EX.NO : 1	**Molecular Visualization: PYmol**
Date : 13/08/24	

Aim: To visualize the structure of a protein and its properties using PYmol.

Material/tools:

- PDB, PDB ID: 1NDZ
- PyMOL

INTRODUCTION:

PDB

PDB is a worldwide central repository of structural information of biological macromolecules and is currently managed by the Research Collaboratory for Structural Bioinformatics (RCSB).

The Protein Data Bank (PDB) was established at Brookhaven National Laboratories (BNL) as an archive for biological macromolecular crystal structures. In the 1980s the number of deposited structures began to increase dramatically. This was due to the improved technology for all aspects of the crystallographic process, the addition of structures determined by nuclear magnetic resonance (NMR) methods, and changes in the community views about data sharing.

Initial use of the PDB had been limited to a small group of experts involved in structural research. Today depositors to the PDB have varying expertise in the techniques of X-ray crystal structure determination, NMR, cryoelectron microscopy and theoretical modeling.

PDB ID:

Identifiers or IDs are commonly used in data resources to point to specific data contents. They may also be used to connect different data resources and indicate their relationships. In the PDB, identifiers are used at all levels of the structural hierarchy in the entry. This includes:

- 4-character PDB ID for the entry
- Numeric ID for the assemblies in the entry
- or 2-character chain ID for instances of entities
- 3-character IDs for residues and small molecules
- “ATOM” or “HETATM” ID, and 4-character atom names for individual atoms

These identifiers are used to specifically select, visualize, locate a specific instance of a ligand, amino acid in a protein chain in a particular PDB entry.

Here we are using PDB ID 1NDZ.

1NDZ - Crystal Structure of Adenosine Deaminase Complexed with FR235999

PyMOL:

PyMOL is an open-source molecular visualization system created by Warren Lyford DeLano. PyMOL can produce high-quality 3D images of small molecules and biological macromolecules, such as proteins. Almost a quarter of all published images of 3D protein structures in the scientific literature were made using PyMOL.

PyMOL is one of the few open-source model visualization tools available for use in structural biology. The Py part of the software's name refers to the program having been written in the programming language Python.

The objects that PyMOL renders in 3D are loaded from coordinate files that describe (in great detail) locations of individual atoms in the molecule. PyMOL can display more than one object at a time, and provides an Object Control Panel to adjust viewing modes, colors, labels, hiding, and just about anything else relating to objects. After each object name is a set of command buttons which control the object. Here are the buttons and some of their options:

- A-Actions: Rename, duplicate, remove, apply presets (like "ball-and-stick" or "publication"), perform computations.
- S-Show: Change the way things appear, eg. change to stick or cartoon view. H-Hide: Things that are shown using S accumulate, and don't automatically replace the last view. H is the opposite of S and hides unwanted representations.
- L-Label: Label atoms, residues, etc.
- C-Colour: Change the color of atoms and groups.

The lower-right corner of the Viewer contains a guide to using the mouse, as well as a powerful selection tool. There is also another command line at the bottom of the Viewer.

Proteins and nucleic acids are composed of a collection of atoms bonded together in specific arrangements. Since they are far too small to see with light, we need to devise artificial ways of representing them.

Three major types of pictures are created:

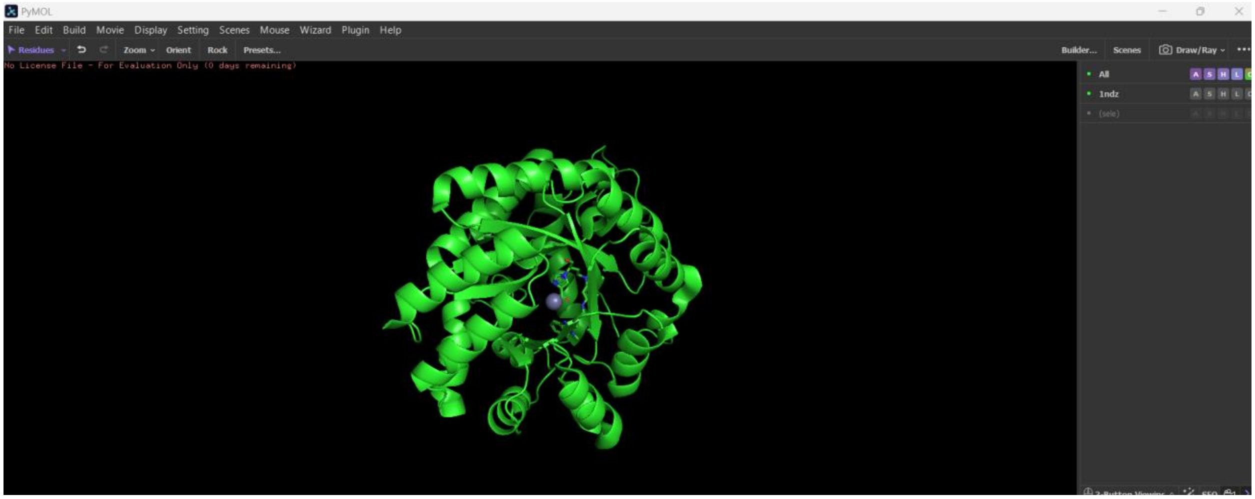
1. Wireframe Diagrams: For these images, a line is drawn for each of the covalent bonds formed between the atoms. In many cases, small balls and sticks are used to make the three-dimensional shape easier to understand.
2. Space filling Diagrams: For these images, a sphere is drawn around each atom, showing the relative size of the atom.
3. Backbone and Ribbon Diagrams: These images highlight the way a protein chain folds. The simplest ones draw a tube that connects the positions of each amino acid. Ribbon diagrams add two special representations: a spring-shaped ribbon for alpha helices and flat arrow that shows beta strands. The familiar ladder diagram is used for nucleic acids, with a smooth ribbon for the backbone and rungs for the bases.

METHOD:

1. Firstly, a protein of interest has to be downloaded from the PDB site. It can be downloaded in several formats like PDB format, fasta format etc.
2. Secondly this protein has to be extracted in the working space of PyMOL by fetch followed by PDB ID or by directly opening the file.
3. The right side of the Viewer shows the loaded PDB as an object, as well as its command buttons. Each button contains a submenu with more options.
4. Firstly, remove the water molecules. Click A, then select remove water.
5. To change the color of each protein chain, click C then select any color. To color each chain differently i.e., a single color to each chain. Click C then select by chain from the by chain menu.
6. Click S, then lines to show the protein's secondary structure in lines form. Currently the cartoon's view is still visible on top-of-the-line view. To hide the cartoon, click H then cartoon.
7. To view the amino acid sequence, click S in the bottom right corner. To remove the chain or residues selecting the sequence then right click and select remove.
8. Now to see the ligand sites, click on A then preset and select ligand sites and then cartoon to see the bonds.
9. Then to label the residues bonded to the ligand. Select the residue by left click on it and then right click and click on label and then select residues.
10. Next step is to measure the distance between the residues and the ligand. Click wizard from top

menu bar and select measurements. Then click on first atom at the end of the bond and then click on the second atom at the other end of the bond.

Results:



```
No License File - For Evaluation Only (0 days remaining)
```

Figure: Protein loaded in Pymol

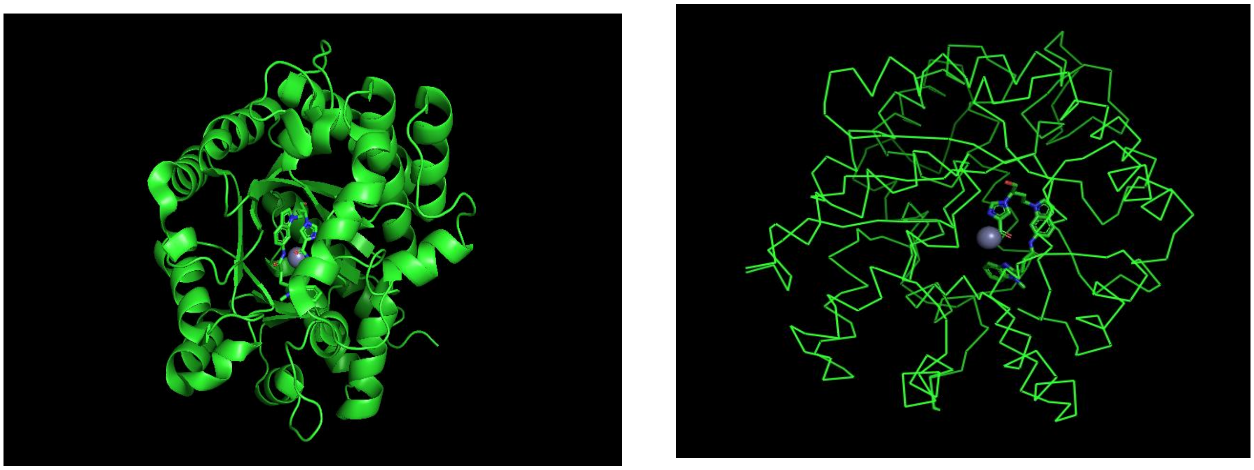


Figure: 3D Structure of protein was shown as Cartoon and ribbon

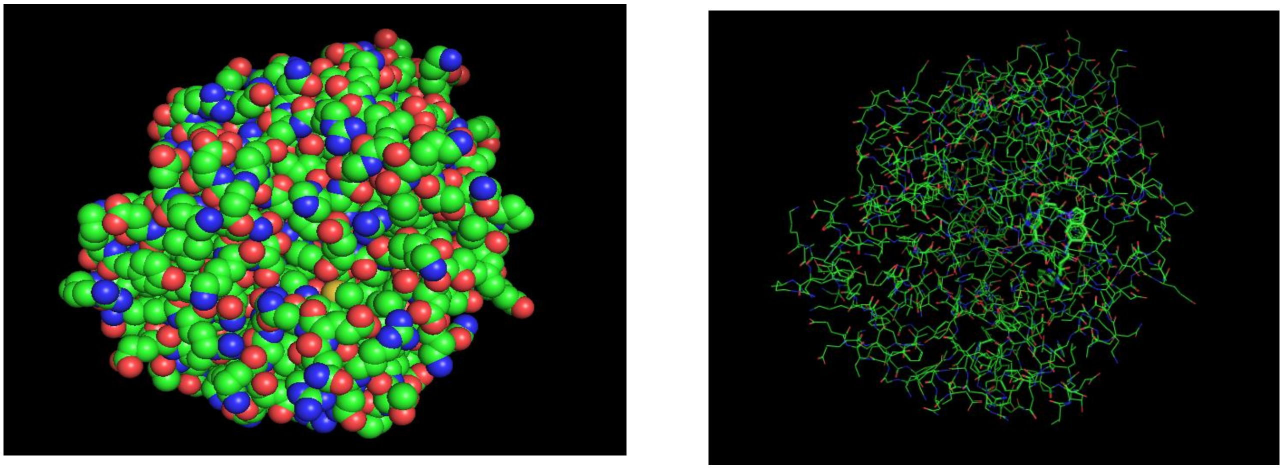


Figure: 3D Structure of protein was shown as Sphere and wire

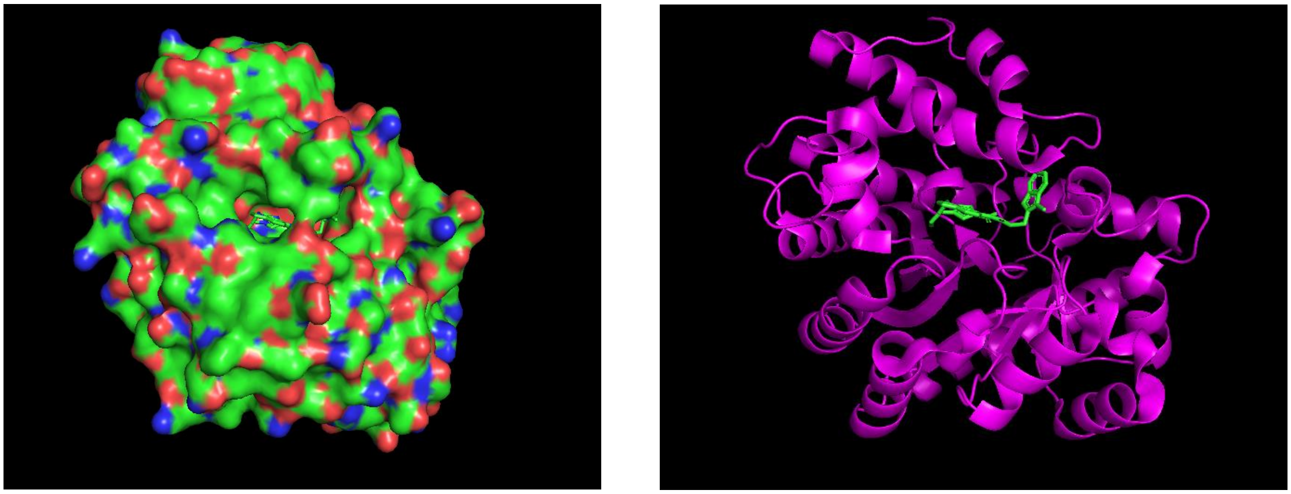


Figure: 1NDZ surface view and water deleted

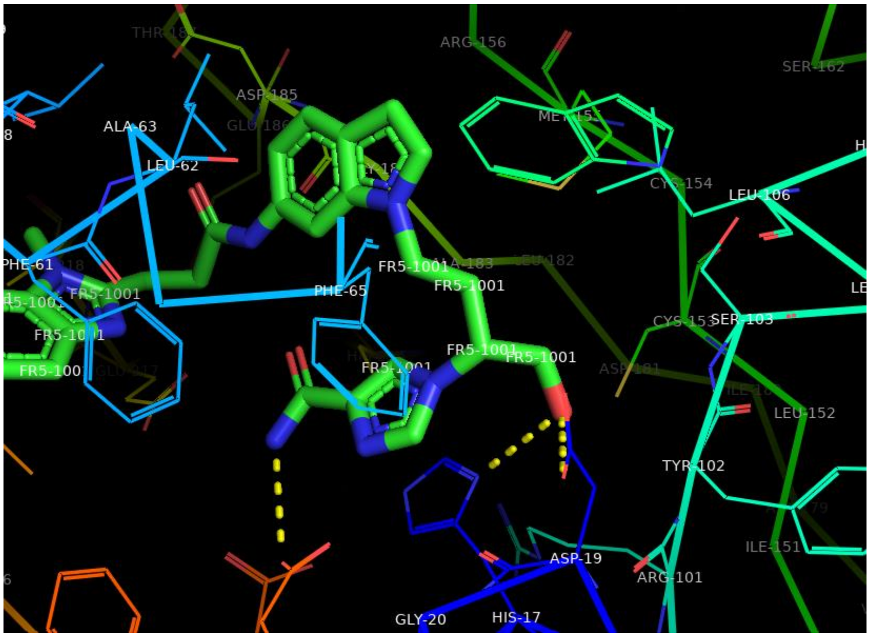


Figure: 1NDZ-A measurement

Conclusion: The secondary structure information and the molecular surface information of the protein with the PDB ID 1NDZ.A was viewed using PyMOL. It shows three hydrogen bonds between the ligand showing interaction with the 184 GLY, 269 THR and 296 ASP residues.

EX.NO :2	SKETCHING OF MOLECULES USING MARVIN SKETCH
Date : 07/08/24	

AIM:

To analyse small molecules and draw the molecules using marvin sketch.

INTRODUCTION:

Marvin Sketch is an advanced chemical editor for drawing chemical structures, queries and reactions. It has a rich (and growing) list of editing features, is chemically aware and is able to call ChemAxon's structure based calculation plugins for structures on the canvas. It supports a wide range of file types such as SDF, RDF (V2000/V3000), RXN, MOL, MOL2, SMILES,SMARTS, InChi, FASTA, etc. The structures can be cleaned in 2D and 3D geometry.

