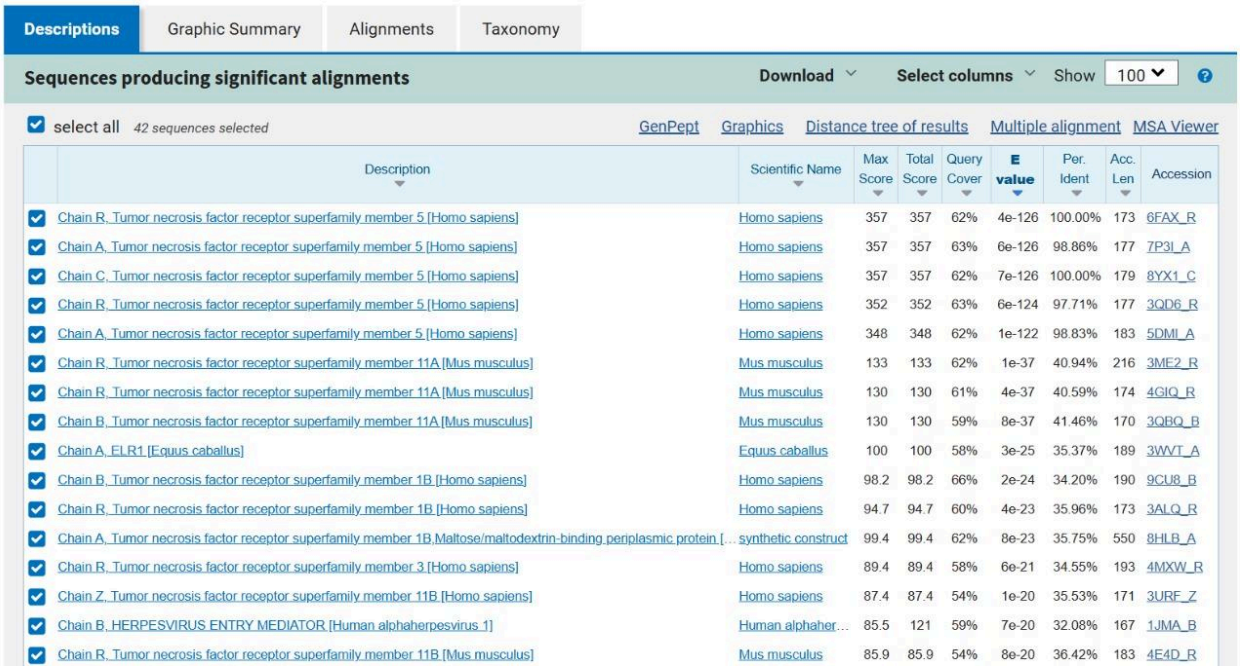
CADD ASSIGNMENT 1 REPORT
HARSH VISHWAKARMA
2022205

Identify the Template for Modeling (1 mark)

- Find a suitable template for modeling the given protein sequence (1 mark)



Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments — Download — Select columns — Show 100

