Project Module 2

By Yash Dhiman

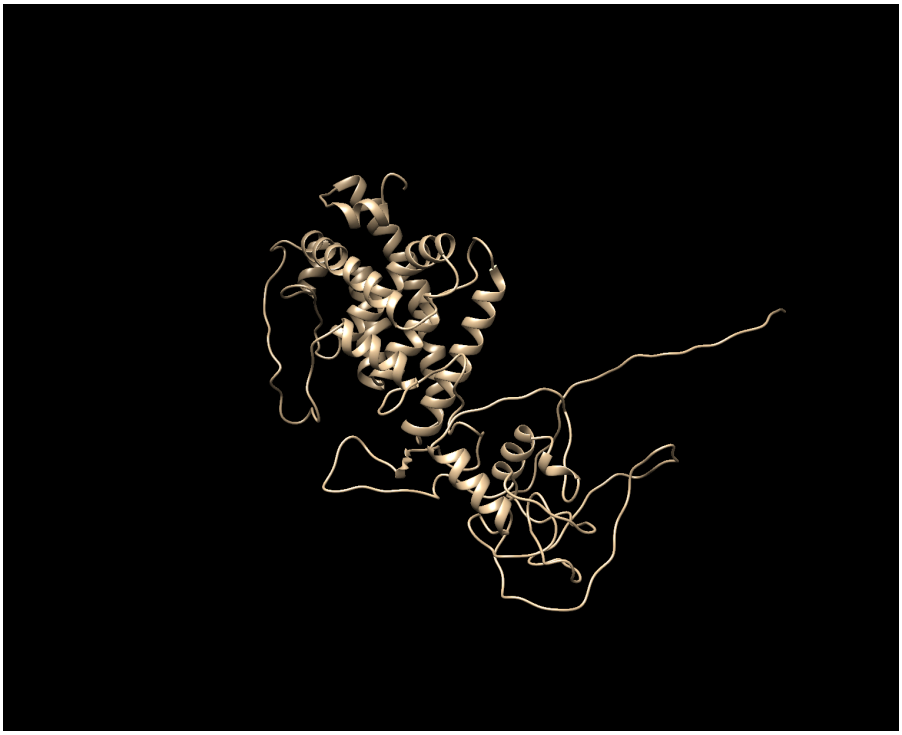
Steps involved in finding the mutated protein based on the CHEMBL ID:

1. Ligand Preparation
    a. The CHEMBL IDs provided were used to search for the corresponding ligand molecules in a chemical database like ChEMBL.
    b. For each ligand identified by its CHEMBL ID, the SDF (Structure-Data File) format was downloaded. This format is commonly used to represent molecular structures in 2D or 3D.
    c. The SDF files were then converted into PDBQT format, which includes information about the partial charges (PDB) and atom types and affinities (QT) necessary for docking. This conversion is crucial for compatibility with many docking software.
2. Protein Preparation
    a. Protein structures (both native and mutant versions) relevant to the project were prepared, possibly involving steps like removing water molecules, adding hydrogen atoms, and identifying the active site or binding pocket.
    b. The protein structures were also converted into a format compatible with the chosen docking software, often PDBQT as well.
3. Virtual Screening
    a. Using docking software (e.g., AutoDock Vina, Dock, or others), each prepared ligand molecule was docked into the active site of both the native and mutant protein structures.
    b. The docking process involves computationally simulating how each ligand fits into the protein's binding site,

evaluating the binding affinity, and predicting the most likely bound conformation.

4. Analysis of Docking Results
    a. The docking results were analyzed to identify ligands with the highest binding affinity to the protein targets. This involves looking at docking scores, binding poses, and interactions between the ligand and amino acids in the binding site.
    b. Comparisons were likely made between how each ligand interacted with the native versus mutant protein structures to assess the impact of mutations on binding affinity.
5. Data Organization and Submission
    a. All the ligand files (in PDBQT format) and the docked structures (also typically in PDBQT or a similar format showing the ligand-protein complexes) were compiled.

