BioPhysics
Assignment 1
Report

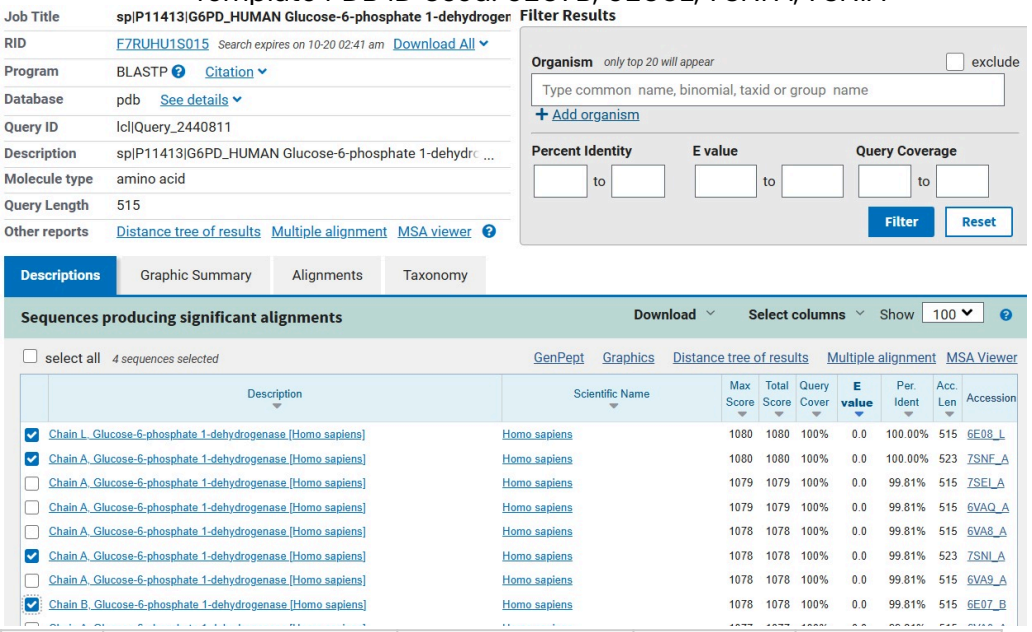
Group number : 5
Pragya Singh 2023379
Aditya Verma 2023051
Abhijaya Pal 2023019
Arhan Jain 2023118
Akshat Lakhera 2023061

Protein Assigned : Glucose-6-phosphate dehydrogenase
UniProt ID : P11413
Github Link : https://github.com/Adit1414/BioPhysics_Assignment_Group5

---

1. Template Identification

- For the G6PD protein sequence in humans, wild type protein sequence with uniprot ID P11413 was identified.
- Template PDB ID Used: 6E07B, 6E08L, 7SNFA, 7SNIA



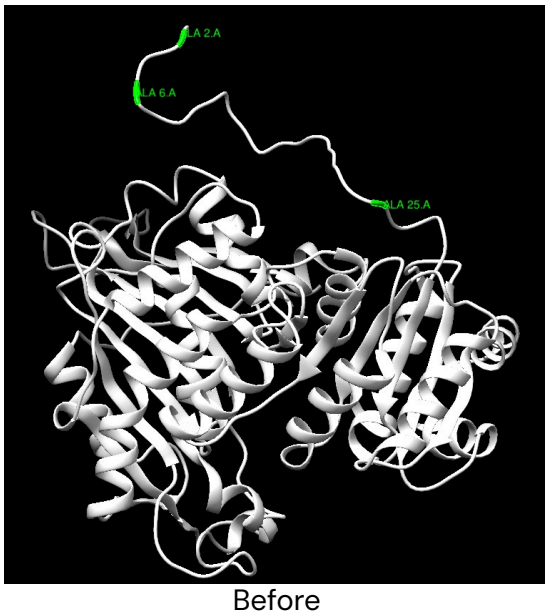
PDB ID	Sequence Identity	Query Coverage	Resolution	Organism
6E08	100%	100%	2.28A	Homo Sapiens
7SNF	100%	100%	3.5A	Homo Sapiens
7SNI	99.81%	100%	2.5A	Homo Sapiens
E07	99.81%	100%	2.6A	Homo Sapiens

4 templates chosen on the basis of Query coverage, Percentage Identity, Resolution. Smaller resolution is better.