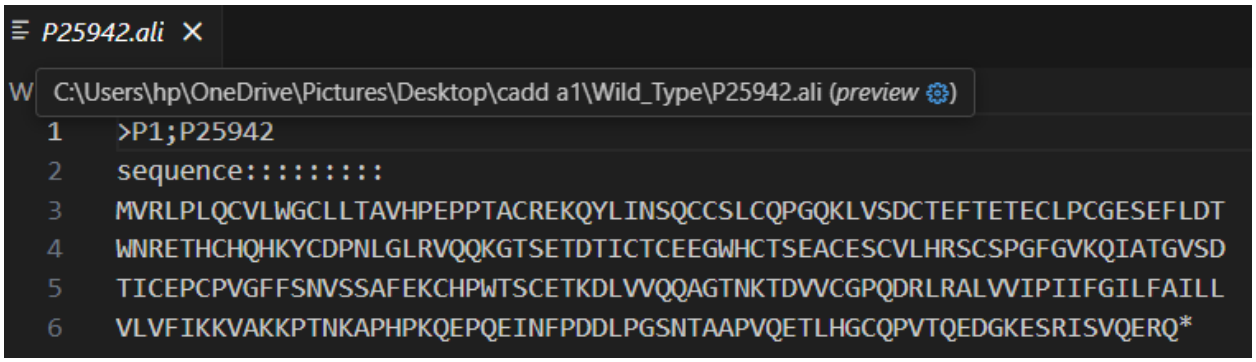
select all 42 sequences selected — GenPept | Graphics | Distance tree of results | Multiple alignment | MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Chain R, Tumor necrosis factor receptor superfamily member 5 [Homo sapiens]	Homo sapiens	357	357	62%	4e-126	100.00%	173	6FAX_R
Chain A, Tumor necrosis factor receptor superfamily member 5 [Homo sapiens]	Homo sapiens	357	357	63%	6e-126	98.86%	177	7P3I_A
Chain C, Tumor necrosis factor receptor superfamily member 5 [Homo sapiens]	Homo sapiens	357	357	62%	7e-126	100.00%	179	8YX1_C
Chain R, Tumor necrosis factor receptor superfamily member 5 [Homo sapiens]	Homo sapiens	352	352	63%	6e-124	97.71%	177	3QD6_R
Chain A, Tumor necrosis factor receptor superfamily member 5 [Homo sapiens]	Homo sapiens	348	348	62%	1e-122	98.83%	183	5DMI_A
Chain R, Tumor necrosis factor receptor superfamily member 11A [Mus musculus]	Mus musculus	133	133	62%	1e-37	40.94%	216	3ME2_R
Chain R, Tumor necrosis factor receptor superfamily member 11A [Mus musculus]	Mus musculus	130	130	61%	4e-37	40.59%	174	4GIQ_R
Chain B, Tumor necrosis factor receptor superfamily member 11A [Mus musculus]	Mus musculus	130	130	59%	8e-37	41.46%	170	3QBQ_B
Chain A, ELR1 [Equus caballus]	Equus caballus	100	100	58%	3e-25	35.37%	189	3WVT_A
Chain B, Tumor necrosis factor receptor superfamily member 1B [Homo sapiens]	Homo sapiens	98.2	98.2	66%	2e-24	34.20%	190	9CU8_B
Chain R, Tumor necrosis factor receptor superfamily member 1B [Homo sapiens]	Homo sapiens	94.7	94.7	60%	4e-23	35.96%	173	3ALQ_R
Chain A, Tumor necrosis factor receptor superfamily member 1B,Maltose/maltodextrin-binding periplasmic protein [...	synthetic construct	99.4	99.4	62%	8e-23	35.75%	550	8HLB_A
Chain R, Tumor necrosis factor receptor superfamily member 3 [Homo sapiens]	Homo sapiens	89.4	89.4	58%	6e-21	34.55%	193	4MXW_R
Chain Z, Tumor necrosis factor receptor superfamily member 11B [Homo sapiens]	Homo sapiens	87.4	87.4	54%	1e-20	35.53%	171	3URF_Z
Chain B, HERPESVIRUS ENTRY MEDIATOR [Human alphaherpesvirus 1]	Human alphaher...	85.5	121	59%	7e-20	32.08%	167	1JMA_B
Chain R, Tumor necrosis factor receptor superfamily member 11B [Mus musculus]	Mus musculus	85.9	85.9	54%	8e-20	36.42%	183	4E4D_R

After performing BLAST on the given uniprot, This was the result. The top 3 sequences were chosen because of their overall High per. Identity, High query cover, Low E-value.

Modeling the Mutated Proteins (9.5 marks)

- Introduce the specified point mutations in the given wild type sequence. (1.5 marks: 0.5 marks per correct mutation)



P25942.ali

C:\Users\hp\OneDrive\Pictures\Desktop\cadd a1\Wild_Type\P25942.ali (preview)

```
1  >P1;P25942
2  sequence::::::::
3  MVRLPLQCVLWGCLLTAVHPEPPTACREKQYLINSQCCSLCQPGQKLVSDCTEFTETECLPCGESEFLDT
4  WNRETHCHQHKYCDPNLGLRVQQKGTSETDTICTCEEGWHCTSEACESCVLHRSCSPGFGVKQIATGVSD
5  TICEPCPVGFFSNVSSAFEKCHPWTSCETKDLVVQQAGTNKTDVVCGPQDRLRALVVIPIIFGILFAILL
6  VLVFIKKVAKKPTNKAPHPKQEPQEINFPDDLPGSNTAAPVQETLHGCQPVTQEDGKESRISVQERQ*
```

The above sequence is the WILD TYPE TEMPLATE. BELOW are the mutated versions highlighted by Blue Dots at the bottom of the position.

P227A.ali

Wild_Type > P227A.ali

```
1  >P1;P25942
2  sequence:::::::::
3  MVRLPLQCVLWGCLLTAVHPEPPTACREKQYLINSQCCSLCQPGQKLVSDCTEFTETECLPCGESEFLDT
4  WNRETHCHQHKYCDPNLGLRVQQKGTSETDTICTCEEGWHCTSEACESCVLHRSCSPGFGVKQIATGVSD
5  TICEPCPVGFFSNVSSAFEKCHPWTSCETKDLVVQQAGTNKTDVVCGPQDRLRALVVIPIIFGILFAILL
6  VLVFIKKVAKKPTNKAAHPKQEPQEINFPDDLPGSNTAAPVQETLHGCQPVTQEDGKESRISVQERQ*
7
```