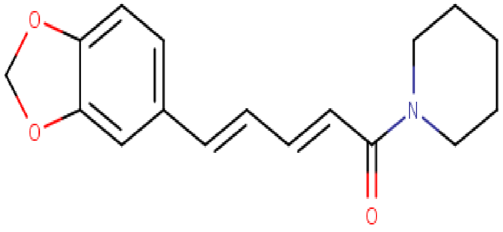
TOOLS: Marvin sketch

METHODOLOGY:

i) The structure of small molecules were drawn using Marvin Sketch tool.
ii) The elementary analysis ,polar surface area conformation were done and the molecule saved in mol, SDF and smiles file format.

RESULTS:

1) Name : Piperine
2) Biological significance: Bioperine has been used in trials studying the treatment of Multiple Myeloma and Deglutition Disorders.
3) SMILES : O=C(\C=C\C=C\C1=CC2=C(OCO2)C=C1)N1CCCCC1



Molecular weight: 285.343
Exact molecular weight: 285.136493476
Formula: $C_{17}H_{19}NO_3$
Dot-disconnected formula: $C_{17}H_{19}NO_3$
Composition: C (71.56%), H (6.71%), N (4.91%), O (16.82%)
Atom count: 40
Mass spectrum [m/z: relative abundance]:
285: 1.00 286: 0.19 287: 0.02

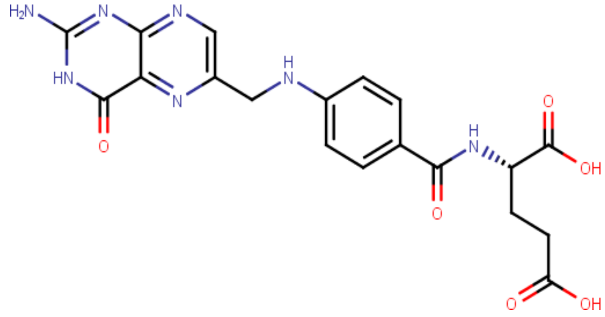
Figure 1 : Structure of Piperine

Figure 1.1 : Elemental analysis of Piperine

1) Name : Folic acid

2) Biological significance: Folic acid is required by the body for the synthesis of purines, pyrimidines, and methionine before incorporation into DNA or protein. Folic acid is particularly important during phases of rapid cell division, such as infancy, pregnancy, and erythropoiesis, and plays a protective factor in the development of cancer.

3) SMILES: NC1=NC(=O)C2=NC(CNC3=CC=C(C=C3)C(=O)N[C@@H](CCC(O)=O)C(O)=O)=CN=C2N1



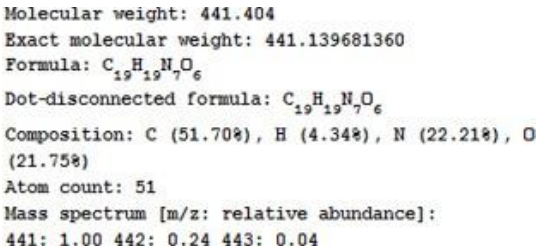

Molecular weight: 441.404
Exact molecular weight: 441.139681360
Formula: $C_{19}H_{19}N_7O_6$
Dot-disconnected formula: $C_{19}H_{19}N_7O_6$
Composition: C (51.70%), H (4.34%), N (22.21%), O (21.75%)
Atom count: 51
Mass spectrum [m/z: relative abundance]:
441: 1.00 442: 0.24 443: 0.04

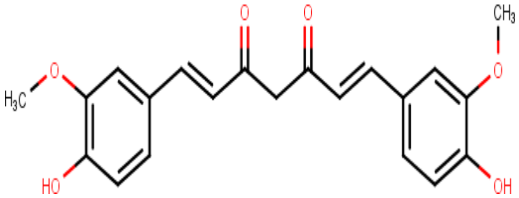
Figure 2 : Folic acid structure

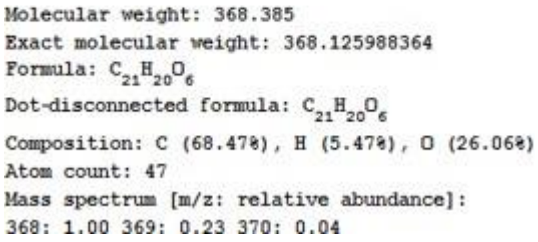
Figure 2.1 : Elemental analysis

1) Name : Curcumin

2) Biological significance: It is a highly pleiotropic molecule that exhibits antibacterial, anti-inflammatory, hypoglycemic, antioxidant, wound-healing, and antimicrobial activities. Due to these properties, curcumin has been investigated for the treatment and supportive care of clinical conditions including proteinuria, breast cancer, multiple myeloma, depression, and Non Small Cell Lung Cancer (NSCLC).

3) SMILES : COC1=CC(\C=C\C(=O)CC(=O)\C=C\C2=CC(OC)=C(O)C=C2)=CC=C1O





Molecular weight: 368.385

Exact molecular weight: 368.125988364

Formula: $C_{21}H_{20}O_6$

Dot-disconnected formula: $C_{21}H_{20}O_6$

Composition: C (68.47%), H (5.47%), O (26.06%)

Atom count: 47

Mass spectrum [m/z: relative abundance]:

368: 1.00 369: 0.23 370: 0.04

Figure 3 : Curcumin structure

Figure 3.1 : Elemental analysis

INFERENCE:

The small molecules were drawn using Marvin Sketch. Elementary analysis, polar surface area, conformers were done using this tool.SDF is a more general-purpose format suitable for representing a variety of chemical data, including multiple structures in a single file. MOL2, on the other hand, is specialized for 3D molecular structures and is often used in the context of molecular modeling and simulations.

Date : 24/07/24

AIM :

To view and understand the file formats.

INTRODUCTION:

File Format: A standard way that data is encoded for storage in a computer file which determines how the data are used to encode information in a digital storage medium.

Types of file format: There is a number of file formats for representation and communication of chemical information. A wide range of file types such as SDF, RDF (V2000/V3000), RXN, MOL, MOL2, SMILES, SMARTS, InChi, FASTA, etc.

PROCEDURE :

i) The structure of the molecule was drawn using the tools available in the toolbar of Marvin Sketch.

ii) It was saved in mol format.

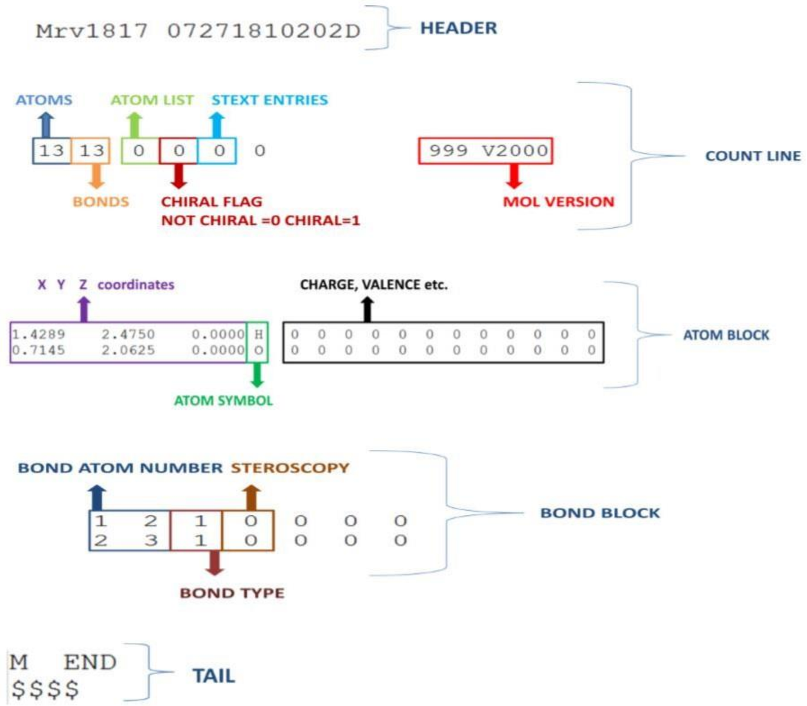
iii) Using Marvin Sketch the saved mol file was converted to .mol2, .pdb and .smile formats.

SDF/MOL file format :

- An MDL Molfile is a file format for holding information about the atoms, bonds, connectivity, and coordinates of a molecule.
- An MDL SDF (Structure –data file) includes one or more molfiles. Another feature is the ability to include data associated with the molecules.

Examples of SDF/MOL file format :

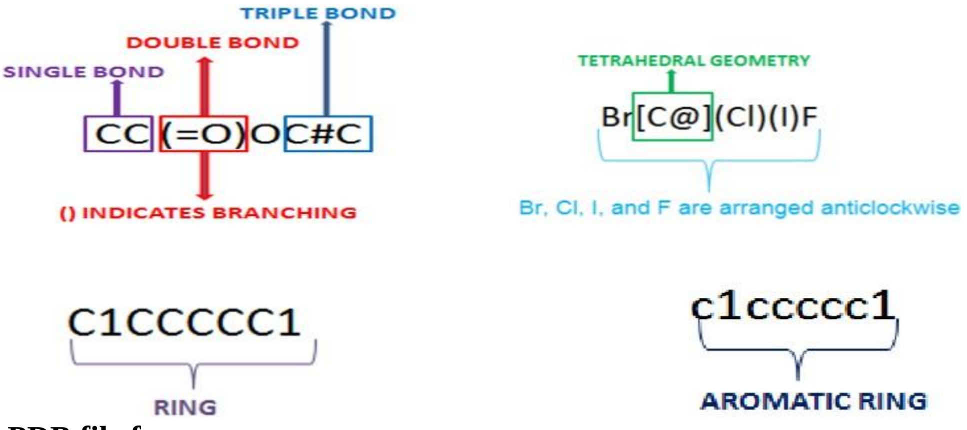
PARACETAMOL



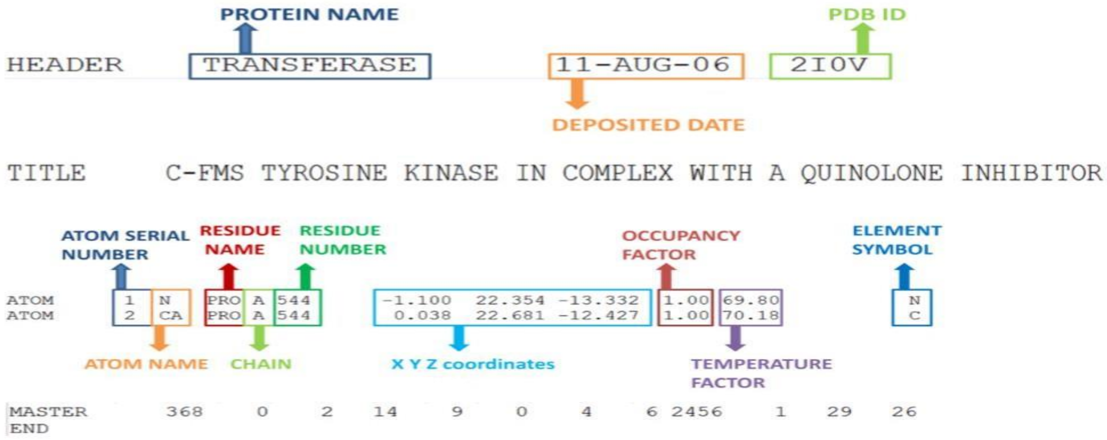
```
Mrv1817 07271810202D

 13 13  0  0  0  0            999 V2000

1.4289    2.4750    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
0.7145    2.0625    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0

  1  2  1  0  0  0  0
  2  3  1  0  0  0  0

M  END
$$$$
```

Example of SMILES file format:

SMILES -Simplified Molecular-Input Line-Entry System



Example of PDB file format:



```
HEADER     TRANSFERASE          11-AUG-06   2I0V

TITLE     C-FMS TYROSINE KINASE IN COMPLEX WITH A QUINOLONE INHIBITOR

ATOM       1  N   PRO A 544    -1.100  22.354 -13.332  1.00 69.80           N
ATOM       2  CA  PRO A 544     0.038  22.681 -12.427  1.00 70.18           C

MASTER      368    0    2   14    9    0    4    6 2456    1   29   26
END
```

INFERENCE :

The various file formats (.mol, .pdb, .smiles) were studied and the comparative analysis has been performed.

EX.NO : 3	Geometry Optimiztion using SWISSPdb Viewer
Date : 28/08/24	

AIM:
To complete the protein structure and minimize the total internal energy of the given protein using Swiss PDB Viewer.

TOOLS USED:

- WHAT IF Web Interface
- Swiss-PDB Viewer v4.10

INTRODUCTION:

WHAT IF Web Interface
The program WHAT IF provides nearly 2000 options in fields as diverse as homology modeling, drug docking, electrostatics calculations, structure validation, and visualization. It has many programs for users to perform freely having internet access. In our experiment, we use WHAT IF Web Interface is used to complete and optimize the structure of the protein molecule, Where it helps to model missing side chains in the protein molecule.

STEPS:
i) Open **WHAT IF** Web Interface online Server.
Classes → Build/check/repair model →Complete a structure
ii) Upload protein .pdb file and run the program.
iii) Download the fixed.pdb as a result file (complete structure).

Swiss-PdbViewer v4.10
Swiss-PdbViewer is an application that provides a user-friendly interface allowing to analyze several proteins at the same time. The proteins can be superimposed in order to deduce structural alignments and compare their active sites or any other relevant parts. Amino acid mutations, H-bonds, angles and distances between atoms are easy to obtain thanks to the intuitive graphic and menu interface.
Swiss-PdbViewer can also read electron density maps, and provides various tools to build into the density. In addition, various modeling tools are integrated and residues can be mutated. Finally, as a special bonus, POV-Ray scenes can be generated from the current view in order to make stunning ray-traced quality images.

PROCEDURE:
i) Swiss-PdbViewer - tool for optimizing and energy minimizing of the protein molecule by Steepest Descent and Conjugate Gradient Methods.
ii) Open Swiss-Pdb-Viewer tool for Energy Minimization.

iii) Load the output file of What if -fixed.pdb in Swiss-Pdb Viewer tool.

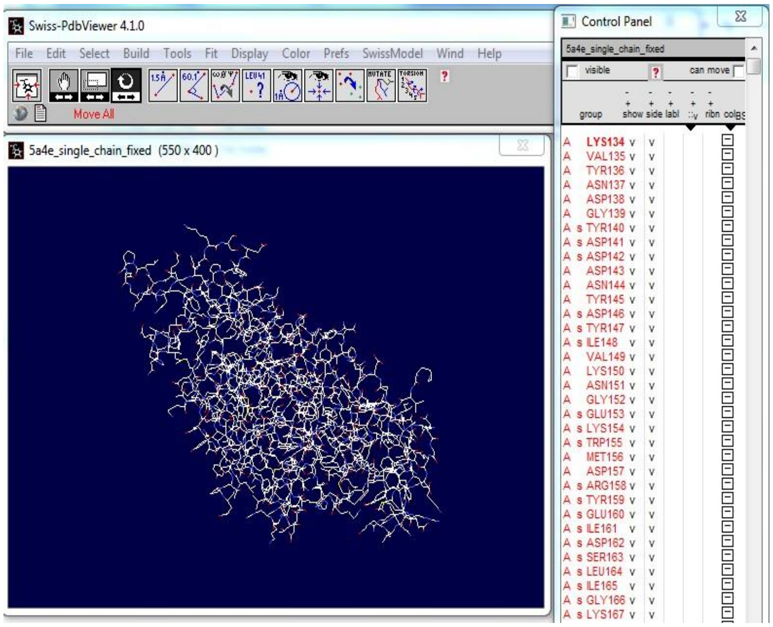
iv) Click on the wind --->Control panel and select all the amino acids.

v) Click on Prefs →Energy Minimization select the number of cycles for Steepest Descent method and Conjugate Gradient method.

vi) Run the Energy Minimization. Tools → Energy minimization or Ctrl+N and save the optimized protein molecule as minimized.pdb for further studies.