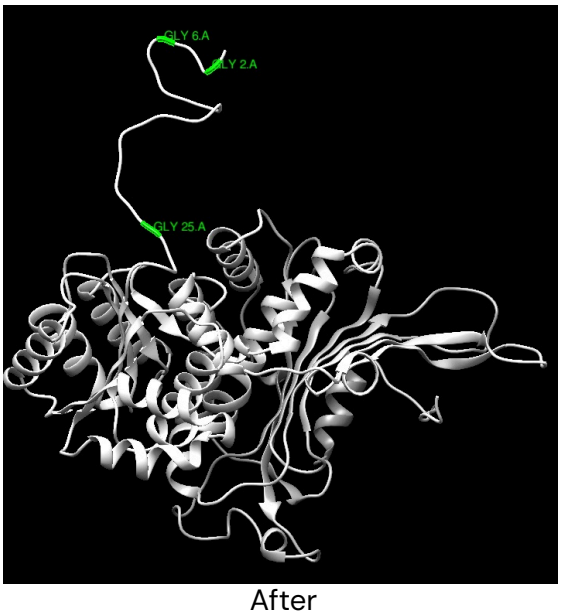
---

2. Modeling the Wild-Type and Mutant

a) Mutation Introduction



Before



After

Serial Number	Residue Number	Original Residue	Mutated Residue	Remark
1	2	Alanine	Glycine	Hydrophobic Non-Polar -> Small Non-Polar
2	6	Alanine	Glycine	Hydrophobic Non-Polar -> Small Non-Polar
3	25	Alanine	Glycine	Hydrophobic Non-Polar -> Small Non-Polar

b) Homology Modeling

Wild Type Model ID	Molpdf Score	Dope Score
P11413.B99990001.pdb	22870.92383	-62459.09766
P11413.B99990002.pdb	22728.62305	-62513.58594
P11413.B99990003.pdb	22961.27344	-62620.66016

The **best structure** is the **third** model, ie, **P11413.B99990003.pdb**.

This is evaluated by the **least dope score.**

As shown in the table, in WT, the third model has a dope score of **-62620.66016**, which is the least of the three. A low dope score predicts how likely is the existence of the protein model.

Mutated Type Model ID	Molpdf Score	Dope Score
P11413M.B99990001.pdb	22941.26172	-62384.14063
P11413M.B99990002.pdb	22595.25195	-62538.85938
P11413M.B99990003.pdb	22842.00586	-62341.17578

The **best structure** is the **second** model, ie **P11413M.B99990002.pdb.**
This is evaluated by the **least dope score.**
As shown in the table, in MT, the second model has a dope score of **-62538.85938**, which is the least of the three.

- Wild Type Models

```
<< end of ENERGY.

>> Summary of successfully produced models:
Filename                          molpdf
----------------------------------------
P11413.B99990001.pdb          22870.92383
P11413.B99990002.pdb          22728.62305
P11413.B99990003.pdb          22961.27344

Total CPU time [seconds]                                  :      80.28
```

```
openf___224_> Open           dope_shope_P11413_B99990002.profile
# Energy of each residue is written to: dope_shope_P11413_B99990002.profile
# The profile IS normalized by the number of restraints.
# The profiles are smoothed over a window of residues:    13
# The sum of all numbers in the file:      -18.6481

<< end of ENERGY.
DOPE score               : -62513.585938
```

```
openf___224_> Open           dope_shope_P11413_B99990001.profile
# Energy of each residue is written to: dope_shope_P11413_B99990001.profile
# The profile IS normalized by the number of restraints.
# The profiles are smoothed over a window of residues:    13
# The sum of all numbers in the file:      -18.8778

<< end of ENERGY.
DOPE score               : -62459.097656
```

```
openf___224_> Open           dope_shope_P11413_B99990003.profile
# Energy of each residue is written to: dope_shope_P11413_B99990003.profile
# The profile IS normalized by the number of restraints.
# The profiles are smoothed over a window of residues:    13
# The sum of all numbers in the file:      -18.5453

<< end of ENERGY.
DOPE score               : -62620.660156
```