I208V.ali

Wild_Type > I208V.ali

```
1  >P1;P25942
2  sequence:::::::::
3  MVRLPLQCVLWGCLLTAVHPEPPTACREKQYLINSQCCSLCQPGQKLVSDCTEFTETECLPCGESEFLDT
4  WNRETHCHQHKYCDPNLGLRVQQKGTSETDTICTCEEGWHCTSEACESCVLHRSCSPGFGVKQIATGVSD
5  TICEPCPVGFFSNVSSAFEKCHPWTSCETKDLVVQQAGTNKTDVVCGPQDRLRALVVIPIIFGILFAVLL
6  VLVFIKKVAKKPTNKAPHPKQEPQEINFPDDLPGSNTAAPVQETLHGCQPVTQEDGKESRISVQERQ*
7
```

T112M.ali

Wild_Type > T112M.ali

```
1  >P1;P25942
2  sequence:::::::::
3  MVRLPLQCVLWGCLLTAVHPEPPTACREKQYLINSQCCSLCQPGQKLVSDCTEFTETECLPCGESEFLDT
4  WNRETHCHQHKYCDPNLGLRVQQKGTSETDTICTCEEGWHCMSEACESCVLHRSCSPGFGVKQIATGVSD
5  TICEPCPVGFFSNVSSAFEKCHPWTSCETKDLVVQQAGTNKTDVVCGPQDRLRALVVIPIIFGILFAILL
6  VLVFIKKVAKKPTNKAPHPKQEPQEINFPDDLPGSNTAAPVQETLHGCQPVTQEDGKESRISVQERQ*
7
```

- Based on the identified template, generate four models: one wild-type and three mutated versions. (8 marks: 2 marks per model)

(SEE VS CODE FOR THE 4 MODELS)

SELECTED MODEL BASED ON LOWEST MOLPDF SCORE (FOR EVALUATION)

For the wild type: P25942.B99990001.pdb

MUTATION 1: P25942.B99990001.pdb

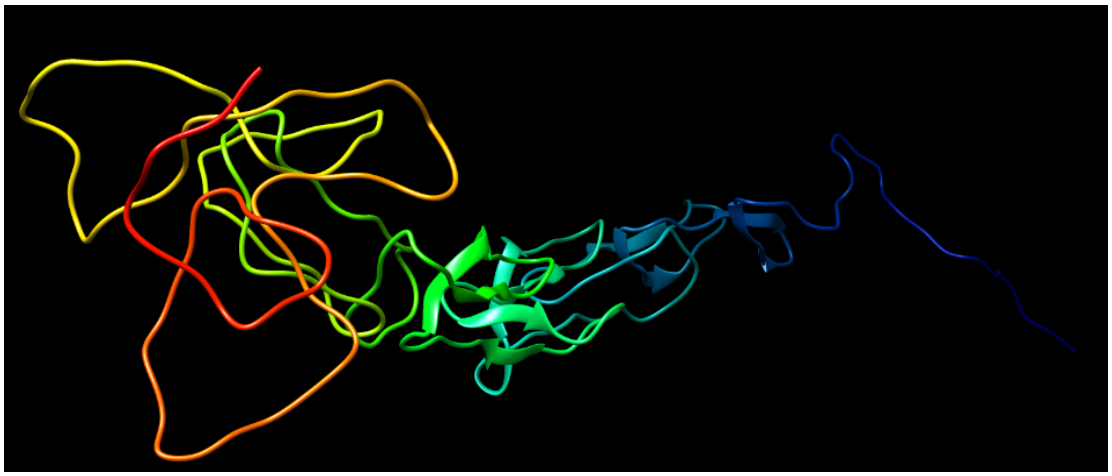
MUTATION 2: P25942.B99990001.pdb

MUTATION 3: P25942.B99990004.pdb

Visualization in Chimera (2 marks)

- Visualize all four models in Chimera, highlighting the mutated residues. (0.25*4)