RESULTS :

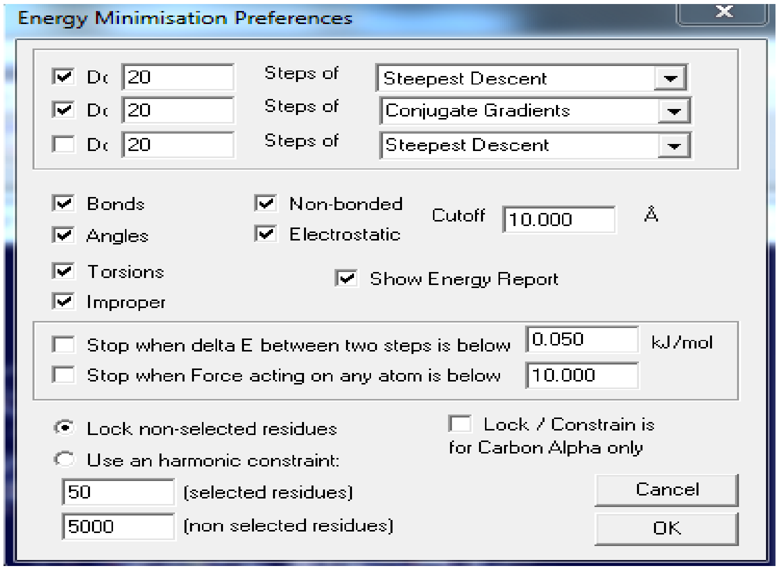
1) 5A4E



5a4e_single_chain_fixed

□ visible ? can move □

group	show	side	labl	::v	ribn	col
A LYS134	v	v				⊟
A VAL135	v	v				⊟
A TYR136	v	v				⊟
A ASN137	v	v				⊟
A ASP138	v	v				⊟
A GLY139	v	v				⊟
A s TYR140	v	v				⊟
A s ASP141	v	v				⊟
A s ASP142	v	v				⊟
A ASP143	v	v				⊟
A ASN144	v	v				⊟
A TYR145	v	v				⊟
A s ASP146	v	v				⊟
A s TYR147	v	v				⊟
A s ILE148	v	v				⊟
A VAL149	v	v				⊟
A LYS150	v	v				⊟
A ASN151	v	v				⊟
A GLY152	v	v				⊟
A s GLU153	v	v				⊟
A s LYS154	v	v				⊟
A s TRP155	v	v				⊟
A MET156	v	v				⊟
A ASP157	v	v				⊟
A s ARG158	v	v				⊟
A s TYR159	v	v				⊟
A s GLU160	v	v				⊟
A s ILE161	v	v				⊟
A s ASP162	v	v				⊟
A s SER163	v	v				⊟
A s LEU164	v	v				⊟
A s ILE165	v	v				⊟
A s GLY166	v	v				⊟
A s LYS167	v	v				⊟



Energy Minimisation Preferences

☑ Do 20 Steps of Steepest Descent

☑ Do 20 Steps of Conjugate Gradients

□ Do 20 Steps of Steepest Descent

☑ Bonds ☑ Non-bonded Cutoff 10.000 Å

☑ Angles ☑ Electrostatic

☑ Torsions ☑ Show Energy Report

☑ Improper

□ Stop when delta E between two steps is below 0.050 kJ/mol

□ Stop when Force acting on any atom is below 10.000

● Lock non-selected residues

○ Use an harmonic constraint:

50 (selected residues)

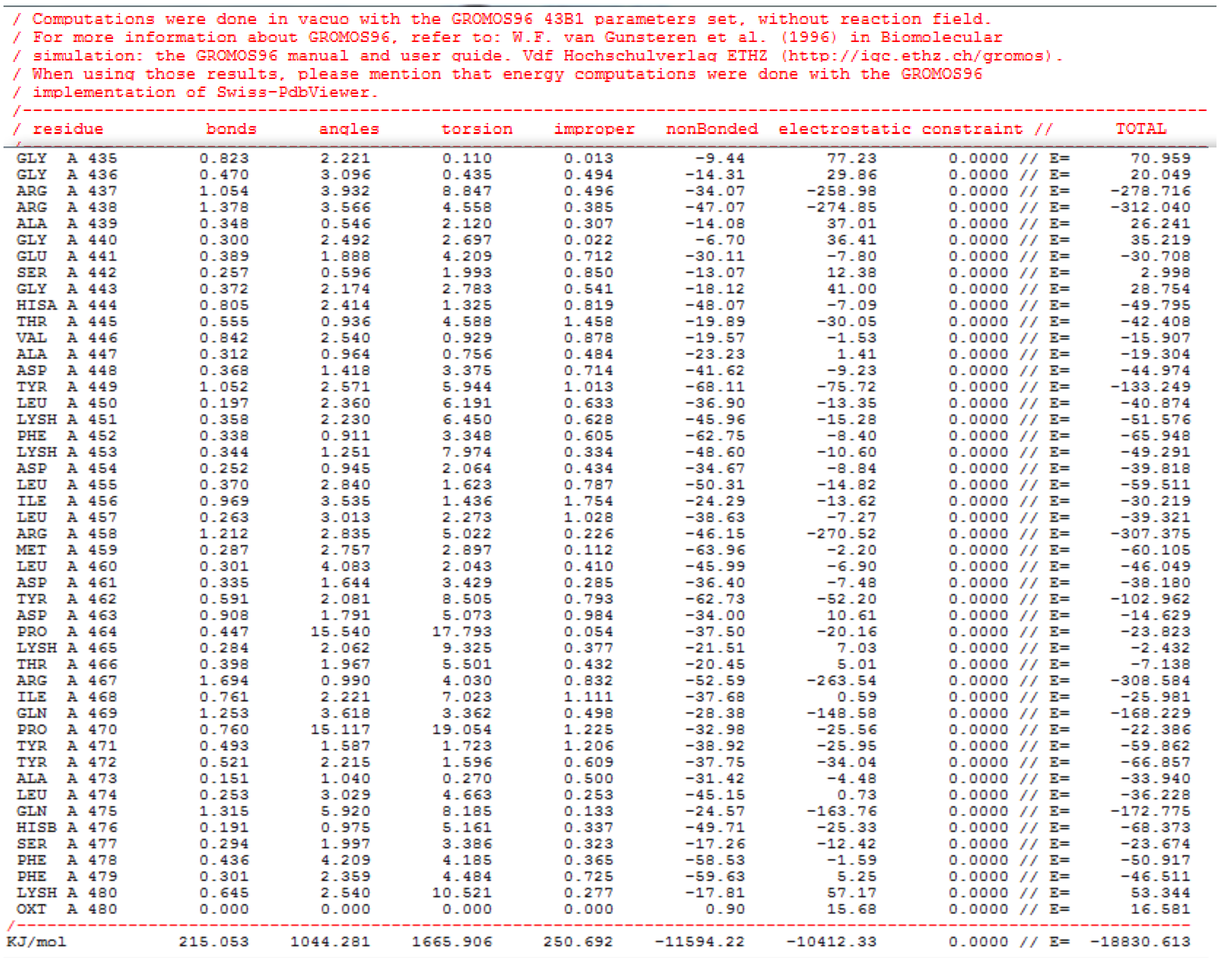
5000 (non selected residues)