Initial protein Structure:



The structure of the protein (without ligand and mutation)

The config.txt for the grid:

```
receptor = yd1_B99990001.pdbqt

center_x = 3.433
center_y = 57.875
center_z = 4.014

size_x = 50
size_y = 40
size_z = 40

num_modes = 10
energy_range = 4
```

Virtual Screening using AutoDock Vina

```
WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)
Output will be CHEMBL1269025pdbqt_out.pdbqt
Detected 8 CPUs
Reading input ... done.
Setting up the scoring function ... done.
Analyzing the binding site ... done.
Using random seed: -1000378636
Performing search ... done.
Refining results ... done.

mode |   affinity | dist from best mode
     | (kcal/mol) | rmsd l.b.| rmsd u.b.
-----+------------+----------+----------
   1        -10.2      0.000      0.000
   2        -10.1     23.196     24.778
   3         -9.8     23.554     25.219
   4         -9.7      1.381      1.735
   5         -9.6      2.033      2.701
   6         -9.5     23.454     24.993
   7         -9.5     15.724     17.301
   8         -9.5      2.318      4.897
   9         -9.5      1.562      3.236
  10         -9.5     21.771     23.354
```

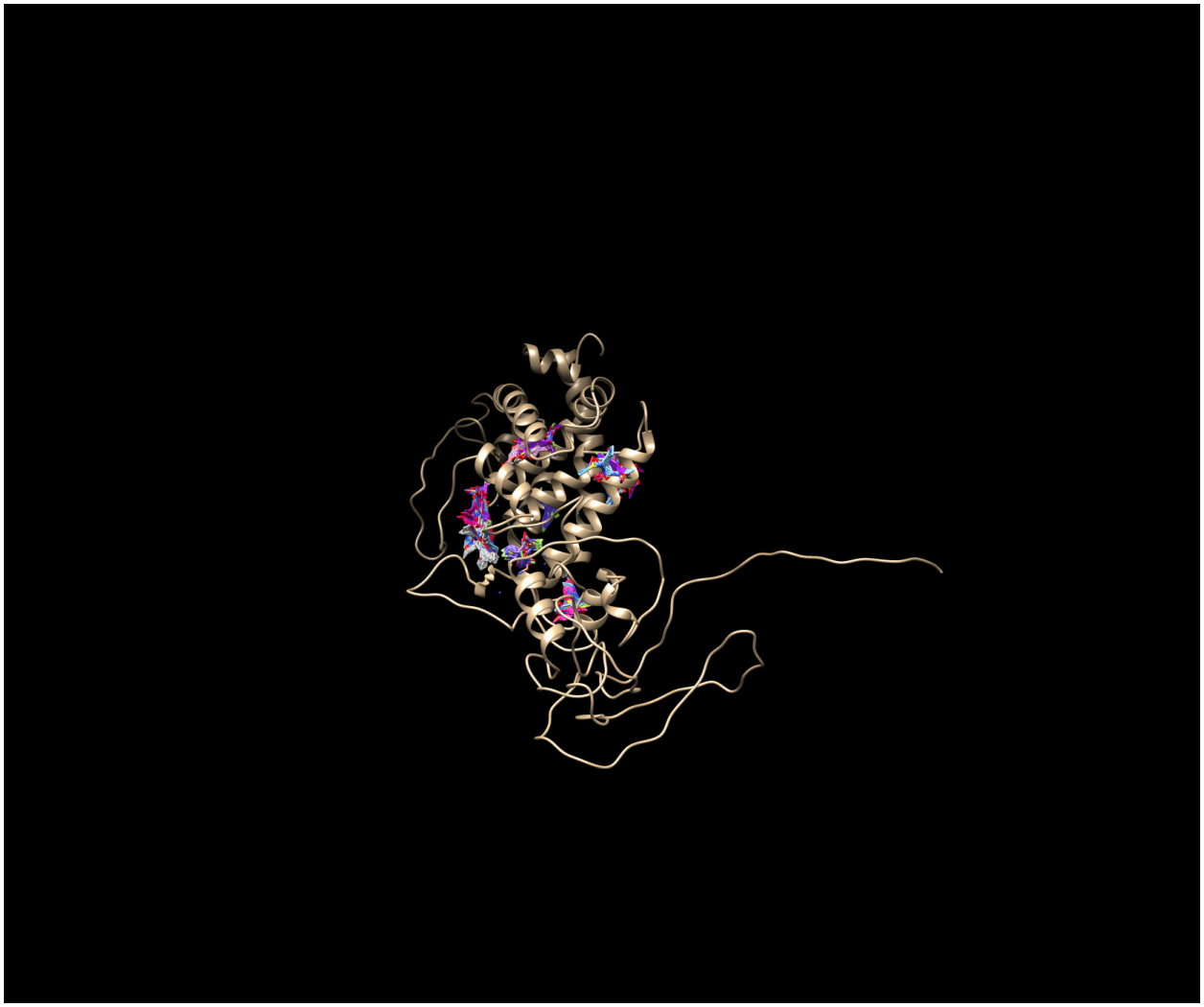
```
WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)
Output will be CHEMBL456237pdbqt_out.pdbqt
Detected 8 CPUs
Reading input ... done.
Setting up the scoring function ... done.
Analyzing the binding site ... done.
Using random seed: 1314865152
Performing search ... done.
Refining results ... done.

mode |   affinity | dist from best mode
     | (kcal/mol) | rmsd l.b.| rmsd u.b.
-----+------------+----------+----------
   1        -13.1      0.000      0.000
   2        -12.7     21.591     23.282
   3        -12.3      1.161      3.364
   4        -12.0     21.227     23.080
   5        -11.4      1.291      4.154
   6        -11.2      2.171      3.197
   7        -11.1      1.180      3.256
   8        -10.6     21.329     23.560
   9        -10.4     21.968     24.226
  10        -10.3      1.640      3.361
Writing output ... done.
```

```
WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)
Output will be CHEMBL573677pdbqt_out.pdbqt
Detected 8 CPUs
Reading input ... done.
Setting up the scoring function ... done.
Analyzing the binding site ... done.
Using random seed: 1848553424
Performing search ... done.
Refining results ... done.

mode |   affinity | dist from best mode
     | (kcal/mol) | rmsd l.b.| rmsd u.b.
-----+------------+----------+----------
   1         -1.6      0.000      0.000
   2         -1.5     11.292     11.292
   3         -1.4     34.006     34.006
   4         -1.4     33.323     33.323
   5         -1.3     22.713     22.713
   6         -1.3     35.683     35.683
   7         -1.3     19.047     19.047
