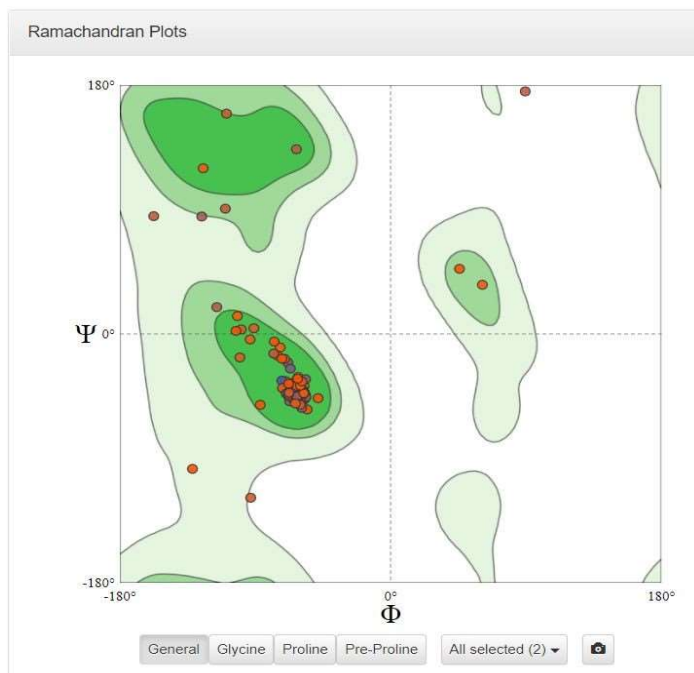
General | Glycine | Proline | Pre-Proline | All selected (2) ▾ | 📷

MolProbity Results ︿

MolProbity Score	0.85	
☐ Clash Score	1.26	(B467 ASP-B470 ARG)
Ramachandran Favoured	98.93%	
Ramachandran Outliers	0.00%	
☐ Rotamer Outliers	0.59%	A557 GLU, A376 ASP
☐ C-Beta Deviations	1	A570 GLN
☐ Bad Bonds	9 / 3216	A402 HIS, B388 HIS, A511 HIS, A388 HIS, B514 HIS, B511 HIS, A457 HIS, B507 HIS, B457 HIS
☐ Bad Angles	40 / 4317	A448 ASN, B448 ASN, (A569 GLU-A570 GLN), B434 HIS, B467 ASP, A467 ASP, A376 ASP, A533 HIS, B457 HIS, B399 GLU, B390 HIS, B533 HIS, B427 ASP, A434 HIS, B504 ASP, B388 HIS, A501 GLU, (B486 ARG-B487 PRO), B402 HIS, (A486 ARG-A487 PRO), (A381 THR-A382 PRO), A452 GLN, A507 HIS, A402 HIS, B489 HIS, A511 HIS, (A375 VAL -A376
☐ Cis Prolines	2 / 13	(A483 VAL-A484 PRO), (B483 VAL-B484 PRO)

Results obtained using MolProbity version 4.4

Structure Prediction: (Model 2):

Ramachandran Plots: Ramachandran Plots: Here Ramachandran favored are 98.93%

Ramachandran Plots ∧

General | Glycine | Proline | Pre-Proline | All selected (2) ▾ | 📷

MolProbity Results ∧

MolProbity Score	0.85	
☐ Clash Score	1.26	(B467 ASP-B470 ARG)
Ramachandran Favoured	98.93%	
Ramachandran Outliers	0.00%	
☐ Rotamer Outliers	0.59%	A557 GLU, A376 ASP
☐ C-Beta Deviations	1	A570 GLN
☐ Bad Bonds	9 / 3216	A402 HIS, B388 HIS, A511 HIS, A388 HIS, B514 HIS, B511 HIS, A457 HIS, B507 HIS, B457 HIS
☐ Bad Angles	40 / 4317	B467 ASP, A467 ASP, A376 ASP, A533 HIS, B457 HIS, B399 GLU, B390 HIS, B533 HIS, B427 ASP, A434 HIS, B504 ASP, B388 HIS, A501 GLU, (B486 ARG-B487 PRO), B402 HIS, (A486 ARG-A487 PRO), (A381 THR-A382 PRO), A452 GLN, A507 HIS, A402 HIS, B489 HIS, A511 HIS, (A375 VAL-A376 ASP), A514 HIS, A388 HIS, B511 HIS, A457 HIS, B452 GLN
☐ Cis Prolines	2 / 13	(A483 VAL-A484 PRO), (B483 VAL-B484 PRO)