□ Lock / Constrain is for Carbon Alpha only

Cancel

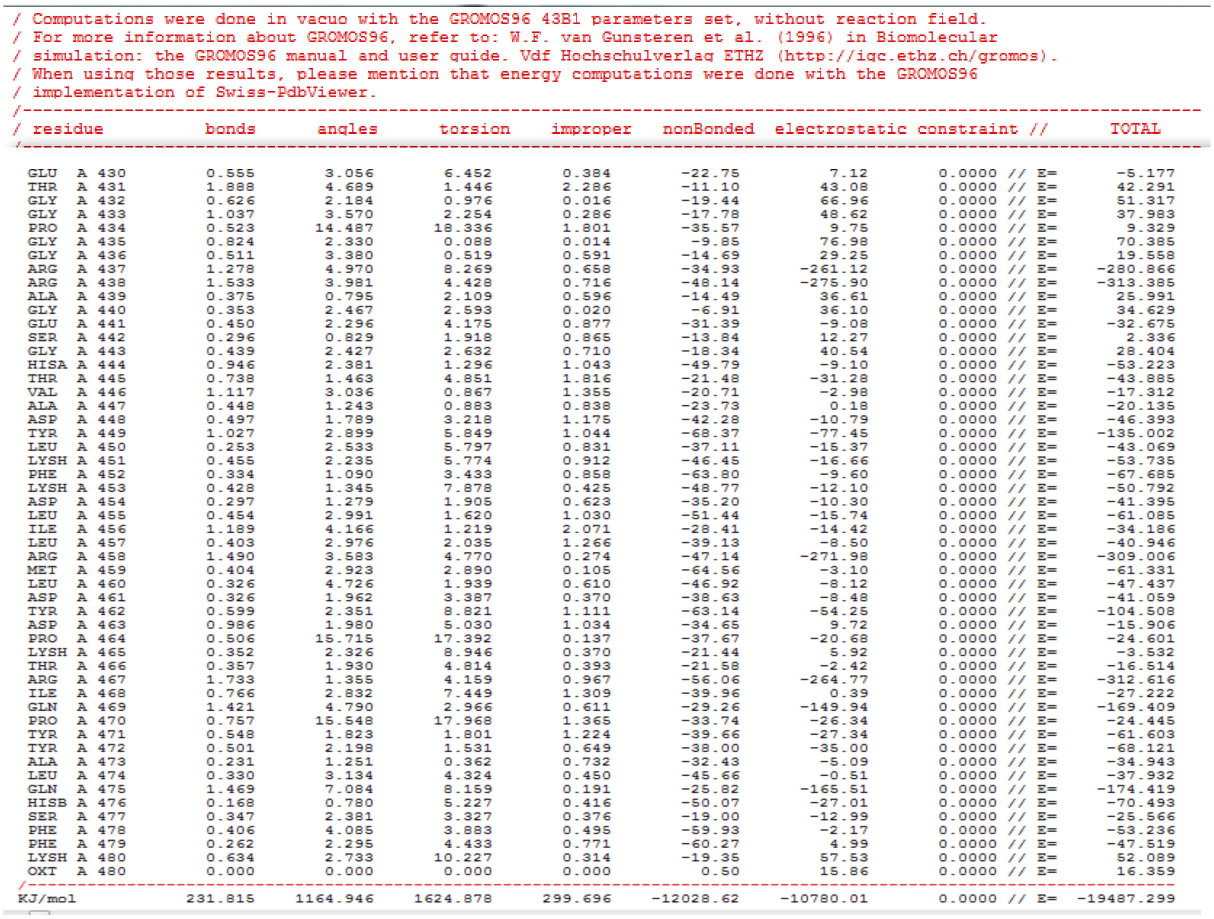
OK

Steepest at 20



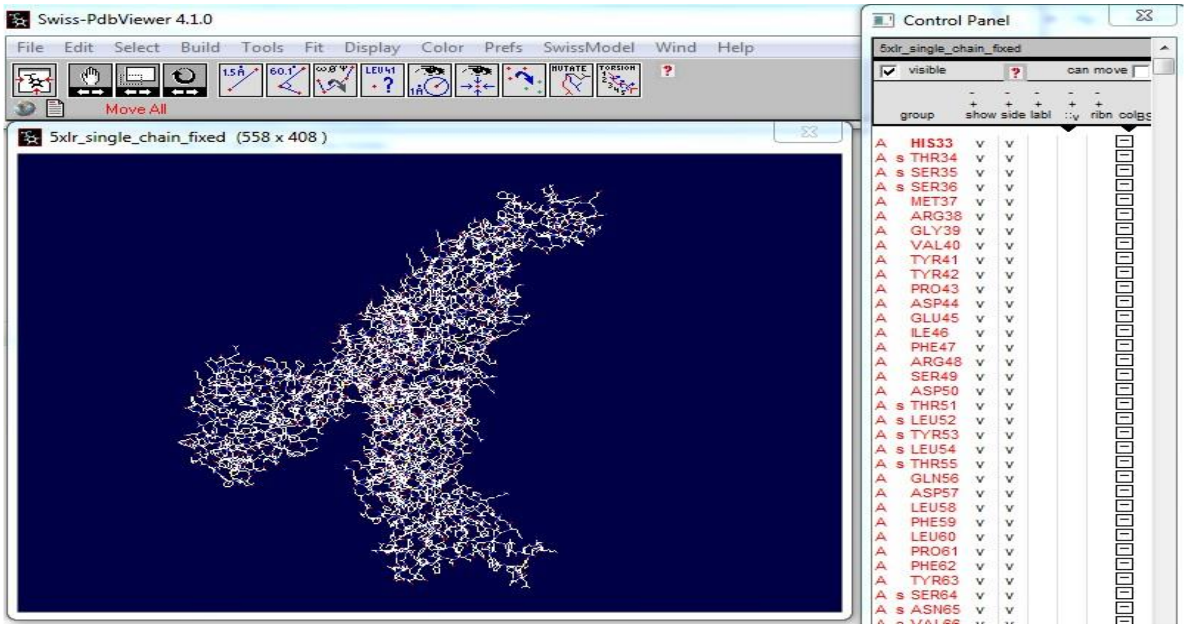
```
/ Computations were done in vacuo with the GROMOS96 43B1 parameters set, without reaction field.
/ For more information about GROMOS96, refer to: W.F. van Gunsteren et al. (1996) in Biomolecular
/ simulation: the GROMOS96 manual and user guide. Vdf Hochschulverlag ETHZ (http://igc.ethz.ch/gromos).
/ When using those results, please mention that energy computations were done with the GROMOS96
/ implementation of Swiss-PdbViewer.
/-----------------------------------------------------------------------------------------------------
/ residue            bonds      angles     torsion    improper   nonBonded  electrostatic constraint //      TOTAL
/-----------------------------------------------------------------------------------------------------
 GLY  A 435          0.823      2.221      0.110      0.013      -9.44      77.23      0.0000 // E=    70.959
 GLY  A 436          0.470      3.096      0.435      0.494     -14.31      29.86      0.0000 // E=    20.049
 ARG  A 437          1.054      3.932      8.847      0.496     -34.07    -258.98      0.0000 // E=  -278.716
 ARG  A 438          1.378      3.566      4.558      0.385     -47.07    -274.85      0.0000 // E=  -312.040
 ALA  A 439          0.348      0.546      2.120      0.307     -14.08      37.01      0.0000 // E=    26.241
 GLY  A 440          0.300      2.492      2.697      0.022      -6.70      36.41      0.0000 // E=    35.219
 GLU  A 441          0.389      1.888      4.209      0.712     -30.11      -7.80      0.0000 // E=   -30.708
 SER  A 442          0.257      0.596      1.993      0.850     -13.07      12.38      0.0000 // E=     2.998
 GLY  A 443          0.372      2.174      2.783      0.541     -18.12      41.00      0.0000 // E=    28.754
 HISA A 444          0.805      2.414      1.325      0.819     -48.07      -7.09      0.0000 // E=   -49.795
 THR  A 445          0.555      0.936      4.588      1.458     -19.89     -30.05      0.0000 // E=   -42.408
 VAL  A 446          0.842      2.540      0.929      0.878     -19.57      -1.53      0.0000 // E=   -15.907
 ALA  A 447          0.312      0.964      0.756      0.484     -23.23       1.41      0.0000 // E=   -19.304
 ASP  A 448          0.368      1.418      3.375      0.714     -41.62      -9.23      0.0000 // E=   -44.974
 TYR  A 449          1.052      2.571      5.944      1.013     -68.11     -75.72      0.0000 // E=  -133.249
 LEU  A 450          0.197      2.360      6.191      0.633     -36.90     -13.35      0.0000 // E=   -40.874
 LYSH A 451          0.358      2.230      6.450      0.628     -45.96     -15.28      0.0000 // E=   -51.576
 PHE  A 452          0.338      0.911      3.348      0.605     -62.75      -8.40      0.0000 // E=   -65.948
 LYSH A 453          0.344      1.251      7.974      0.334     -48.60     -10.60      0.0000 // E=   -49.291
 ASP  A 454          0.252      0.945      2.064      0.434     -34.67      -8.84      0.0000 // E=   -39.818
 LEU  A 455          0.370      2.840      1.623      0.787     -50.31     -14.82      0.0000 // E=   -59.511
 ILE  A 456          0.969      3.535      1.436      1.754     -24.29     -13.62      0.0000 // E=   -30.219
 LEU  A 457          0.263      3.013      2.273      1.028     -38.63      -7.27      0.0000 // E=   -39.321
 ARG  A 458          1.212      2.835      5.022      0.226     -46.15    -270.52      0.0000 // E=  -307.375
 MET  A 459          0.287      2.757      2.897      0.112     -63.96      -2.20      0.0000 // E=   -60.105
 LEU  A 460          0.301      4.083      2.043      0.410     -45.99      -6.90      0.0000 // E=   -46.049
 ASP  A 461          0.335      1.644      3.429      0.285     -36.40      -7.48      0.0000 // E=   -38.180
 TYR  A 462          0.591      2.081      8.505      0.793     -62.73     -52.20      0.0000 // E=  -102.962
 ASP  A 463          0.908      1.791      5.073      0.984     -34.00      10.61      0.0000 // E=   -14.629
 PRO  A 464          0.447     15.540     17.793      0.054     -37.50     -20.16      0.0000 // E=   -23.823
 LYSH A 465          0.284      2.062      9.325      0.377     -21.51       7.03      0.0000 // E=    -2.432
 THR  A 466          0.398      1.967      5.501      0.432     -20.45       5.01      0.0000 // E=    -7.138
 ARG  A 467          1.694      0.990      4.030      0.832     -52.59    -263.54      0.0000 // E=  -308.584
 ILE  A 468          0.761      2.221      7.023      1.111     -37.68       0.59      0.0000 // E=   -25.981
 GLN  A 469          1.253      3.618      3.362      0.498     -28.38    -148.58      0.0000 // E=  -168.229
 PRO  A 470          0.760     15.117     19.054      1.225     -32.98     -25.56      0.0000 // E=   -22.386
 TYR  A 471          0.493      1.587      1.723      1.206     -38.92     -25.95      0.0000 // E=   -59.862
 TYR  A 472          0.521      2.215      1.596      0.609     -37.75     -34.04      0.0000 // E=   -66.857
 ALA  A 473          0.151      1.040      0.270      0.500     -31.42      -4.48      0.0000 // E=   -33.940
 LEU  A 474          0.253      3.029      4.663      0.253     -45.15       0.73      0.0000 // E=   -36.228
 GLN  A 475          1.315      5.920      8.185      0.133     -24.57    -163.76      0.0000 // E=  -172.775
 HISB A 476          0.191      0.975      5.161      0.337     -49.71     -25.33      0.0000 // E=   -68.373
 SER  A 477          0.294      1.997      3.386      0.323     -17.26     -12.42      0.0000 // E=   -23.674
 PHE  A 478          0.436      4.209      4.185      0.365     -58.53      -1.59      0.0000 // E=   -50.917
 PHE  A 479          0.301      2.359      4.484      0.725     -59.63       5.25      0.0000 // E=   -46.511
 LYSH A 480          0.645      2.540     10.521      0.277     -17.81      57.17      0.0000 // E=    53.344
 OXT  A 480          0.000      0.000      0.000      0.000       0.90      15.68      0.0000 // E=    16.581
/-----------------------------------------------------------------------------------------------------
KJ/mol             215.053   1044.281   1665.906    250.692   -11594.22  -10412.33      0.0000 // E= -18830.613
```

Conjugate at 20

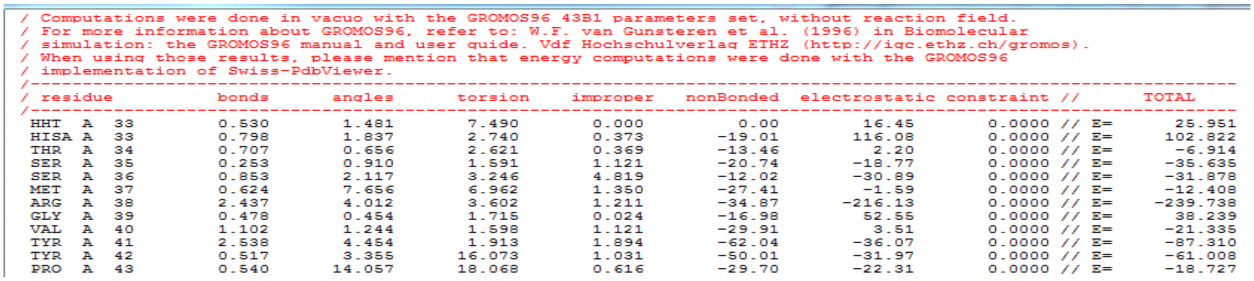


```
/ Computations were done in vacuo with the GROMOS96 43B1 parameters set, without reaction field.
/ For more information about GROMOS96, refer to: W.F. van Gunsteren et al. (1996) in Biomolecular
/ simulation: the GROMOS96 manual and user guide. Vdf Hochschulverlag ETHZ (http://igc.ethz.ch/gromos).
/ When using those results, please mention that energy computations were done with the GROMOS96
/ implementation of Swiss-PdbViewer.
/-------------------------------------------------------------------------------------------------------------
/ residue          bonds       angles      torsion     improper    nonBonded  electrostatic constraint //      TOTAL
/-------------------------------------------------------------------------------------------------------------

  GLU  A 430       0.555       3.056       6.452       0.384      -22.75        7.12     0.0000 // E=     -5.177
  THR  A 431       1.888       4.689       1.446       2.286      -11.10       43.08     0.0000 // E=     42.291
  GLY  A 432       0.626       2.184       0.976       0.016      -19.44       66.96     0.0000 // E=     51.317
  GLY  A 433       1.037       3.570       2.254       0.286      -17.78       48.62     0.0000 // E=     37.983
  PRO  A 434       0.523      14.487      18.336       1.801      -35.57        9.75     0.0000 // E=      9.329
  GLY  A 435       0.824       2.330       0.088       0.014       -9.85       76.98     0.0000 // E=     70.385
  GLY  A 436       0.511       3.380       0.519       0.591      -14.69       29.25     0.0000 // E=     19.558
  ARG  A 437       1.278       4.970       8.269       0.658      -34.93     -261.12     0.0000 // E=   -280.866
  ARG  A 438       1.533       3.981       4.428       0.716      -48.14     -275.90     0.0000 // E=   -313.385
  ALA  A 439       0.375       0.795       2.109       0.596      -14.49       36.61     0.0000 // E=     25.991
  GLY  A 440       0.353       2.467       2.593       0.020       -6.91       36.10     0.0000 // E=     34.629
  GLU  A 441       0.450       2.296       4.175       0.877      -31.39       -9.08     0.0000 // E=    -32.675
  SER  A 442       0.296       0.829       1.918       0.865      -13.84       12.27     0.0000 // E=      2.336
  GLY  A 443       0.439       2.427       2.632       0.710      -18.34       40.54     0.0000 // E=     28.404
  HISA A 444       0.946       2.381       1.296       1.043      -49.79       -9.10     0.0000 // E=    -53.223
  THR  A 445       0.738       1.463       4.851       1.816      -21.48      -31.28     0.0000 // E=    -43.885
  VAL  A 446       1.117       3.036       0.867       1.355      -20.71       -2.98     0.0000 // E=    -17.312
  ALA  A 447       0.448       1.243       0.883       0.838      -23.73        0.18     0.0000 // E=    -20.135
  ASP  A 448       0.497       1.789       3.218       1.175      -42.28      -10.79     0.0000 // E=    -46.393
  TYR  A 449       1.027       2.899       5.849       1.044      -68.37      -77.45     0.0000 // E=   -135.002
  LEU  A 450       0.253       2.533       5.797       0.831      -37.11      -15.37     0.0000 // E=    -43.069
  LYSH A 451       0.455       2.235       5.774       0.912      -46.45      -16.66     0.0000 // E=    -53.735
  PHE  A 452       0.334       1.090       3.433       0.858      -63.80       -9.60     0.0000 // E=    -67.685
  LYSH A 453       0.428       1.345       7.878       0.425      -48.77      -12.10     0.0000 // E=    -50.792
  ASP  A 454       0.297       1.279       1.905       0.623      -35.20      -10.30     0.0000 // E=    -41.395
  LEU  A 455       0.454       2.991       1.620       1.030      -51.44      -15.74     0.0000 // E=    -61.085
  ILE  A 456       1.189       4.166       1.219       2.071      -28.41      -14.42     0.0000 // E=    -34.186
  LEU  A 457       0.403       2.976       2.035       1.266      -39.13       -8.50     0.0000 // E=    -40.946
  ARG  A 458       1.490       3.583       4.770       0.274      -47.14     -271.98     0.0000 // E=   -309.006
  MET  A 459       0.404       2.923       2.890       0.105      -64.56       -3.10     0.0000 // E=    -61.331
  LEU  A 460       0.326       4.726       1.939       0.610      -46.92       -8.12     0.0000 // E=    -47.437
  ASP  A 461       0.326       1.962       3.387       0.370      -38.63       -8.48     0.0000 // E=    -41.059
  TYR  A 462       0.599       2.351       8.821       1.111      -63.14      -54.25     0.0000 // E=   -104.508
  ASP  A 463       0.986       1.980       5.030       1.034      -34.65        9.72     0.0000 // E=    -15.906
  PRO  A 464       0.506      15.715      17.392       0.137      -37.67      -20.68     0.0000 // E=    -24.601
  LYSH A 465       0.352       2.326       8.946       0.370      -21.44        5.92     0.0000 // E=     -3.532
  THR  A 466       0.357       1.930       4.814       0.393      -21.58       -2.42     0.0000 // E=    -16.514
  ARG  A 467       1.733       1.355       4.159       0.967      -56.06     -264.77     0.0000 // E=   -312.616
  ILE  A 468       0.766       2.832       7.449       1.309      -39.96        0.39     0.0000 // E=    -27.222
  GLN  A 469       1.421       4.790       2.966       0.611      -29.26     -149.94     0.0000 // E=   -169.409
  PRO  A 470       0.757      15.548      17.968       1.365      -33.74      -26.34     0.0000 // E=    -24.445
  TYR  A 471       0.548       1.823       1.801       1.224      -39.66      -27.34     0.0000 // E=    -61.603
  TYR  A 472       0.501       2.198       1.531       0.649      -38.00      -35.00     0.0000 // E=    -68.121
  ALA  A 473       0.231       1.251       0.362       0.732      -32.43       -5.09     0.0000 // E=    -34.943
  LEU  A 474       0.330       3.134       4.324       0.450      -45.66       -0.51     0.0000 // E=    -37.932
  GLN  A 475       1.469       7.084       8.159       0.191      -25.82     -165.51     0.0000 // E=   -174.419
  HISB A 476       0.168       0.780       5.227       0.416      -50.07      -27.01     0.0000 // E=    -70.493
  SER  A 477       0.347       2.381       3.327       0.376      -19.00      -12.99     0.0000 // E=    -25.566
  PHE  A 478       0.406       4.085       3.883       0.495      -59.93       -2.17     0.0000 // E=    -53.236
  PHE  A 479       0.262       2.295       4.433       0.771      -60.27        4.99     0.0000 // E=    -47.519
  LYSH A 480       0.634       2.733      10.227       0.314      -19.35       57.53     0.0000 // E=     52.089
  OXT  A 480       0.000       0.000       0.000       0.000        0.50       15.86     0.0000 // E=     16.359
 /-------------------------------------------------------------------------------------------------------------
 KJ/mol          231.815    1164.946    1624.878     299.696    -12028.62   -10780.01     0.0000 // E= -19487.299
```

2) 5XLR

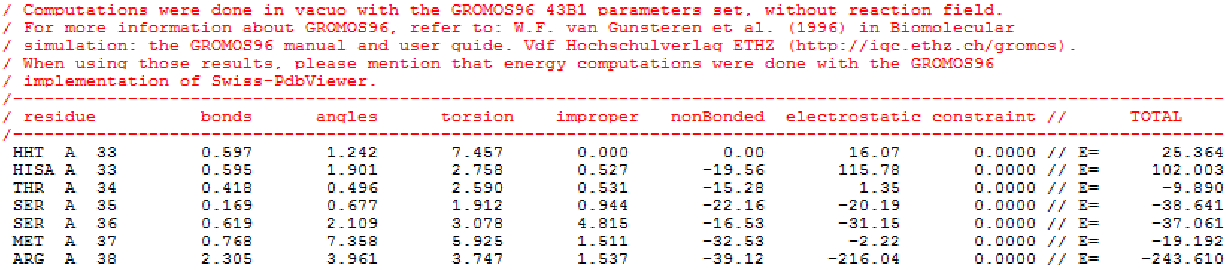


Steepest at 20



```
/ Computations were done in vacuo with the GROMOS96 43B1 parameters set, without reaction field.
/ For more information about GROMOS96, refer to: W.F. van Gunsteren et al. (1996) in Biomolecular
/ simulation: the GROMOS96 manual and user guide. Vdf Hochschulverlag ETHZ (http://igc.ethz.ch/gromos).
/ When using those results, please mention that energy computations were done with the GROMOS96
/ implementation of Swiss-PdbViewer.
/-----------------------------------------------------------------------------------------------------
/ residue          bonds      angles      torsion     improper   nonBonded  electrostatic constraint //      TOTAL
/-----------------------------------------------------------------------------------------------------
 HHT  A  33        0.530       1.481       7.490       0.000        0.00        16.45     0.0000 // E=      25.951
 HISA A  33        0.798       1.837       2.740       0.373      -19.01       116.08     0.0000 // E=     102.822
 THR  A  34        0.707       0.656       2.621       0.369      -13.46         2.20     0.0000 // E=      -6.914
 SER  A  35        0.253       0.910       1.591       1.121      -20.74       -18.77     0.0000 // E=     -35.635
 SER  A  36        0.853       2.117       3.246       4.819      -12.02       -30.89     0.0000 // E=     -31.878
 MET  A  37        0.624       7.656       6.962       1.350      -27.41        -1.59     0.0000 // E=     -12.408
 ARG  A  38        2.437       4.012       3.602       1.211      -34.87      -216.13     0.0000 // E=    -239.738
 GLY  A  39        0.478       0.454       1.715       0.024      -16.98        52.55     0.0000 // E=      38.239
 VAL  A  40        1.102       1.244       1.598       1.121      -29.91         3.51     0.0000 // E=     -21.335
 TYR  A  41        2.538       4.454       1.913       1.894      -62.04       -36.07     0.0000 // E=     -87.310
 TYR  A  42        0.517       3.355      16.073       1.031      -50.01       -31.97     0.0000 // E=     -61.008
 PRO  A  43        0.540      14.057      18.068       0.616      -29.70       -22.31     0.0000 // E=     -18.727
```

Conjugate at 20



```
/ Computations were done in vacuo with the GROMOS96 43B1 parameters set, without reaction field.
/ For more information about GROMOS96, refer to: W.F. van Gunsteren et al. (1996) in Biomolecular
/ simulation: the GROMOS96 manual and user guide. Vdf Hochschulverlag ETHZ (http://igc.ethz.ch/gromos).
/ When using those results, please mention that energy computations were done with the GROMOS96
/ implementation of Swiss-PdbViewer.
/-----------------------------------------------------------------------------------------------------
/ residue          bonds      angles      torsion     improper   nonBonded  electrostatic constraint //      TOTAL
/-----------------------------------------------------------------------------------------------------
 HHT  A  33        0.597       1.242       7.457       0.000        0.00        16.07     0.0000 // E=      25.364
 HISA A  33        0.595       1.901       2.758       0.527      -19.56       115.78     0.0000 // E=     102.003
 THR  A  34        0.418       0.496       2.590       0.531      -15.28         1.35     0.0000 // E=      -9.890
 SER  A  35        0.169       0.677       1.912       0.944      -22.16       -20.19     0.0000 // E=     -38.641
 SER  A  36        0.619       2.109       3.078       4.815      -16.53       -31.15     0.0000 // E=     -37.061
 MET  A  37        0.768       7.358       5.925       1.511      -32.53        -2.22     0.0000 // E=     -19.192
 ARG  A  38        2.305       3.961       3.747       1.537      -39.12      -216.04     0.0000 // E=    -243.610
```

INFERENCE:

Minimization of the protein has been done using Steepest descent method which reduced the energy to as low as -18830.613 KJ/Mol for 5A4E and -44141.633 KJ/Mol for 5XLR and again with Conjugate gradient method the overall energy was reduced to -19487.299 KJ/Mol for 5A4E and -46664.551 KJ/Mol for 5XLR ,which shows that algorithm was able to minimize the total energy of the protein.