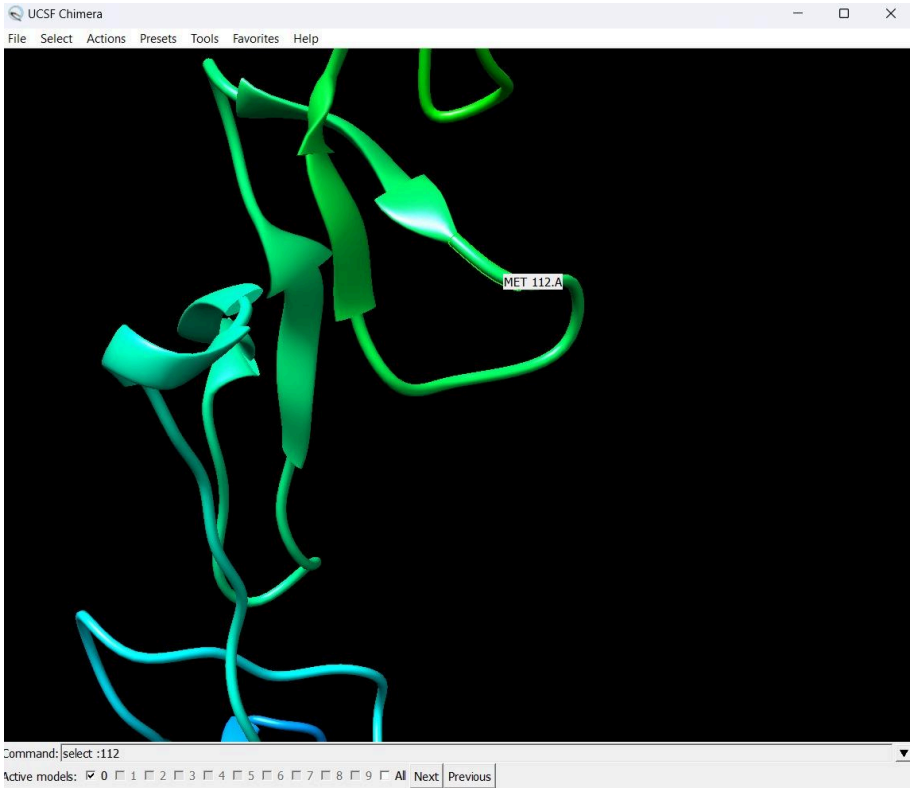
THE FIRST ONE IS THE WILD-TYPE SEQUENCE IN CHIMERA.



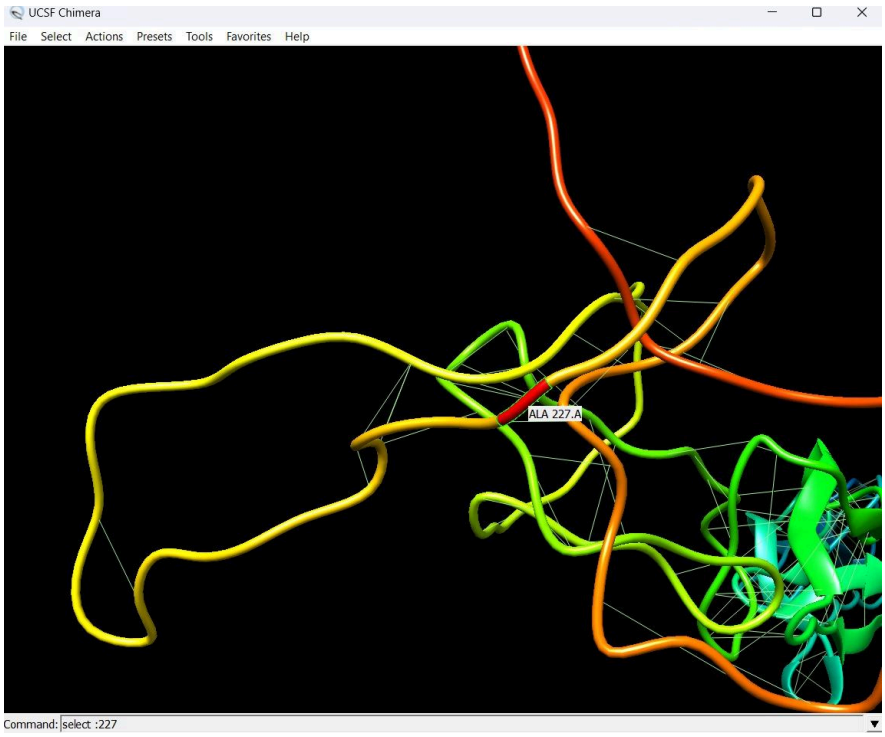
THE SECOND ONE IS P227A MUTATED SEQUENCE.

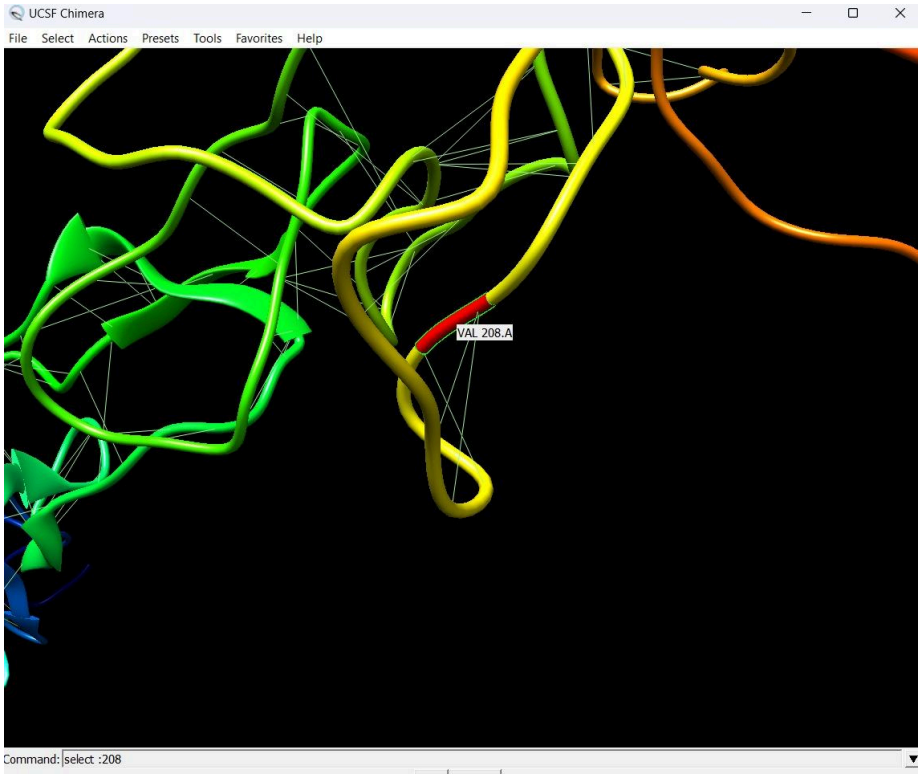
THIRD ONE IS I208V MUTATED SEQUENCE.