Results obtained using MolProbity version 4.4

Code snippets:

```
!pip install Biopython


# Global sequence alignment with ClustalW

# import the commandline software

from Bio.Align.Applications import ClustalwCommandline

in_file = "unaligned.fasta"

# parsing DNA file (Unaligned)

clustalw_cline = ClustalwCommandline("clustalw2", infile= in_file)

#performing the alignment

# printing the output

print(clustalw_cline)


#import clustalw from Biopython

#step1)

from Bio.Align.Applications import ClustalwCommandline

cline = ClustalwCommandline("clustalw2", infile = "unaligned.fasta")

print(cline)
```

```python
# path
# C:\Program Files (x86)\ClustalW2

# A) assertion
#B) Run Clustal W fasta
import os
clustalw_exe = r"C:\Program Files (x86)\ClustalW2\clustalw2.exe"
clustalw_cline = ClustalwCommandline(clustalw_exe, infile = "unaligned.fasta")
assert os.path.isfile(clustalw_exe), "Clustal_W executable is missing or not found"
stdout, stderr = clustalw_cline()
print(clustalw_cline)

#Generate Alignment output
from Bio import AlignIO
ClustalAlign = AlignIO.read("unaligned.aln", "clustal")
print(ClustalAlign)
```

```python
# Generate Dendogram (Tree) output

from Bio import Phylo

tree = Phylo.read("unaligned.dnd", "newick")

Phylo.draw_ascii(tree)
```

```
In [6]: !pip install Biopython

        Requirement already satisfied: Biopython in c:\users\hii\anaconda3\lib\site-packages (1.79)
        Requirement already satisfied: numpy in c:\users\hii\anaconda3\lib\site-packages (from Biopython) (1.18.5)

In [7]: # Global sequence alignment with ClustalW
        # import the commandline software
        from Bio.Align.Applications import ClustalwCommandline
        in_file = "unaligned.fasta"
        # parsing DNA file (Unaligned)
        clustalw_cline = ClustalwCommandline("clustalw2", infile= in_file)
        #performing the alignment
        # printing the output
        print(clustalw_cline)

        clustalw2 -infile=unaligned.fasta

In [8]: #import clustalw from Biopython
        #step1)
        from Bio.Align.Applications import ClustalwCommandline
        cline = ClustalwCommandline("clustalw2", infile = "unaligned.fasta")
        print(cline)
        # path
        # C:\Program Files (x86)\ClustalW2

        clustalw2 -infile=unaligned.fasta
```

```
In [9]:  # A) assertion
         #B) Run Clustal W fasta
         import os
         clustalw_exe = r"C:\Program Files (x86)\ClustalW2\clustalw2.exe"
         clustalw_cline = ClustalwCommandline(clustalw_exe, infile = "unaligned.fasta")
         assert os.path.isfile(clustalw_exe), "Clustal_W executable is missing or not found"
         stdout, stderr = clustalw_cline()
         print(clustalw_cline)

         "C:\Program Files (x86)\ClustalW2\clustalw2.exe" -infile=unaligned.fasta


In [10]: #Generate Alignment output
         from Bio import AlignIO
         ClustalAlign = AlignIO.read("unaligned.aln", "clustal")
         print(ClustalAlign)

         Alignment with 100 rows and 938 columns
         --------------------------------------------...MQN NP_000475.1
         --------------------------------------------...MQN ANN47479.1
         --------------------------------------------...MQN NP_001191230.1
         --------------------------------------------...MQN XP_009449632.1
         --------------------------------------------...MQN XP_004062684.1
         --------------------------------------------...MQN XP_004062681.1
         --------------------------------------------...MQN NP_001013036.1
         --------------------------------------------...MQN XP_030662027.1
         --------------------------------------------...MQN XP_032030716.1
         --------------------------------------------...MQN XP_005548940.1
         --------------------------------------------...MQN XP_011853699.1
         --------------------------------------------...MQN XP_023045769.1
         --------------------------------------------...MQN XP_010381738.1
         --------------------------------------------...MQN XP_011781877.1
         --------------------------------------------...MQN XP_011781879.1
         --------------------------------------------...MQN XP_004062682.1
         --------------------------------------------...MQN NP_958816.1
         --------------------------------------------...MQN XP_005548942.1
```