EX.NO : 4	Homology modelling of Protein 3D structure
Date : 04/09/24	

AIM:
To perform homology modelling
TOOLS USED:
Modeller, Pymol, UCLA-DOE lab saves(https://saves.mbi.ucla.edu/)

INTRODUCTION:

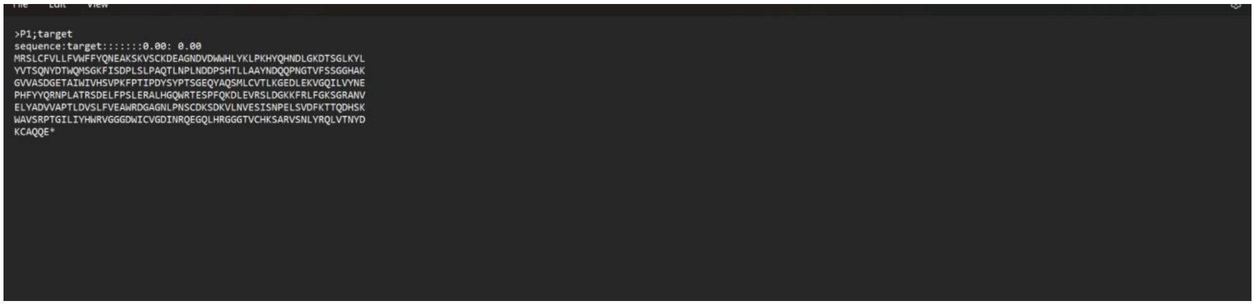
Homology modeling aims to build three-dimensional protein structure models using experimentally determined structures of related family members as templates. For a given target protein, a library of experimental protein structures is searched to identify suitable templates. On the basis of a sequence alignment between the target protein and the template structure, a three-dimensional model for the target protein is generated. Model quality assessment tools are used to estimate the reliability of the resulting models. Homology modeling is currently the most accurate computational method to generate reliable structural models and is routinely used in many biological applications Typically, the computational effort for a modeling project is less than 2 h. The homology modeling procedure can be broken down into four sequential steps: template selection, target-template alignment, model construction, and model assessment. The first two steps are often essentially performed together, as the most common methods of identifying templates rely on the production of sequence alignments; however, these alignments may not be of sufficient quality because database search techniques prioritize speed over alignment quality.
PyMOL:
PyMOL is an open-source molecular visualization system created by Warren Lyford DeLano. PyMOL can produce high-quality 3D images of small molecules and biological macromolecules, such as proteins. Almost a quarter of all published images of 3D protein structures in the scientific literature were made using PyMOL.
PyMOL is one of the few open-source model visualization tools available for use in structural biology. The Py part of the software's name refers to the program having been written in the programming language Python.
The objects that PyMOL renders in 3D are loaded from coordinate files that describe (in great detail) locations of individual atoms in the molecule. PyMOL can display more than one object at a time, and provides an Object Control Panel to adjust viewing modes, colors, labels, hiding, and just about anything else relating to objects.

Modeller:

It used for homology or comparative modeling of protein three-dimensional structures. From a sequence alignment with known related structures, MODELLER automatically calculates a model containing all non-hydrogen atoms using comparative protein structure modeling by satisfaction of spatial restraints. It can also perform de novo modeling of loops in protein structures and optimize various models of protein structure with respect to a flexibly defined objective function, multiple alignment of protein sequences and/or structures, clustering, searching of sequence databases, comparison of protein structures, etc.

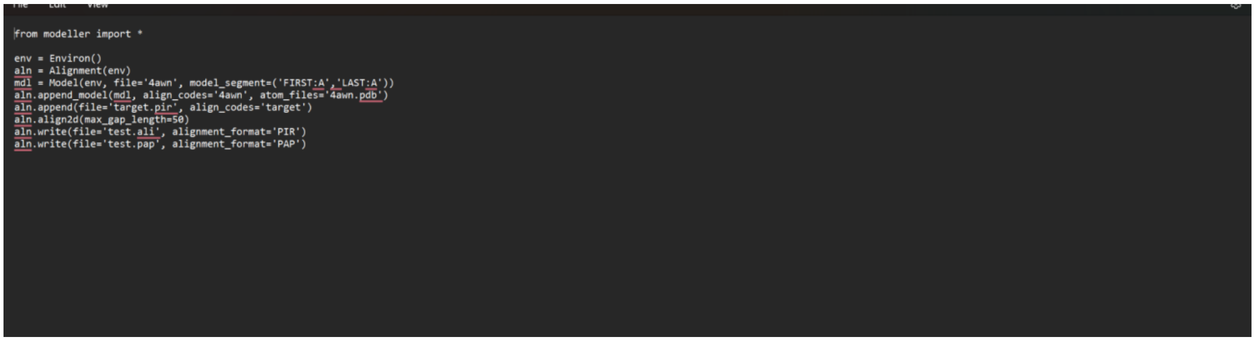
PROCEDURE:

1. Go to NCBI and search for the required protein.
2. Copy the sequence and BLAST the sequence to get similar sequences the sequence of protein kinase.
3. From blastp results, check query coverage, %identity, organism and for any missing residues. Select the top-hit result *4AWN*.
4. Open RCSB-PDB and download the PDB structure associated with ID *4AWN*. Convert the fasta format to PIR format for the target required protein and save the file in the folder as target.pir.
5. Convert the fasta format to PIR format for the target required protein and save the file in the folder



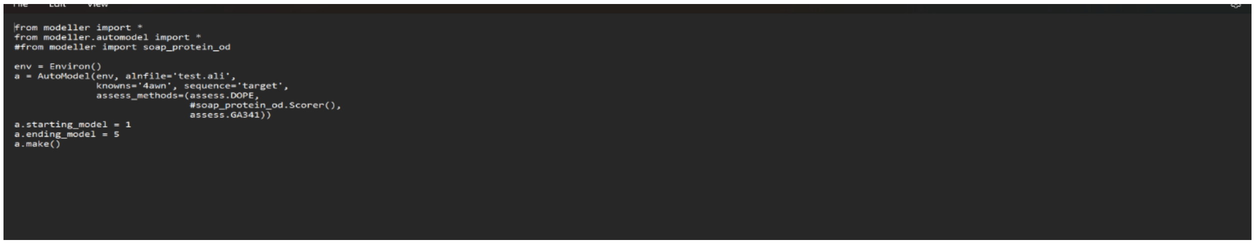
```
>P1;target
sequence:target:::::::0.00: 0.00
MRSLCFVLLFVWFFYQNEAKSKVSCKDEAGNDVDWWHLYKLPKHYQHNDLGKDTSGLKYL
YVTSQNYDTWQMSGKFISDPLSLPAQTLNPLNDDPSHTLLAAYNDQQPNGTVFSSGGHAK
GVVASDGETAIWIVHSVPKFPTIPDYSYPTSGEQYAQSMLCVTLKGEDLEKVGQILVYNE
PHFYYQRNPLATRSDELFPSLERALHGQWRTESPFQKDLEVRSLDGKKFRLFGKSGRANV
ELYADVVAPTLDVSLFVEAWRDGAGNLPNSCDKSDKVLNVESISNPELSVDFKTTQDHSK
WAVSRPTGILIYHWRVGGGDWICVGDINRQEGQLHRGGGTVCHKSARVSNLYRQLVTNYD
KCAQQE*
```

6. Open align2d.py in Notepad



```
from modeller import *

env = Environ()
aln = Alignment(env)
mdl = Model(env, file='4awn', model_segment=('FIRST:A','LAST:A'))
aln.append_model(mdl, align_codes='4awn', atom_files='4awn.pdb')
aln.append(file='target.pir', align_codes='target')
aln.align2d(max_gap_length=50)
aln.write(file='test.ali', alignment_format='PIR')
aln.write(file='test.pap', alignment_format='PAP')
```

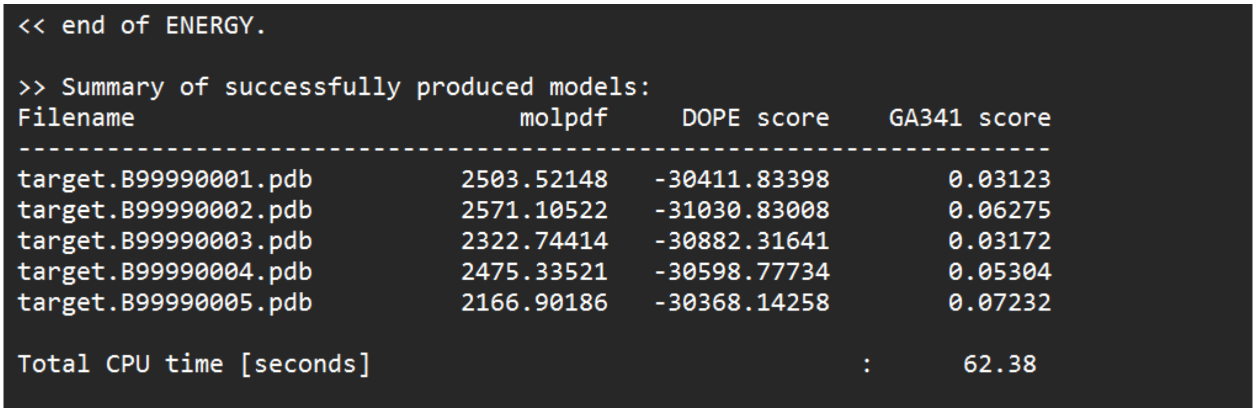
7. Change (env, file='4awn', (align_codes='4awn', atom_files='4awn.pdb'); (file='target.pir', align_codes='target)
8. Open the modeller app and set the working directory using cd.
9. Then give the command mod10.5 align2d.py and run it.
10. Again, from tutorial get the script for model building and modify accordingly and save as model.py



```
from modeller import *
from modeller.automodel import *
#from modeller import soap_protein_od

env = Environ()
a = AutoModel(env, alnfile='test.ali',
              knowns='4awn', sequence='target',
              assess_methods=(assess.DOPE,
                              #soap_protein_od.Scorer(),
                              assess.GA341))
a.starting_model = 1
a.ending_model = 5
a.make()
```

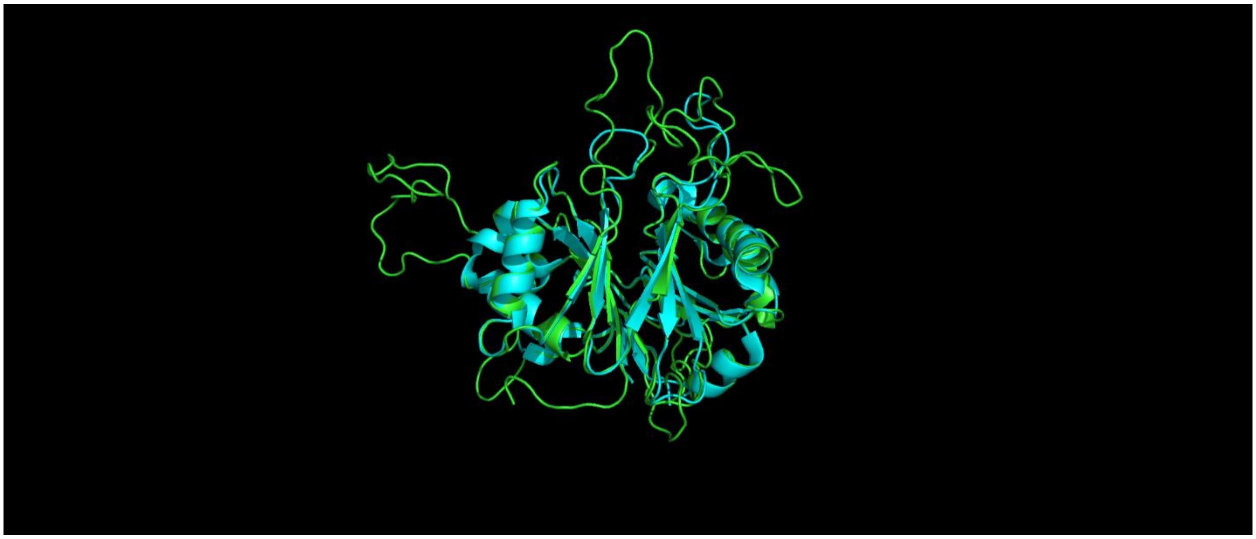
11. In modeller give the command mod10.5 model.py and run to generate the models.

12. In the document named model check the Dope score and select the model with the lowest score. Select model 5 that is having lowest dope score.