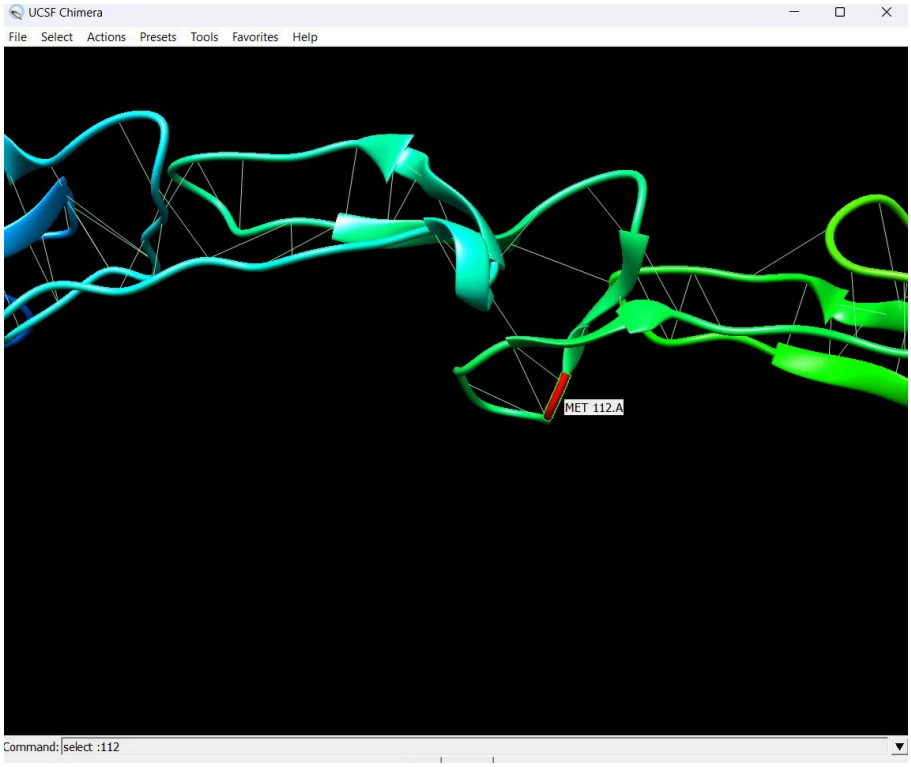
FOURTH ONE IS T112M MUTATED SEQUENCE.



- Clearly represent the interactions involving the mutated residues, such as hydrogen bonds. (0.25*4, Marks awarded for clarity and uniform representation)