Results:

```
In [11]: # Generate Dendogram (Tree) output
         from Bio import Phylo
         tree = Phylo.read("unaligned.dnd", "newick")
         Phylo.draw_ascii(tree)
```

```
                              , NP_000475.1
                              |
                              |_ ANN47479.1
                            , |
                            |, XP_004062681.1
                            |||
                            ||| _ NP_001191230.1
                          , | ||
                          || ||| XP_009449632.1
                          || |
                          || |_ XP_004062684.1
                        _ ||
                        | || NP_001013036.1
                        | |
                        | |        , XP_030662027.1
                        | |_____|
                        |          |____ XP_032030716.1
                        |
                        |  , XP_005548940.1
                        |  |
```

Jupyter unaligned.aln✔ 17 hours ago

Logout

File   Edit   View   Language

Plain Text

```
310  NP_000475.1          RALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWDSDPSGTKTCIDTKEGILQYCQE
311  ANN47479.1           RALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWDSDPSGTKTCIDTKEGILQYCQE
312  NP_001191230.1       RALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWDSDPSGTKTCIDTKEGILQYCQE
313  XP_009449632.1       RALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWDSDPSGTKTCIDTKEGILQYCQE
314  XP_004062684.1       RALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWDSDPSGTKTCIDTKEGILQYCQE
315  XP_004062681.1       RALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWDSDPSGTKTCIDTKEGILQYCQE
316  NP_001013036.1       RALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWDSDPSGTKTCIDTKEGILQYCQE
317  XP_030662027.1       RALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWDSDPSGTKTCIDTKEGILQYCQE
318  XP_032030716.1       RALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWDSDPSGTKTCIDTKEGILQYCQE
319  XP_005548940.1       RALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWDSDPSGTKTCIDTKEGILQYCQE
320  XP_011853699.1       ---HVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWDSDPSGTKTCIDTKEGILQYCQE
321  XP_023045769.1       RALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWDSDPSGTKTCIDTKEGILQYCQE
322  XP_010381738.1       RALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWDSDPSGTKTCIDTKEGILQYCQE
323  XP_011781877.1       RALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWDSDPSGTKTCIDTKEGILQYCQE
324  XP_011781879.1       RALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWDSDPSGTKTCIDTKEGILQYCQE
325  XP_004062682.1       RALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWDSDPSGTKTCIDTKEGILQYCQE
326  NP_958816.1          RALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWDSDPSGTKTCIDTKEGILQYCQE
327  XP_005548942.1       RALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWDSDPSGTKTCIDTKEGILQYCQE