- Mutated Type Models

```
<< end of ENERGY.

>> Summary of successfully produced models:
Filename                          molpdf
----------------------------------------
P11413M.B99990001.pdb         22941.26172
P11413M.B99990002.pdb         22595.25195
P11413M.B99990003.pdb         22842.00586

Total CPU time [seconds]                                  :     121.16
```

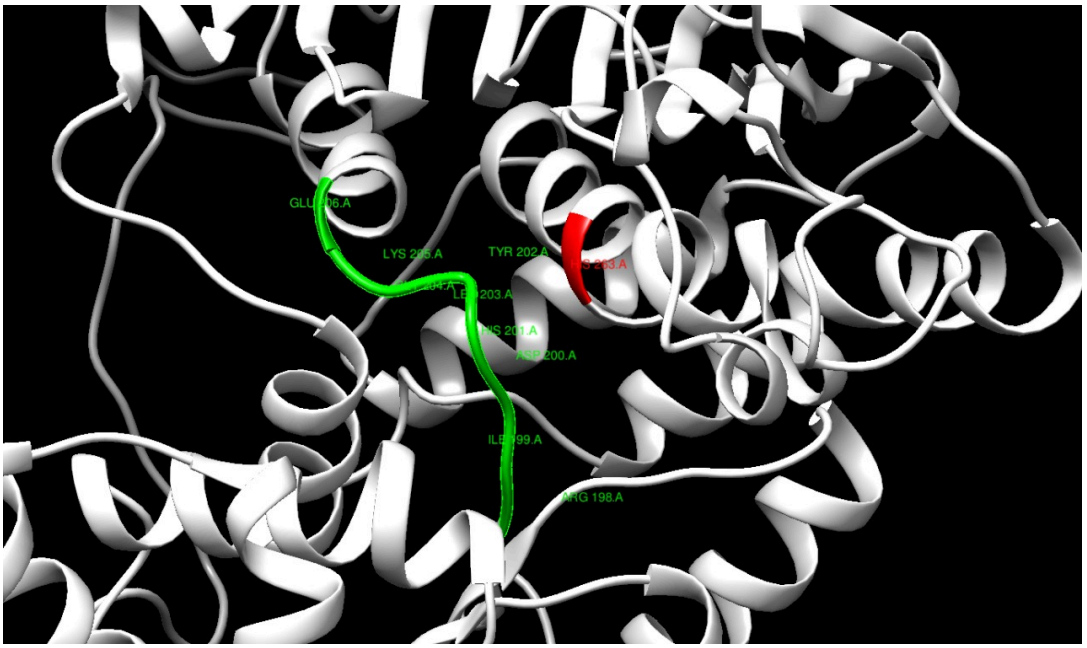
```
openf___224_> Open           dope_shope_P11413M_B99990001.profile
# Energy of each residue is written to: dope_shope_P11413M_B99990001.profile
# The profile IS normalized by the number of restraints.
# The profiles are smoothed over a window of residues:    13
# The sum of all numbers in the file:      -18.7355

<< end of ENERGY.
DOPE score               : -62384.140625
```

```
openf___224_> Open           dope_shope_P11413M_B99990002.profile
# Energy of each residue is written to: dope_shope_P11413M_B99990002.profile
# The profile IS normalized by the number of restraints.
# The profiles are smoothed over a window of residues:    13
# The sum of all numbers in the file:      -18.7591

<< end of ENERGY.
DOPE score               : -62538.859375
```