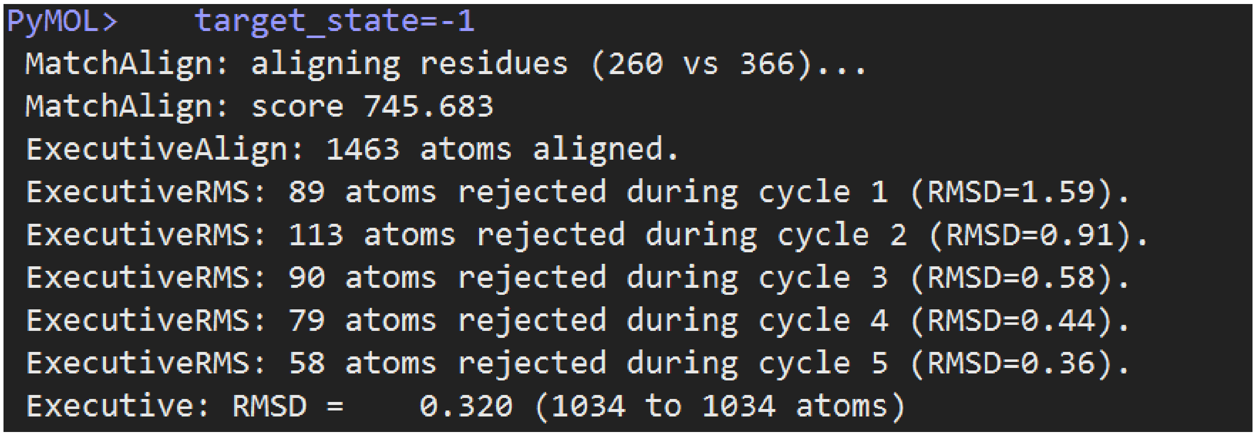


```
<< end of ENERGY.

>> Summary of successfully produced models:
Filename                          molpdf     DOPE score    GA341 score
----------------------------------------------------------------------
target.B99990001.pdb          2503.52148   -30411.83398        0.03123
target.B99990002.pdb          2571.10522   -31030.83008        0.06275
target.B99990003.pdb          2322.74414   -30882.31641        0.03172
target.B99990004.pdb          2475.33521   -30598.77734        0.05304
target.B99990005.pdb          2166.90186   -30368.14258        0.07232

Total CPU time [seconds]                                 :      62.38
```

13. Open Pymol app and select file and then open. Go to the folder created and select the model2 and the 4AWN structure.

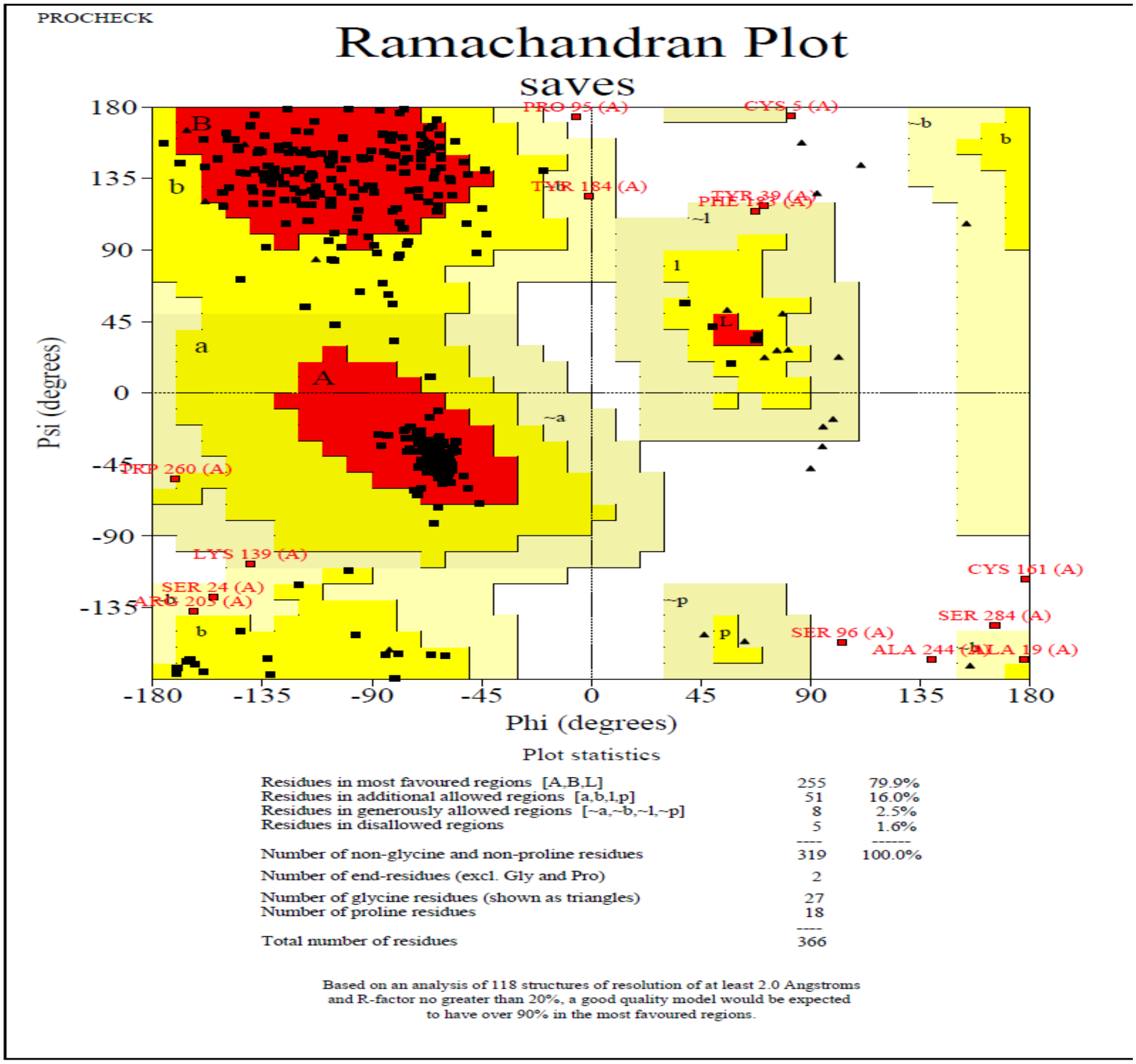


14. Then go to tools and select structure comparison and then match maker.
15. Select the reference sequence and structure to match here as 4AWN .pdb and apply.



```
PyMOL>    target_state=-1
 MatchAlign: aligning residues (260 vs 366)...
 MatchAlign: score 745.683
 ExecutiveAlign: 1463 atoms aligned.
 ExecutiveRMS: 89 atoms rejected during cycle 1 (RMSD=1.59).
 ExecutiveRMS: 113 atoms rejected during cycle 2 (RMSD=0.91).
 ExecutiveRMS: 90 atoms rejected during cycle 3 (RMSD=0.58).
 ExecutiveRMS: 79 atoms rejected during cycle 4 (RMSD=0.44).
 ExecutiveRMS: 58 atoms rejected during cycle 5 (RMSD=0.36).
 Executive: RMSD =    0.320 (1034 to 1034 atoms)
```

16. Then to validate the structure, go to UCLA-DOE lab saves and load the file of model 2.


PROCHECK

Ramachandran Plot

saves

```
                         Plot statistics

Residues in most favoured regions  [A,B,L]                  255     79.9%
Residues in additional allowed regions  [a,b,l,p]            51     16.0%
Residues in generously allowed regions  [~a,~b,~l,~p]         8      2.5%
Residues in disallowed regions                                5      1.6%
                                                            ----   ------
Number of non-glycine and non-proline residues              319    100.0%

Number of end-residues (excl. Gly and Pro)                    2

Number of glycine residues (shown as triangles)              27
Number of proline residues                                   18
                                                            ----
Total number of residues                                    366
```

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms
and R-factor no greater than 20%, a good quality model would be expected
to have over 90% in the most favoured regions.

Residues in most favoured regions- 255

Residues in additional allowed regions- 51

Residues in generously allowed regions-8

Residues in disallowed regions- 5

INFERENCE:

The structure of unknown protein was predicted using modeller.

EX.NO : 5	Binding Site Identification
Date : 21/08/24	

AIM:

To identify the binding site of a protein using Cast-P, POCASA, 3D ligand site.

MOLECULE:

Protein name: DYRK1A in complex with methoxy benzothiazole fragment Organism: Homo sapiens

PDB Id: 5A4E

Length: 368

TOOL : CASTp 3.0

TOOL INFORMATION:

Computed Atlas of Surface Topography of proteins (CASTp) provides an online resource for locating, delineating and measuring concave surface regions on three-dimensional structures of proteins.

These include pockets located on protein surfaces and voids buried in the interior of

proteins. The measurement includes the area and volume of pocket or void by solvent accessible surface model (Richards' surface) and by molecular surface model (Connolly's surface), all calculated analytically. CASTp can be used to study surface features and functional regions of proteins. CASTp includes a graphical user interface, flexible interactive visualization, as well as on-the-fly calculation for user uploaded structures.

METHODOLOGY:

i) CASTp was accessed from the source

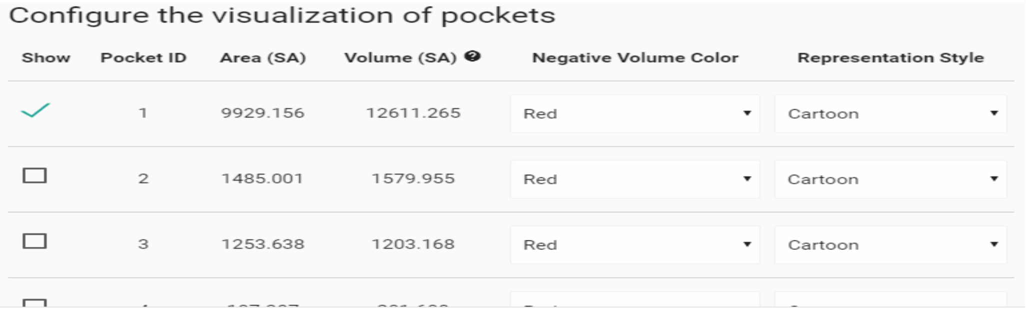
ii) The protein was given in PDB format as input and submitted.

iii) The largest pocket was selected from the results and note down the residue.

iv) Visualize the protein with the pocket.

RESULTS:

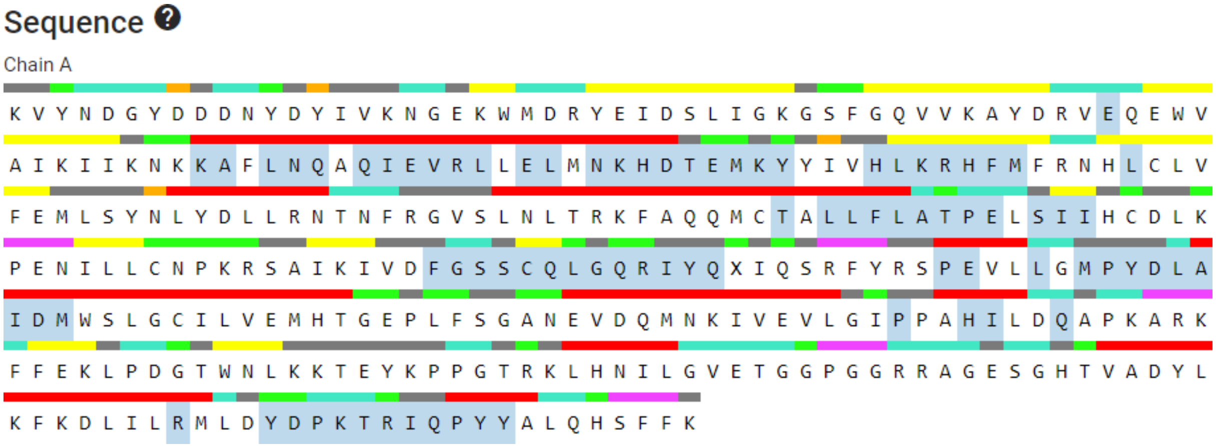
Number of pockets and the amino acids residues in the pocket



Configure the visualization of pockets

Show	Pocket ID	Area (SA)	Volume (SA)	Negative Volume Color	Representation Style
✓	1	9929.156	12611.265	Red	Cartoon
□	2	1485.001	1579.955	Red	Cartoon
□	3	1253.638	1203.168	Red	Cartoon

Protein Sequence :

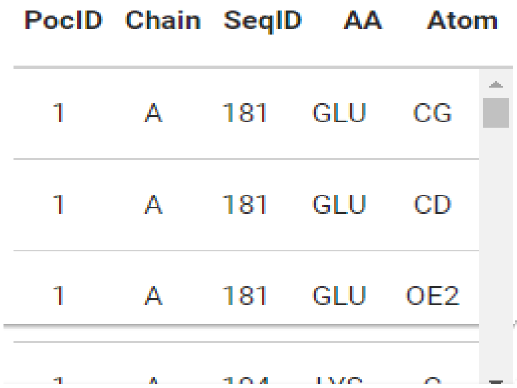


Sequence

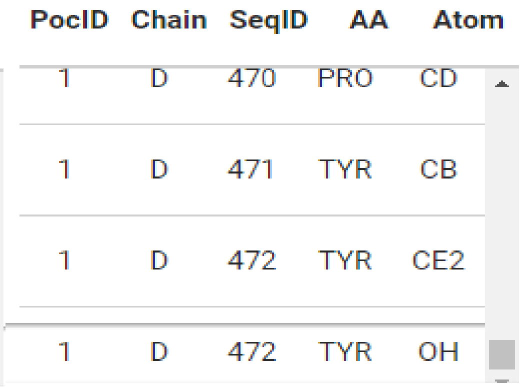
Chain A

KVYNDGYDDDNYDYIVKNGEKWMDRYEIDSLIGKGSFGQVVKAYDRVEQEWV
AIKIIKNKKAFLNQAQIEVRLLELMNKHDTEMKYYIVHLKRHFMFRNHLCLV
FEMLSYNLYDLLRNTNFRGVSLNLTRKFAQQMCTALLFLATPELSIIHCDLK
PENILLCNPKRSAIKIVDFGSSCQLGQRIYQXIQSRFYRSPEVLLGMPYDLA
IDMWSLGCILVEMHTGEPLFSGANEVDQMNKIVEVLGIPPAHILDQAPKARK
FFEKLPDGTWNLKKTEYKPPGTRKLHNILGVETGGPGGRRAGESGHTVADYL
KFKDLILRMLDYDPKTRIQPYYALQHSFFK

Binding site residues :

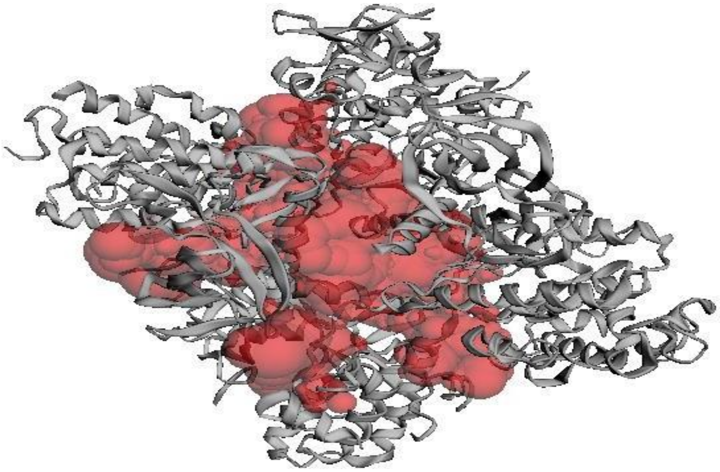


PocID	Chain	SeqID	AA	Atom
1	A	181	GLU	CG
1	A	181	GLU	CD
1	A	181	GLU	OE2
1	A	184	LYS	C



PocID	Chain	SeqID	AA	Atom
1	D	470	PRO	CD
1	D	471	TYR	CB
1	D	472	TYR	CE2
1	D	472	TYR	OH

Visualization of protein with pocket :



Tool/Server: POCASA 1.1

TOOL INFORMATION:

POCASA (POcket-CAvity Search Application) is an automatic program that implements the algorithm named Roll which can predict binding sites by detecting pockets and cavities of proteins of the known 3D structure. First, a 3D grid system is created and filled with atoms in theprotein molecule. Second, a probe sphere is adapted to roll along the protein surface to generate a"probe surface" based on the inner border tracing algorithm in the image processing field. Then,the regions between the protein and the probe surface or those surrounded by the protein surfaceare defined as pockets and cavities, respectively.

To remove noise points, two parameters were designed: Single-Point Flag (SPF) and Protein-Depth Flag (PDF). Moreover, POCASA can predict pockets differing in shape and volume by adjusting the radius of the probe sphere.