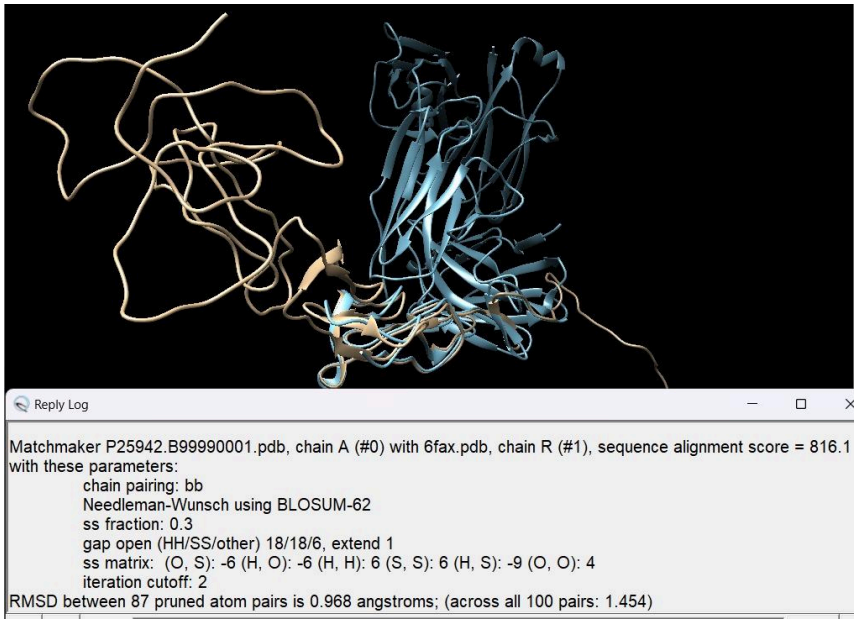


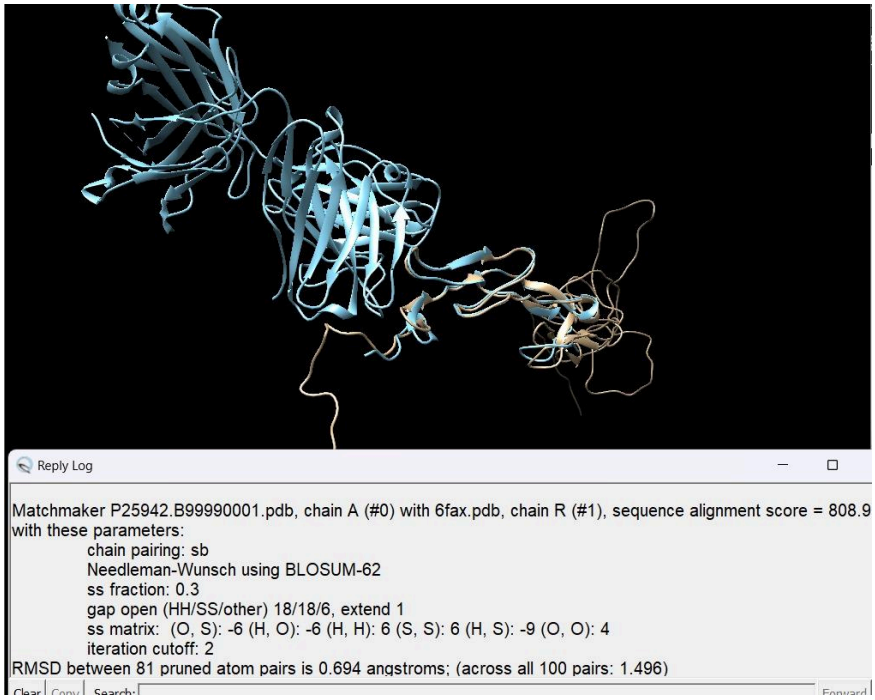
SEQUENCE OF IMAGES: WILD-TYPE, MUTATION 1, MUTATION 2, MUTATION 3

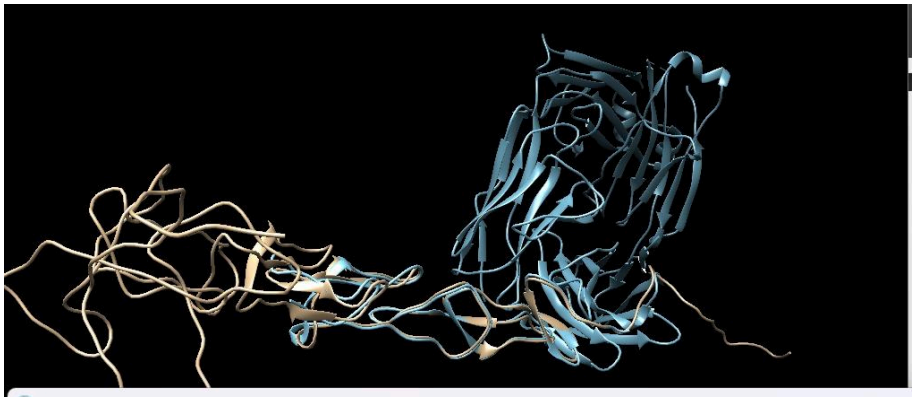
Comparison Using DOPE Score & RMSD (5.5 marks)

- Compare the wild-type and mutated models using DOPE score (1 marks) and RMSD analysis (1marks)