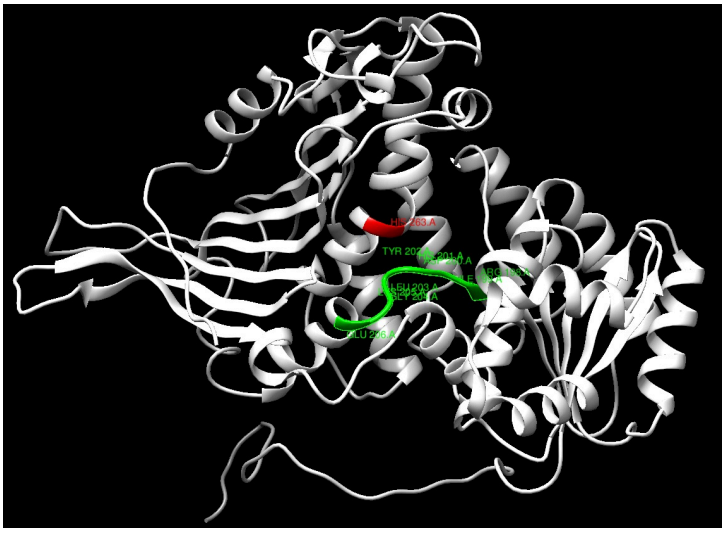
```
openf___224_> Open           dope_shope_P11413M_B99990003.profile
# Energy of each residue is written to: dope_shope_P11413M_B99990003.profile
# The profile IS normalized by the number of restraints.
# The profiles are smoothed over a window of residues:    13
# The sum of all numbers in the file:      -18.8601

<< end of ENERGY.
DOPE score               : -62341.175781
```

3. Visualization in Chimera

- **Catalytic Site Residues**



HIS 263 is primary chemical actor, it relies directly on the resides in the binding site at position 198-206 RIDHYLGKE, without which it cannot take part in reaction, thus essential for catalytic activity.

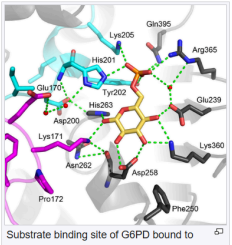


Reference :

[Uniprot protein structure webpage]

[Wikipedia article on G6PD]

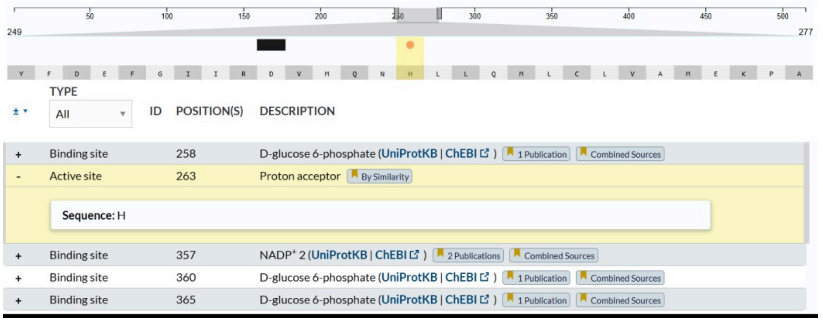
Enzyme structure [edit]