   8         -1.3     27.728     27.728
   9         -1.3     20.045     20.045
  10         -1.3     13.544     13.544
Writing output ... done.
```

```
WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)
Output will be CHEMBL1434513pdbqt_out.pdbqt
Detected 8 CPUs
Reading input ... done.
Setting up the scoring function ... done.
Analyzing the binding site ... done.
Using random seed: 1916086576
Performing search ... done.
Refining results ... done.

mode |   affinity | dist from best mode
     | (kcal/mol) | rmsd l.b.| rmsd u.b.
-----+------------+----------+----------
   1         -5.5      0.000      0.000
   2         -5.5      3.616      5.487
   3         -5.4     22.227     23.440
   4         -5.2     21.108     23.060
   5         -5.2     20.962     22.538
   6         -4.8     19.856     21.081
   7         -4.8     19.766     21.043
   8         -4.7     18.693     20.417
   9         -4.6     25.077     25.561
  10         -4.5     24.960     25.735
Writing output ... done.
```

```
WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)
Output will be CHEMBL2105747pdbqt_out.pdbqt
Detected 8 CPUs
Reading input ... done.
Setting up the scoring function ... done.
Analyzing the binding site ... done.
Using random seed: 1271640180
Performing search ... done.
Refining results ... done.

mode |   affinity | dist from best mode
     | (kcal/mol) | rmsd l.b.| rmsd u.b.
-----+------------+----------+----------
   1         -9.4      0.000      0.000
   2         -9.3     23.281     24.433
   3         -9.1      1.695      2.659
   4         -8.9      1.969      2.523
   5         -8.8     23.166     24.779
   6         -8.8     23.028     24.958
   7         -8.8     23.264     25.187
   8         -8.8     23.304     25.373
   9         -8.7      0.959      2.598
  10         -8.7     23.637     25.452
Writing output ... done.
```

```
WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)
Output will be CHEMBL2105709pdbqt_out.pdbqt
Detected 8 CPUs
Reading input ... done.
Setting up the scoring function ... done.
Analyzing the binding site ... done.
Using random seed: -1088433800
Performing search ... done.
Refining results ... done.

mode |   affinity | dist from best mode
     | (kcal/mol) | rmsd l.b.| rmsd u.b.
-----+------------+----------+----------
   1        -10.9      0.000      0.000
   2        -10.5      3.164      3.936
   3        -10.4      1.394      2.141
   4        -10.4      3.438      4.533
   5        -10.2      2.183      4.253
   6        -10.2      3.499      4.567
   7        -10.2      2.413      3.064
   8        -10.1      1.757      2.120
   9        -10.1      1.480      1.618
  10        -10.0      3.932      4.242
Writing output ... done.
```



Visualization of the protein after mutation