SEQUENCE	WILD_TYPE	P227A	I208V	T112M
DOPE SCORE	-16636.890625	-16992.107422	-16755.902344	-17134.009766

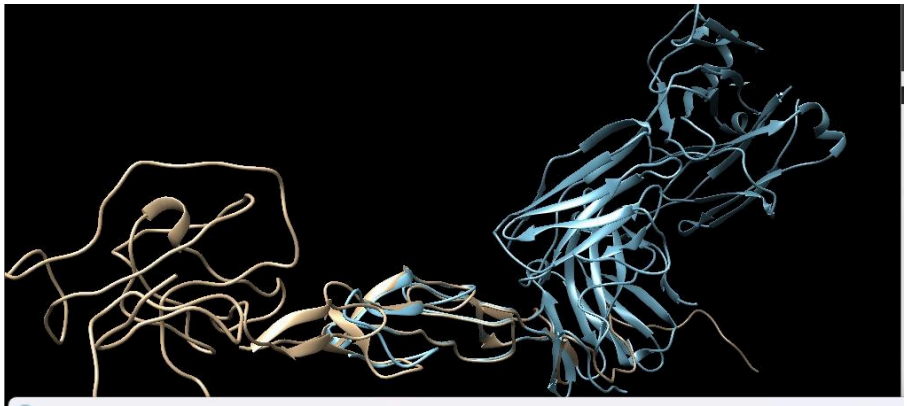






Reply Log

```
Matchmaker P25942.B99990001.pdb, chain A (#0) with 6fax.pdb, chain R (#1), sequence alignment score = 816.1
with these parameters:
        chain pairing: sb
        Needleman-Wunsch using BLOSUM-62
        ss fraction: 0.3
        gap open (HH/SS/other) 18/18/6, extend 1
        ss matrix:  (O, S): -6 (H, O): -6 (H, H): 6 (S, S): 6 (H, S): -9 (O, O): 4
        iteration cutoff: 2
RMSD between 90 pruned atom pairs is 0.853 angstroms; (across all 100 pairs: 1.279)
```



Reply Log

```
Matchmaker P25942.B99990004.pdb, chain A (#0) with 6fax.pdb, chain R (#1), sequence alignment score = 811.9
with these parameters:
        chain pairing: sb
        Needleman-Wunsch using BLOSUM-62
        ss fraction: 0.3
        gap open (HH/SS/other) 18/18/6, extend 1
        ss matrix:  (O, S): -6 (H, O): -6 (H, H): 6 (S, S): 6 (H, S): -9 (O, O): 4
        iteration cutoff: 2
RMSD between 79 pruned atom pairs is 0.792 angstroms; (across all 100 pairs: 1.837)
```

Wild-type RMSD:

- Sequence alignment score: **816.1**
- RMSD between 87 pruned pairs: **0.968 Å**
- RMSD across 100 pairs: **1.454 Å**

P227A mutation:

- Sequence alignment score: **808.9**
- RMSD between 81 pruned pairs: **0.694 Å**

- RMSD across 100 pairs: **1.496 Å**

I208V mutation:

- Sequence alignment score: **816.1**
- RMSD between 90 pruned pairs: **0.853 Å**
- RMSD across 100 pairs: **1.279 Å**

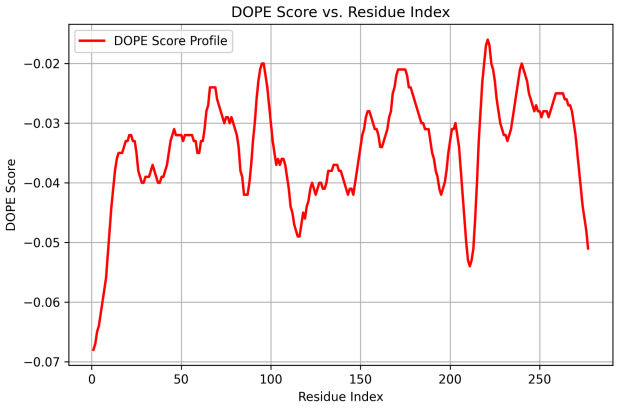
T112M mutation:

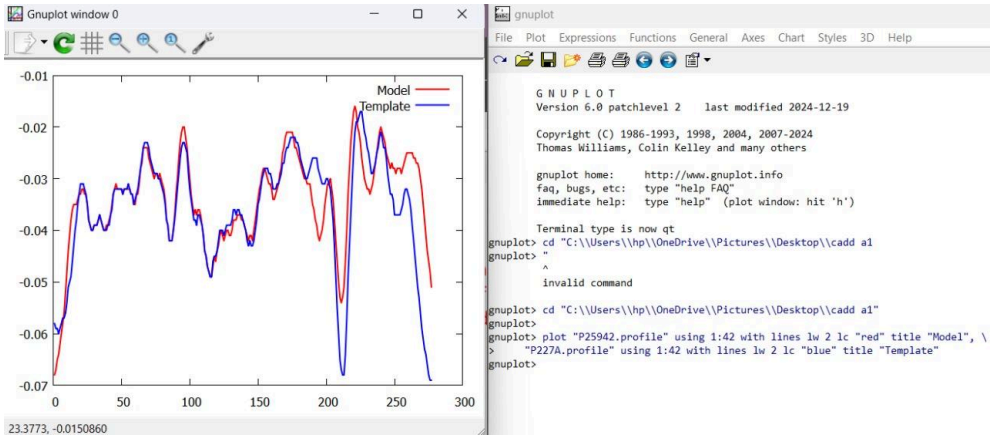
- Sequence alignment score: **811.9**
- RMSD between 79 pruned pairs: **0.792 Å**
- RMSD across 100 pairs: **1.837 Å**