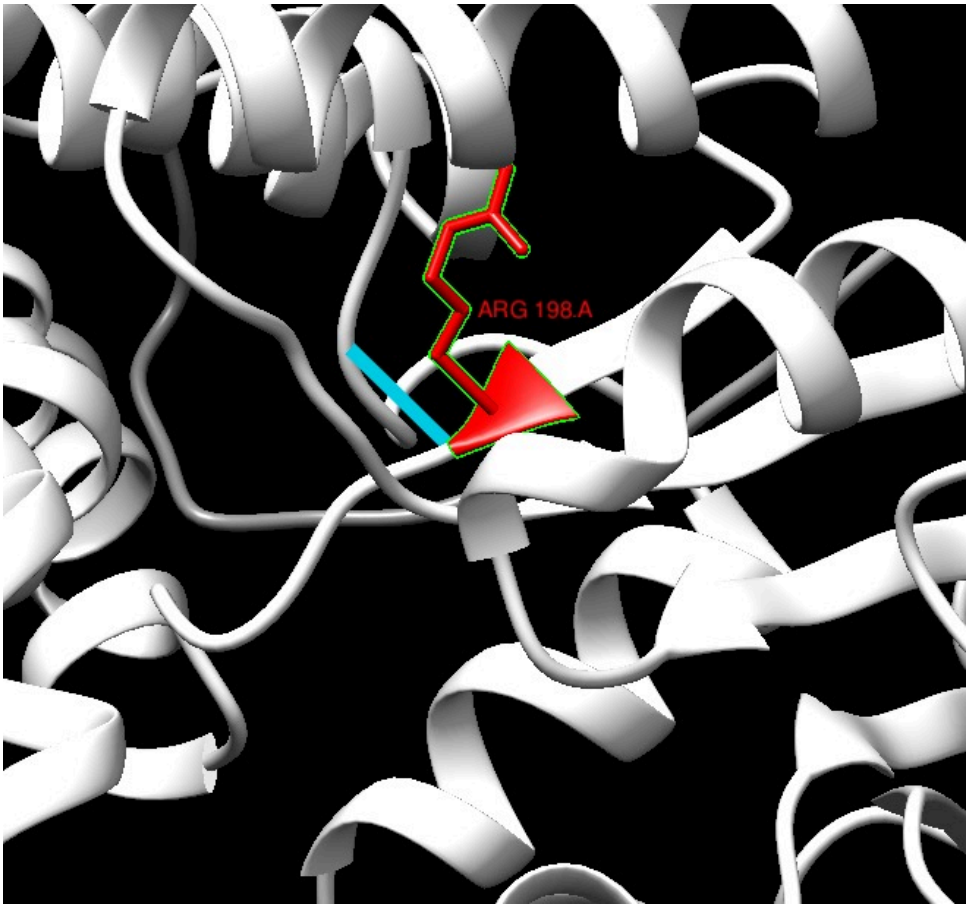
Substrate binding site of G6PD bound to G6P (shown in cream), from 2BHL. Phosphorus is shown in orange. Oxygen atoms of crystallographic waters are shown as red spheres. The conserved nine-peptide sequence of G6PD, and the partially conserved five-residue sequence of G6PD are shown in cyan and magenta respectively. All other amino acids from G6PD are shown in black. Hydrogen bonding and electrostatic interactions are shown by green dashed lines.

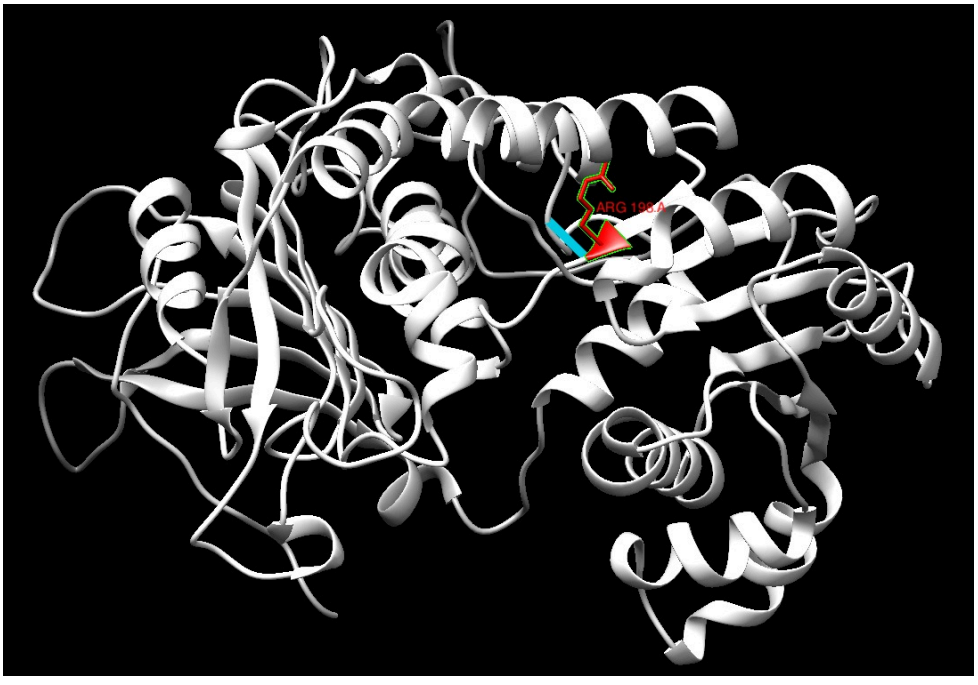
G6PD is generally found as a dimer of two identical monomers (see main thumbnail).[8] Depending on conditions, such as pH, these dimers can themselves dimerize to form tetramers.[5] Each monomer in the complex has a substrate binding site that binds to G6P, and a catalytic coenzyme binding site that binds to $NADP^+$/NADPH using the Rossman fold.[4] For some higher organisms, such as humans, G6PD contains an additional $NADP^+$ binding site, called the $NADP^+$ structural site, that does not seem to participate directly in the reaction catalyzed by G6PD. The evolutionary purpose of the $NADP^+$ structural site is unknown.[4] As for size, each monomer is approximately 500 amino acids long (514 amino acids for humans[5]).