Finally, Volume-Depth (VD) quantitively describing the volume and position information of pockets was designed as a pocket-ranking descriptor.

METHODOLOGY:

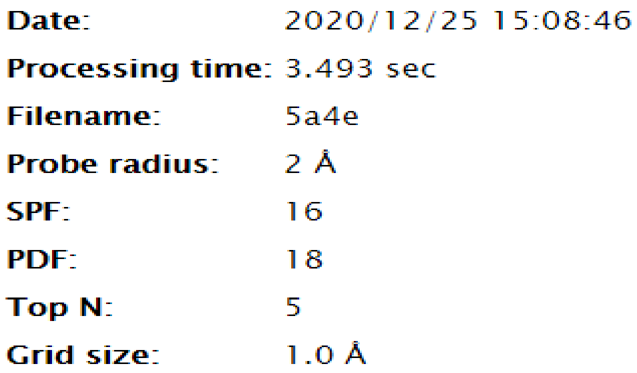
i) POCASA was accessed from the source

ii) The protein was given in PDB format as input and submitted.

iii) It generates four file - the protein PDB file, a parameter text file, a pdb file with the pockets and a PDB file with the geometric center of the pockets.

iv) Visualization was done using PyMol

RESULTS :



Date: 2020/12/25 15:08:46

Processing time: 3.493 sec

Filename: 5a4e

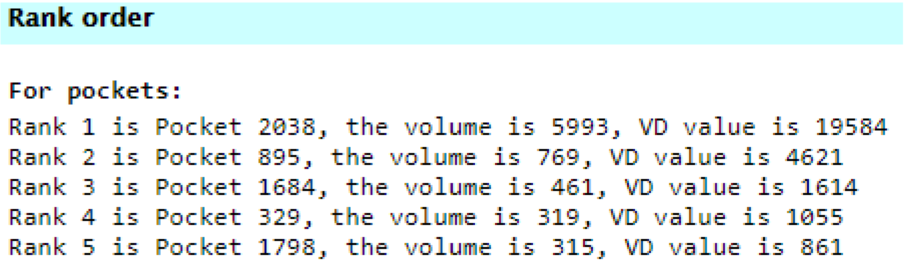
Probe radius: 2 Å

SPF: 16

PDF: 18

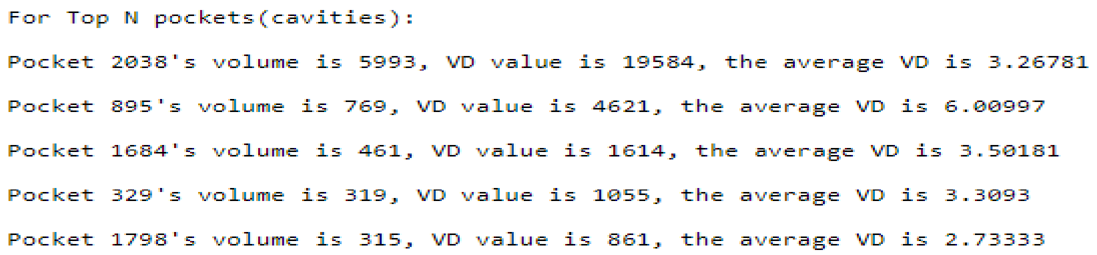
Top N: 5

Grid size: 1.0 Å



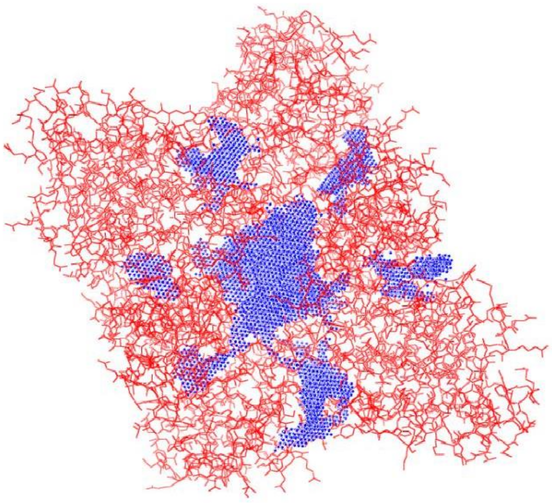
Rank order

```
For pockets:
Rank 1 is Pocket 2038, the volume is 5993, VD value is 19584
Rank 2 is Pocket 895, the volume is 769, VD value is 4621
Rank 3 is Pocket 1684, the volume is 461, VD value is 1614
Rank 4 is Pocket 329, the volume is 319, VD value is 1055
Rank 5 is Pocket 1798, the volume is 315, VD value is 861
```

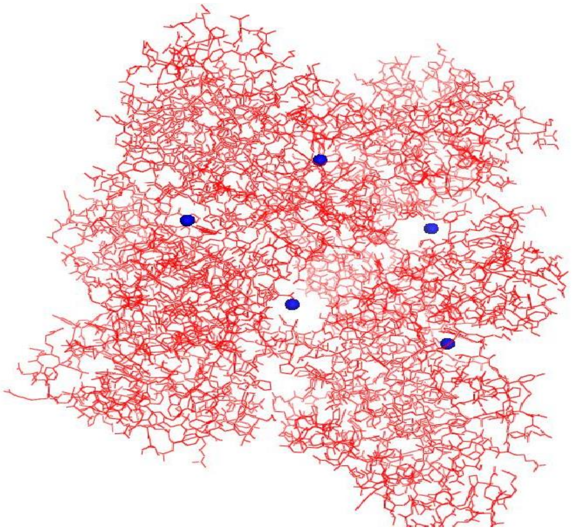


```
For Top N pockets(cavities):

Pocket 2038's volume is 5993, VD value is 19584, the average VD is 3.26781

Pocket 895's volume is 769, VD value is 4621, the average VD is 6.00997

Pocket 1684's volume is 461, VD value is 1614, the average VD is 3.50181

Pocket 329's volume is 319, VD value is 1055, the average VD is 3.3093

Pocket 1798's volume is 315, VD value is 861, the average VD is 2.73333
```

TOP N pockets with 5A4E



Pocket Depth Centers with 5A4E



INFERENCE:

The tools were able to predict the pocket with almost similar numbers of residue for interacting in their top three predicted cavities.

EX.NO: 6	**Structure-Based Drug Design Using AutoDock**
Date: 18/09/24	

AIM:
To retrieve the protein and ligand structure from PDB Database and perform molecular docking studies and find interactions between them.

Required SOFTWARE: Auto dock v1.5.6
MOLECULAR DOCKING:
Molecular docking is a method that predicts the orientation of bonding one molecule with another molecule to form a stable complex. It is a frequently used method in structure-based drug designing to predict the binding conformation of the ligand with the target binding site. Docking is a term used for computational schemes that attempt to find the "best" matching between two molecules: a receptor and a ligand. In the docking, the program obtains the image of the binding site from the molecular surface of the macromolecule and ligand molecules are mapped on to the binding and then docking energies and scores have been evaluated.

Input Macromolecule:
Protein Name: Crystal structure of human CDK2 in complex with the inhibitor olomoucine
- PDB ID : 1W0X
- Ligand Name: Acetaminophen

Tools Required:
1. PDB Database (http://www.rcsb.org)
2. PyMOL Visualization tool
3. WHAT IF Web Interface
4. Swiss-PdbViewer v4.1.0
5. Docking Tool – AutoDockTools-1.5.6

METHODOLOGY:
A) Retrieval of Protein:
□ Open the PDB database in a web browser and give PDB ID in the search bar.
□ Download the protein in .pdb format.
□ Open the downloaded protein in PyMOL and remove crystal water molecules and ligand. And save only the protein as. pdb file.
B) Retrieval of Ligand:
Download the ligand molecule & save ligand in .pdb format.
C) Optimization and Energy Minimization of Protein:
□ WHAT IF Web Interface is used to complete and optimize the structure of the protein molecule, where it helps to model missing side chains in a protein molecule.
□ Open WHAT IF Web Interface online Server.
□ Build/check/repair model → Complete a structure. Upload protein.PDB file and run the program. Download the fixed.pdb as a result file (complete structure).

Swiss-PdbViewer
□ Swiss-PdbViewer tool for optimizing and energy minimizing of the protein molecule by

Steepest Descent and Conjugate Gradient Methods.
□ Open Swiss-Pdb-Viewer tool for Energy Minimization. Load the output file fixed.pdb in Swiss-Pdb Viewer tool.
□ Click on the CONTROL PANEL and select all the amino acids.
□ Click on Pref →Energy Minimization select the number of cycles for Steepest Descent method and Conjugate Gradient method.
□ Run the Energy Minimization Tools→Energy minimization or Ctrl+N and save the optimized protein molecule as minimized .pdb

D) **Molecular Docking:**
□ Open AutoDockTools-4.2.6
□ Set directory path in preferences→set.
□ Keep the autogrid.exe and autodock.exe programming files in the same path.

Protein preparation: Protein preparation is a process keeping protein structure ready for docking. AutoDock is based on the United Atom force-field of AMBER, which uses only polar hydrogens. This helps to reduce the number of atoms that must be modeled explicitly during the docking, thus speeding up the calculations. Polar hydrogens are hydrogen atoms that are bonded to electronegative atoms like oxygen and nitrogen. That makes the protein more interactive to the ligand.

Read the fixed.pdb file in AutoDockTools-4.2.6
• File → Read □Open fixed protein.pdb file Add Hydrogens to the protein molecule.
• Edit →Hydrogens→ Add →Only Polar Add Charges to a protein molecule.
• Edit→ Charges→ Kollman charges Assign Atom type to a protein molecule.
• Edit→ Atoms →AssignAD4
Save the protein molecule file as fixed.pdbqt format.

Ligand preparation: Ligand preparation is a process of making ligand structure ready for docking. This is achieved by detecting root and adding charge. It is done because to make ligand to interact at the specific pocket of protein structure. Read the ligand structure in AutoDockTools-1.5.6
• Ligand→ Input → Open
Detect the root for ligand structure.
• Ligand → Torsion→ Detect
Find out the charges in the ligand.
• Edit → Charges → Compute gasteier
Save the ligand file as ligand.pdbqt format.
Grid Box Preparation:
Grid box preparation is setting the predicted interaction site of ligand to the target pocket. The box is adjusted at the interactive site on X, Y, Z co-ordinations. Here use both fixed.pdbqt and ligand.pdbqt. Read the ligand structure from

ligand.pdbqt to Grid Box Set the Grid box at the predicted binding pocket, by adjusting X, Y, Z dimension parameters. Read protein molecule from fixed.pdbqt to grid.

- Grid → Macromolecule → Open

Read ligand molecule from ligand.pdbqt to the grid

- Grid→ SetMap → Open Ligand

Adjust the Grid Box at possible interaction site.

- Grid → Grid Box

Set the Grid Box in the right position or around the binding site of the protein. File → close saving current.

Grid → Output → Save as grid.gpf

Docking Parameters:

Read protein molecule from fixed.pdbqt file to docking

- Docking → Macromolecules → Set Rigid Read ligand structure from ligand.pdbqt file to docking.
- Ligand → Docking → Open

Set the Search parameters by using the Genetic algorithm.

- Docking → Search Parameter → Genetic Algorithm Set the Docking parameters by using docking options
- Docking →Output → Lamarckian GA (4.2)

Run:

Run auto grid using autogrid.exe programming file.

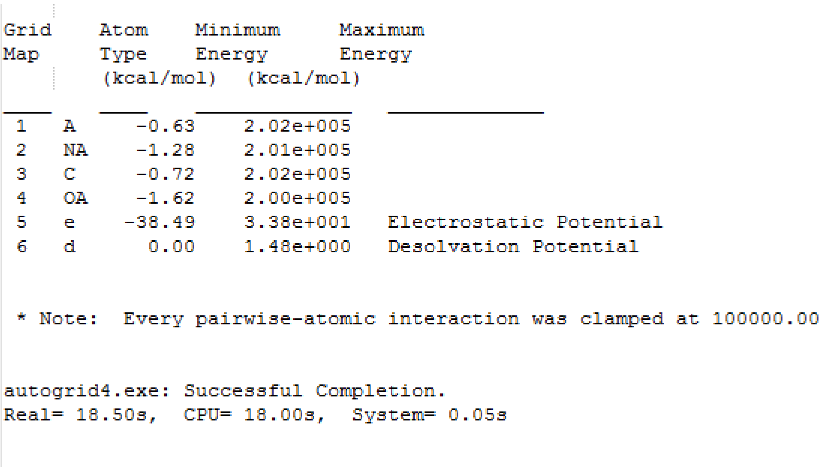
- Run →Autogrid grid.gpf →grid.glg

Run auto dock using autodock.exe programming file.

- Run → Autodock dock.gpf → dock.glg

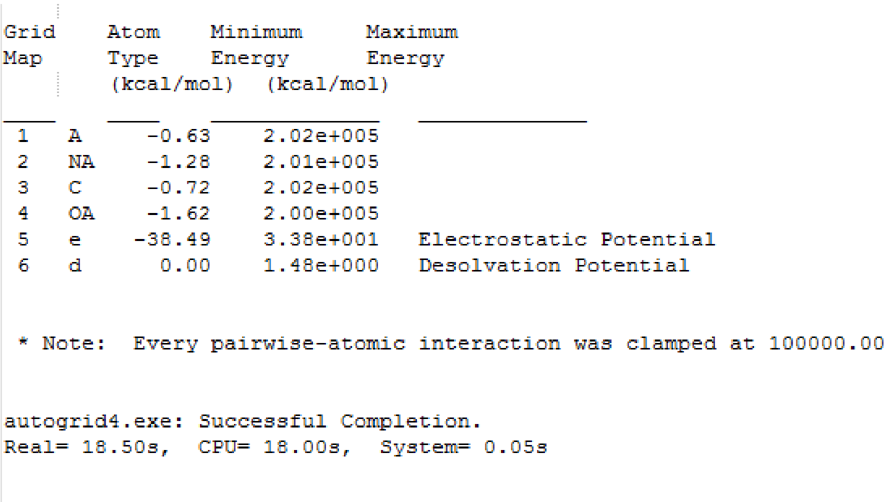
RESULTS :

Autogrid output file :



```
Grid    Atom    Minimum      Maximum
Map     Type    Energy       Energy
         (kcal/mol)  (kcal/mol)

____    ____   ____________    ____________
 1   A      -0.63    2.02e+005
 2   NA     -1.28    2.01e+005
 3   C      -0.72    2.02e+005
 4   OA     -1.62    2.00e+005
 5   e     -38.49    3.38e+001   Electrostatic Potential
 6   d       0.00    1.48e+000   Desolvation Potential


 * Note:  Every pairwise-atomic interaction was clamped at 100000.00


autogrid4.exe: Successful Completion.
Real= 18.50s,  CPU= 18.00s,  System= 0.05s
```

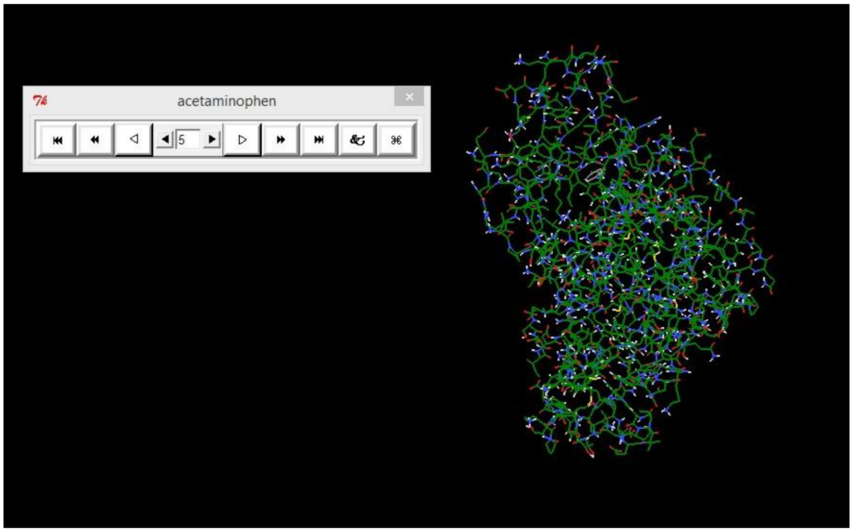
Autodock Output file –



```
Grid    Atom    Minimum     Maximum
Map     Type    Energy      Energy
        (kcal/mol)  (kcal/mol)

____    ____    _____________   _____________
 1   A      -0.63    2.02e+005
 2   NA     -1.28    2.01e+005
 3   C      -0.72    2.02e+005
 4   OA     -1.62    2.00e+005
 5   e     -38.49    3.38e+001   Electrostatic Potential
 6   d       0.00    1.48e+000   Desolvation Potential


 * Note:  Every pairwise-atomic interaction was clamped at 100000.00


autogrid4.exe: Successful Completion.
Real= 18.50s,  CPU= 18.00s,  System= 0.05s
```

Analyze Docking result

Analyze → Conformations →Play



Conclusion:

- Total Kollman Charge → 15.504
- Binding energy → - 3.78 KJ/mol
- Number of hydrogen bonds - 2 bonds
- Name of the residues interacting: Leu and Asp

Inference:

After performing molecular docking, the ligand is not superimposed well. It forms a hydrogen bond with LEU83 and ASP86 residue with distance measuring 1.730Å and 1.973 Å respectively.