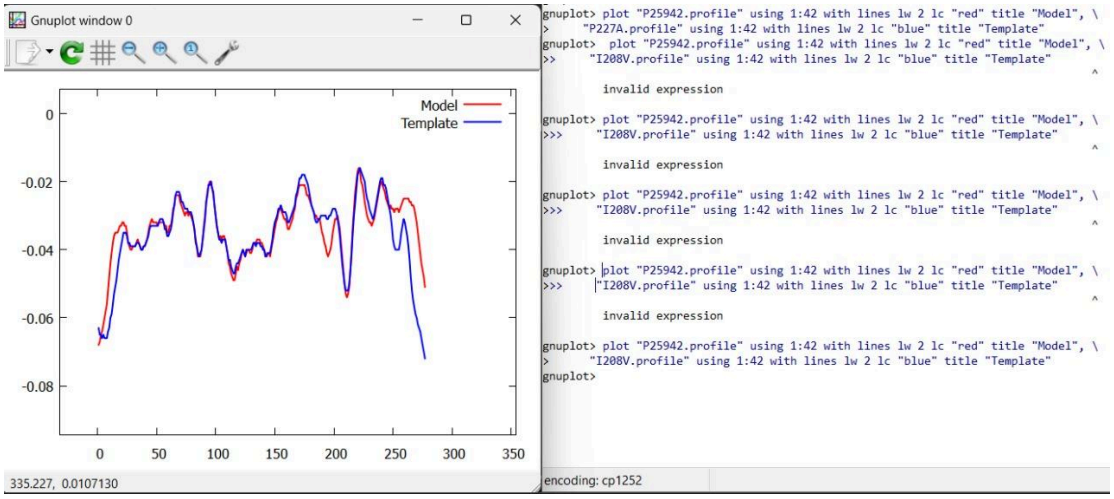
- Plot DOPE score profile for all 4 models (0.5*4=2 marks)

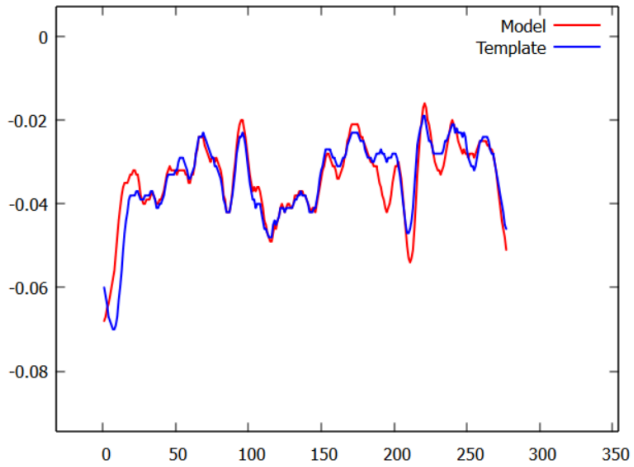
The first one is WILD_TYPE DOPE SCORE PLOT.

The second one is Mutated sequence 1, then 2, then 3.









- Report observations and provide insights based on the analysis. (1.5 mark)

MODEL	DOPE SCORE	RMSD PRUNED PAIR	RMSD 100 PAIRS	OBSERVATIONS
WILD TYPE	-16636.89	0.968 A	1.454 A	Best DOPE score, highest structural stability.
P227A	-16992.11	0.694 A	1.496 A	Slight decrease in stability, minimal structural changes but some distortion.
I208V	-16755.90	0.853 A	1.279 A	Moderate structural impact, moderate RMSD change.
T112M	-17134.01	0.792 A	1.837 A	Significant structural instability, largest structural deviation.

Literature Review on Mutation Significance (2 marks)

- Research and summarize any reported significance of the given mutations (1.5 mark)

The **P25942** gene encodes **CD40**, a receptor in the tumor necrosis factor receptor superfamily, crucial for immune responses. Mutations in CD40 can lead to immunodeficiencies and autoimmune diseases.

P227A Mutation: This mutation involves substituting proline with alanine at position 227. Proline's rigid structure is vital for maintaining protein conformation. Replacing it with alanine may disrupt the receptor's structure, potentially impairing its function. While specific studies on P227A are limited, similar mutations in the CD40 gene have been linked to immunodeficiencies.

I208V Mutation: This mutation replaces isoleucine with valine at position 208. Both are hydrophobic amino acids, but valine is smaller. Such a change could affect the protein's hydrophobic core, potentially altering its stability and function. Mutations in the CD40 gene have been associated with hyper IgM syndrome, an immunodeficiency disorder.

T112M Mutation: This mutation substitutes threonine with methionine at position 112. Methionine is larger and more hydrophobic than threonine, which could disrupt the protein's structure and function. Mutations in the CD40 gene have been linked to immunodeficiencies and autoimmune diseases.

- Provide references to relevant literature (0.5 mark)
- https://pmc.ncbi.nlm.nih.gov/articles/PMC60102/