Functional and structural conservation between human G6PD and *Leuconostoc mesenteroides* G6PD points to three widely conserved regions on the enzyme: a nine-residue peptide in the substrate binding site, RIDHYLGKE (residues 198–206 on human G6PD), a nucleotide-binding fingerprint, GxxGDLA (residues 38–44 on human G6PD), and a partially conserved sequence EKPxG near the substrate binding site (residues 170–174 on human G6PD), where we have use "x" to denote a variable amino acid.[4] The crystal structure of G6PD reveals an extensive network of electrostatic interactions and hydrogen bonding involving G6P, three water molecules, three lysine residues, one arginine, two histidines, two glutamic acids, and other polar amino acids.



❖ **H-bonds of Residue 198**

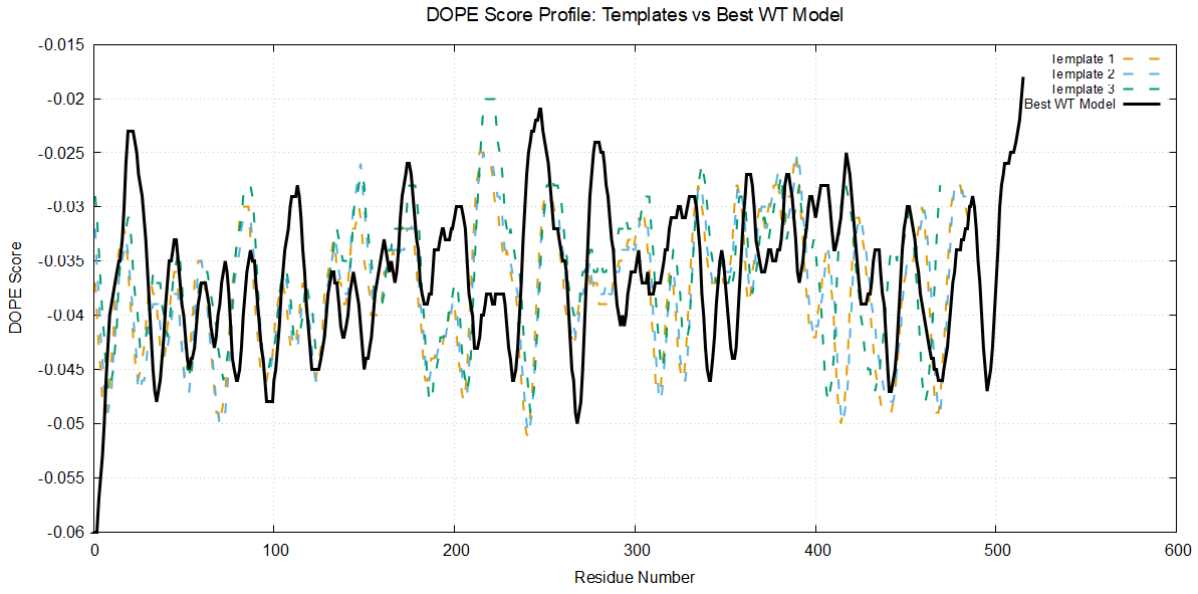




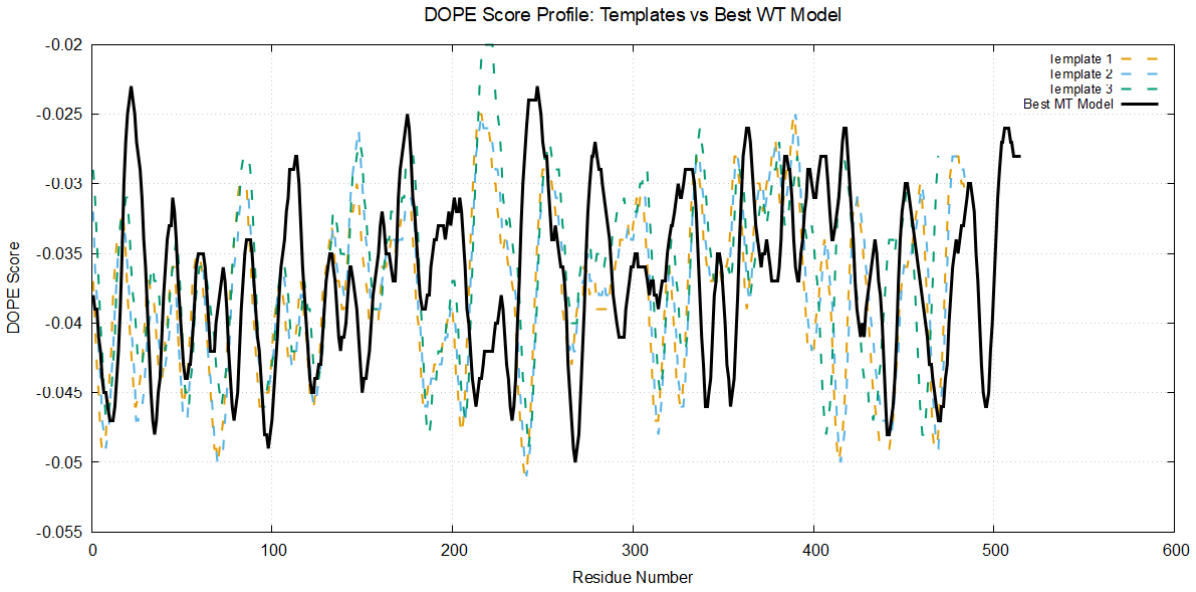
4. Structural Comparison: DOPE Score and RMSD

a) DOPE Score Analysis

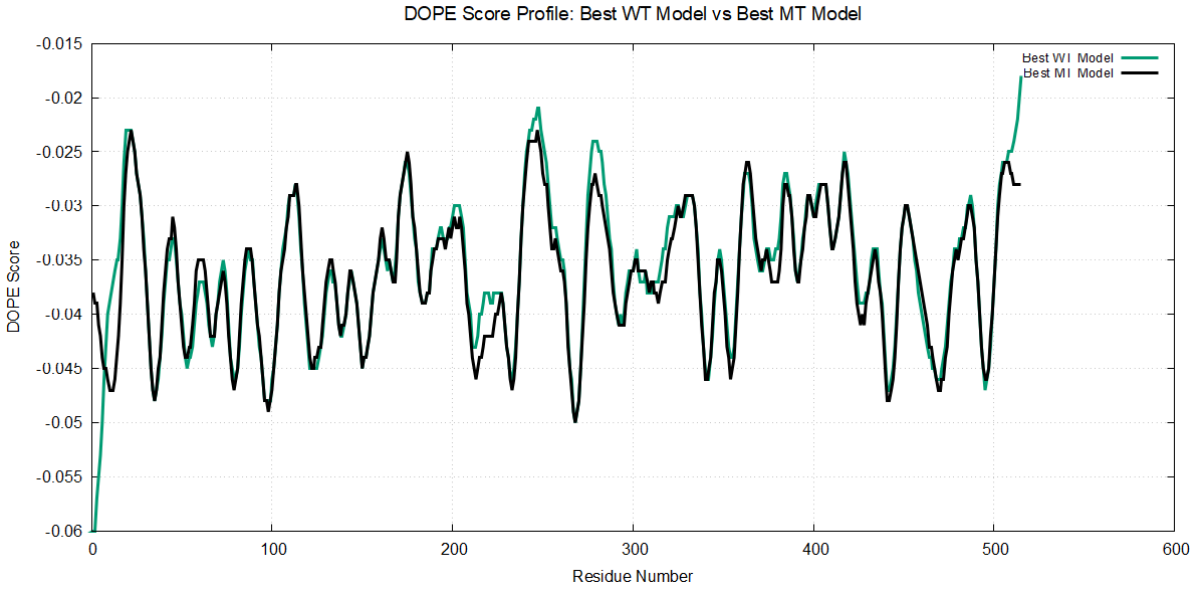
Model	Dope Score
P11413	-62620.66016
P11413M	-62538.85938
6E07	-62609.61719
6E08	-62870.41016
7SNF	-59044.46094
7SNI	-62051.26953