328  BAE01907.1           RALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWDSDPSGTKTCIDTKEGILQYCQE
329  XP_017397391.1       RALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWESDPSGTKTCIDTKEGILQYCQE
330  XP_002761374.2       RALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWESDPSGTKTCIDTKEGILQYCQE
331  XP_032132996.1       RALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWESDPSGTKTCIDTKEGILQYCQE
332  XP_039336411.1       RALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWESDPSGTKTCIDTKEGILQYCQE
333  XP_041590715.1       RALEVPTDGNAGLLAEPQVAMLCGKLRMHMNVQNGKWESDPLGTKTCIGSKEDILQYCQE
334  CAD7693518.1         RALEVPTDGNAGLLAEPQVAMLCGKLRMHMNVQNGKWESDPLGTKTCIGSKEDILQYCQE
335  AAX81908.1           RALEVPTDGNAGLLAEPQVAMLCGKLRMHMNVQNGKWESDPLGTKTCIGSKEDILQYCQE
336  AAR97726.1           RALEVPTDGNAGLLAEPQVAMLCGKLRMHMNVQNGKWESDPLGTKTCIGSKEDILQYCQE
337  NP_001006601.2       RALEVPTDGNAGLLAEPQVAMLCGKLRMHMNVQNGKWESDPLGTKTCIGSKEDILQYCQE
338  XP_026373752.1       RALEVPTDGNAGLLAEPQVAMFCGKLNMHMNVQNGKWESDPSGTKTCIGTKEDILQYCQE
339  XP_002920108.1       RALEVPTDGNAGLLAEPQVAMFCGKLNMHMNVQNGKWESDPSGTKTCIGTKEDILQYCQE
340  XP_032721380.1       RALEVPTDGNAGLLAEPQVAMFCGKLNMHMNVQNGKWESDPSGTKTCIGTKEDILQYCQE
341  XP_022366667.1       RALEVPTDGNAGLLAEPQVAMFCGKLNMHMNVQNGKWESDPSGTKTCIGTKEDILQYCQE
342  XP_044941112.1       RALEVPTDGNAGLLAEPQVAVFCGKLNMHMNVQNGKWESDPSGTKTCIGTKEDILQYCQE
343  XP_044107007.1       RALEVPTDGNAGLLAEPQVAVFCGKLNMHMNVQNGKWESDPSGTKTCIGTKEDILQYCQE
344  XP_032249242.1       RALEVPTDGNAGLLAEPQVAMFCGKLNMHMNVQNGKWESDPSGTKTCIGTKEDILQYCQE
345  XP_035928655.1       RALEVPTDGNAGLLAEPQVAMFCGKLNMHMNVQNGKWESDPSGTKTCIGTKEDILQYCQE
346  XP_021533767.1       RALEVPTDGNAGLLAEPQVAMFCGKLNMHMNVQNGKWESDPSGTKTCIGTKEDILQYCQE
347  XP_034873420.1       RALEVPTDGNAGLLAEPQVAMFCGKLNMHMNVQNGKWESDPSGTKTCIGTKEDILQYCQE
348  XP_019323331.1       RALEVPTDGNAGLLAEPQVAMFCGKLNMHMNVQNGKWESDPSGTKTCIGTKEDILQYCQE
349  XP_042812781.1       RALEVPTDGNAGLLAEPQVAMFCGKLNMHMNVQNGKWESDPSGTKTCIGTKEDILQYCQE
350  XP_039093221.1       RALEVPTDGNAGLLAEPQVAMFCGKLNMHMNVQNGKWESDPSGTKTCIGTKEDILQYCQE
```

File   Edit   View   Language   Plain Text

```
48  (
49  XP_005548940.1:0.00028,
50  XP_011853699.1:0.01802)
51  :0.00097,
52  XP_023045769.1:0.00228)
53  :0.00000,
54  (
55  XP_010381738.1:0.00000,
56  XP_011781877.1:0.00126)
57  :0.00031)
58  :0.00156,
59  XP_011781879.1:0.00101)
60  :0.00028,
61  XP_004062682.1:0.00130)
62  :0.00000,
63  (
64  NP_958816.1:0.00133,
65  XP_005548942.1:0.00000)
66  :0.00000)
67  :0.00012,
68  BAE01907.1:0.00122)
69  :0.00165)
70  :0.00128,
71  (
72  (
73  XP_017397391.1:0.00000,
74  XP_002761374.2:0.00250)
75  :0.00118,
76  (
77  XP_032132996.1:0.00000,
78  XP_039336411.1:0.00121)
79  :0.00012)
80  :0.00279)
81  :0.00708,
82  XP_037688902.1:0.01637)
83  :0.00103,
84  (
85  (
86  XP_012619905.1:0.00268,
87  XP_012495594.1:0.00252)
88  :0.00185,
89  XP_023365467.1:0.00661)
```

References:

https://www.uniprot.org/

https://www.ebi.ac.uk/Tools/msa/clustalo/

https://swissmodel.expasy.org/