Templates vs Best WT Model – shows DOPE profiles of the three chosen templates and the selected wild-type model.



Templates vs Best Mutant Model – shows DOPE profiles of the same templates and the selected mutant model.



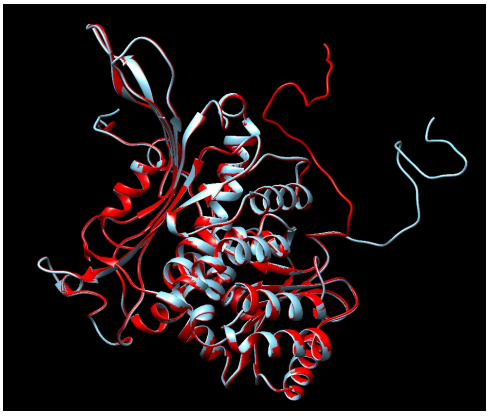
Best Wild Model vs Best Mutant Model - shows DOPE profiles comparison between the best wild and mutant models.

Interpretation

With DOPE scores, a lower (more negative) number means the protein structure is more stable. We saw that the experimental templates had scores ranging from -16.9483 (7SNF) to -18.3051 (6E07), which gave us a good benchmark for a high-quality structure. Our Wild Type (WT) model scored -18.5453 and our Mutant (MT) model scored -18.7591. Both of these scores are right in the same region as (and even slightly better than) the

templates, which confirms that our models are well-folded and physically realistic. We did see a slight *stabilization* in the MT model, as its score was slightly more negative than the WT's. This finding is also backed up by the per-residue energy graphs, where both of our models' energy profiles followed the templates' profiles very closely. Most importantly, when we directly compared the two models, their energy profiles were almost identical. This clearly shows that the mutation was structurally conservative—it didn't introduce any new unstable spots or change the protein's overall fold.

b) RMSD Analysis



```
Computing secondary structure assignments for model(s) #1, #0
using ksdssp (Kabsch and Sander Define Secondary Structure
of Proteins) with the parameters:
  energy cutoff -0.5
  minimum helix length 3
  minimum strand length 3

Matchmaker P11413.B99990003.pdb, chain A (#1) with P11413M.B99990002.pdb, chain A (#0), sequence alignment score = 2675.6
with these parameters:
        chain pairing: bb
        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 463 pruned atom pairs is 0.184 angstroms; (across all 515 pairs: 5.174)
```

RMSD between 463 pruned atom pairs is 0.184 angstroms.

c) Interpretation & Insights

The structural analysis of the models indicate **minor effects** on the protein. The DOPE score for the best wild type model -62620.66, is **slightly lesser** than the DOPE score for the best mutated type model -62538.86. This minor change in the score depicts a minor destabilizing effect on the mutated protein sequence, thus a **slight increase in free energy**.

The RMSD score, by superimposing the best wild type model with the best mutant type model, shows a **very close alignment of the models**. The **low RMSD** value indicates that the energy destabilization due to introduced mutation **doesn't translate into any significant conformation changes**. The protein's **folds are preserved**, thus substituting Aniline with a small and flexible Glycine is structurally conservative and doesn't cause any significant protein rearrangement.