EX.NO : 7	MOLECULAR DYNAMICS SIMULATION USING GROMACS
Date: 09/10/24	

AIM: Molecular Dynamics Simulation using Gromacs

MATERIAL/TOOLS: Linux, GROMACS, GRACE, PyMol

Introduction:

GROMACS is one of the most widely used open-source and free software codes in chemistry, used primarily for dynamical simulations of biomolecules. It provides a rich set of calculation types, preparation and analysis tools. Several advanced techniques for free-energy calculations are supported.

Molecular structures, which were solved by X-ray crystallography, nuclear magnetic resonance, or in any other way are stored in the Protein Data Bank. To make those data easily accessible to researchers, the PDB file format had been developed. This type of file contains information about atom names, their positions in the Cartesian coordinate system, bindings to other atoms, and some other auxiliary information such as locations of the secondary structure elements. The PDB formatted files are plain text, so they can be easily read and manipulated by most programs dedicated to molecular structure analysis. They can also be directly read by any text viewing program.

Molecular dynamics (MD) simulations are powerful computational tools used to explore the structural, dynamic, and thermodynamic properties of biomolecules at an atomic level. In this study, we performed a comprehensive MD simulation of a protein system using GROMACS, a widely used software for simulating the molecular mechanics of proteins, lipids, and other biomolecules. The simulation workflow included preprocessing steps such as cleaning the protein structure, solvating the protein in a water box, ion addition to neutralize the system, energy minimization, and equilibration phases. We then ran a production MD simulation, capturing the dynamic behavior of the protein over time. The results were analyzed to obtain essential properties such as Root Mean Square Deviation (RMSD), Radius of Gyration, and system density over time.

The objectives of this MD simulation were:

1. **To analyze structural stability** of the protein throughout the simulation by calculating RMSD and Radius of Gyration.
2. **To assess system stability** during the equilibration phases (NVT and NPT) by tracking temperature, pressure, and density.

PROCEDURE:

Downloading protein structure:

Go to the RCSB website and download the PDB text for the crystal structure. Once the structure has been downloaded, it can be visualized using a viewing program such as VMD, Chimera, PyMOL, etc.

Deleting water molecules:

use grep to delete these lines very easily

Preparing input file for gromacs using pdb2gmx: Now that the crystal waters are gone and we have verified that all the necessary atoms are present, the PDB file should contain only protein atoms, and is ready to be input into the first GROMACS module, pdb2gmx. The purpose of pdb2gmx is to generate three files:

- The topology for the molecule.
- A position restraint file.
- A post-processed structure file.

The topology (topol.top by default) contains all the information necessary to define the molecule within a simulation. This information includes nonbonded parameters (atom types and charges) as well as bonded parameters (bonds, angles, and dihedrals). Execute pdb2gmx

Generated three new files:

1. **1AKI processed.gro**:
   1AKI_processed.gro is a GROMACS-formatted structure file that contains all the atoms defined within the force field (Le, H atoms have been added to the amino acids in the protein)

1. **Topol.top:** The topol.top file is the system topology.

2. **Posre.itp**: The posre itp file contains information used to restrain the positions atoms.

Defining box for solvation:

It is possible to simulate proteins and other molecules in different solvents, provided that good parameters are available for all species involved. There are two steps to defining the box and filling it with solvent: Define the box dimensions using the editconf module. Fill the box with water using the solvate module.

Solvating the box:

Now that we have defined a box, we can fill it with solvent (water). Solvation is accomplished using solvate

gmx solvate -cp 1AKI_newbox gro -cs spc216.gro -o 1AKI_solv.gro -p topol.top The configuration of the protein (-cp) is contained in the output of the previous editconf step, and the configuration of the solvent (-es) is part of the standard GROMACS installation. We are using spc216.gro, which is a generic equilibrated 3-point solvent model. You can use spe216.gro as the solvent configuration for SPC, SPC/E, or TIP3P water, since they are all three-point water models. The output is called 1AKI solv.gro, and we tell to solvate the name of the topology file (topol.top) to modify it.

Adding ions:

Assemble tpr file
Now we have an atomic-level description of our system in the binary file ions.tpr.

Energy Minimization:

The solvated, electroneutral system is now assembled. Before we can begin dynamics, we must ensure that the system has no steric clashes or inappropriate geometry. The structure is relaxed through a process called energy minimization (EM).

GROMACS Molecular Dynamics Simulation Protocol

1. Remove Water Molecules from the PDB File

```
grep -v HOH 1aki.pdb > output_clean.pdb
```

- 1aki.pdb original PDB file name.
- output_clean.pdb will be the cleaned PDB file with water molecules removed.

---

2. Convert PDB to GROMACS Format

```
gmx pdb2gmx -f output_clean.pdb -o processed.gro -water spce
```

- output_clean.pdb is the input cleaned PDB file.
- processed.gro is the output file in GROMACS format.

---

3. Define the Simulation Box

```
gmx editconf -f processed.gro -o newbox.gro -c -d 1.0 -bt cubic
```

- Adjust -d 1.0 to set the box size (distance in nm between solute and box edge).
- newbox.gro is the output GROMACS file for the box.

---

4. Solvate the System

```
gmx solvate -cp newbox.gro -cs spc216.gro -o solvated.gro -p topol.top
```

- solvated.gro is the output with solvent molecules added.

---

5. Add Ions

Prepare Input for Ion Addition:

```
gmx grompp -f ions.mdp -c solvated.gro -p topol.top -o ions.tpr
```

Neutralize the System by Adding Ions:

```
gmx genion -s ions.tpr -o solvated_ions.gro -p topol.top -pname NA -nname CL -neutral
```

- solvated_ions.gro is the output with ions added.

---

6. Energy Minimization

Prepare for Minimization:

```
gmx grompp -f minim.mdp -c solvated_ions.gro -p topol.top -o em.tpr
```

Run Minimization:

```
gmx mdrun -v -deffnm em
```

Analyze Potential Energy:

```
gmx energy -f em.edr -o potential.xvg
```

---

7. Equilibration

7.1 NVT Equilibration (Constant Volume, Temperature)

Prepare NVT Equilibration:

```
gmx grompp -f nvt.mdp -c em.gro -r em.gro -p topol.top -o nvt.tpr
```

Run NVT Equilibration:

```
gmx mdrun -deffnm nvt
```

Analyze Temperature Profile:

```
gmx energy -f nvt.edr -o temperature.xvg
```

(Select option 16 0 when prompted)

7.2 NPT Equilibration (Constant Pressure, Temperature)

Prepare NPT Equilibration:

```
gmx grompp -f npt.mdp -c nvt.gro -r nvt.gro -t nvt.cpt -p topol.top -o npt.tpr
```

Run NPT Equilibration:

```
gmx mdrun -deffnm npt
```

Analyze Pressure and Density Profiles:

```
gmx energy -f npt.edr -o pressure.xvg
gmx energy -f npt.edr -o density.xvg
```

(Select options 18 0 for pressure and 24 0 for density)

---

8. Production Molecular Dynamics Run

Prepare Production MD Run:

```
gmx grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -o md_0_1.tpr
```

Run Production MD (GPU Enabled):

```
gmx mdrun -deffnm md_0_1 -nb gpu
```

9. Post-Processing and Analysis

Remove Periodic Boundary Conditions:

```
gmx trjconv -s md_0_1.tpr -f md_0_1.xtc -o md_0_1_noPBC.xtc -pbc mol -center
```

Calculate RMSD of Trajectory:

```
gmx rms -s md_0_1.tpr -f md_0_1_noPBC.xtc -o rmsd.xvg -tu ns
gmx rms -s em.tpr -f md_0_1_noPBC.xtc -o rmsd_xtal.xvg -tu ns
```

Calculate Radius of Gyration:

```
gmx gyrate -s md_0_1.tpr -f md_0_1_noPBC.xtc -o gyrate.xvg
```

Results Summary

1. Solvation Box Setup

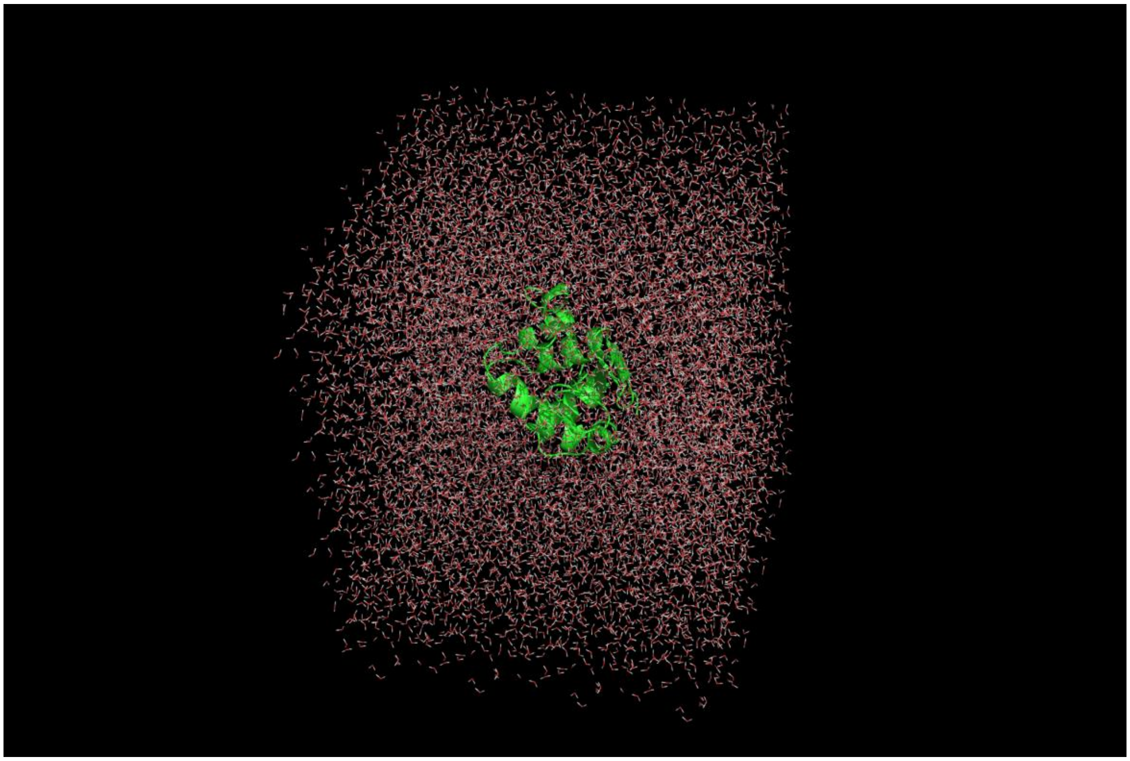


Figure: Solvation Box with Protein and Water Molecules

The protein was placed in a cubic simulation box with a 1 nm buffer distance from the edges, as shown in Figure (Solvation Box). This buffer space allows solvent interactions around the protein without excessive computational overhead. After defining the box, we added water molecules using the SPC/E water model, resulting in a fully solvated system ready for ionization and minimization steps.

2. Energy Minimization

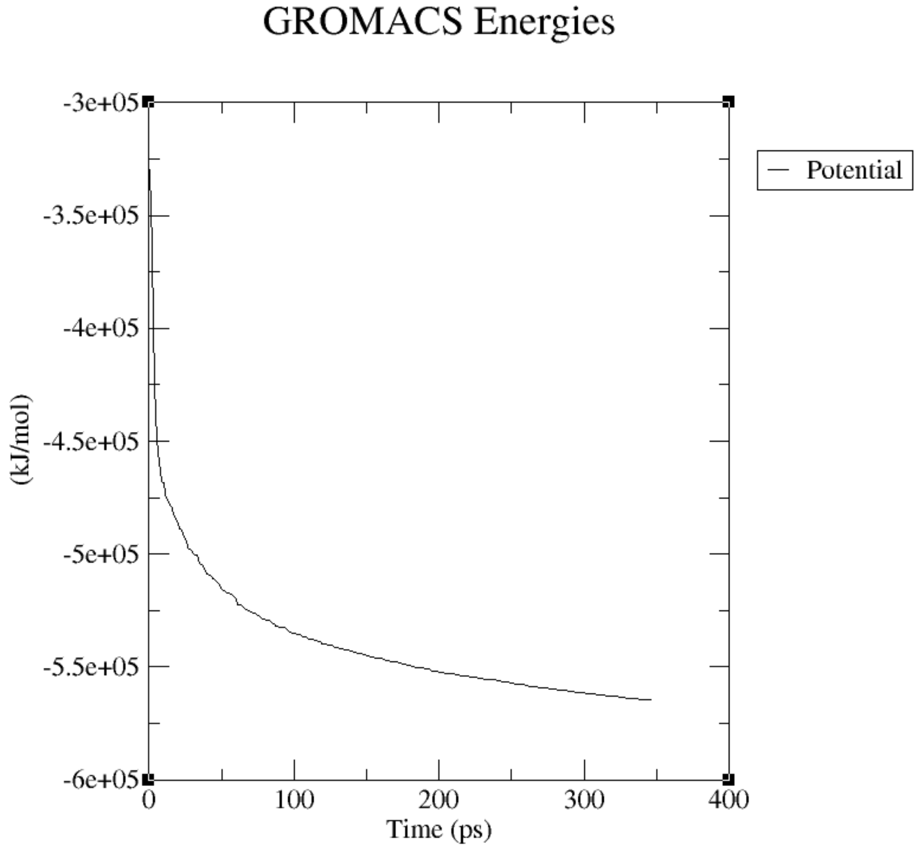


Figure: Potential Energy of the System During Minimization

The energy minimization step is essential to remove steric clashes and unfavorable atomic overlaps in the system. As shown in Figure (Potential Energy), the potential energy decreased significantly over the minimization steps, converging to a stable value. This reduction confirms that the system has achieved a local energy minimum, indicating readiness for the equilibration steps.

3. Equilibration Phase

3.1 NVT Equilibration (Constant Volume, Temperature)

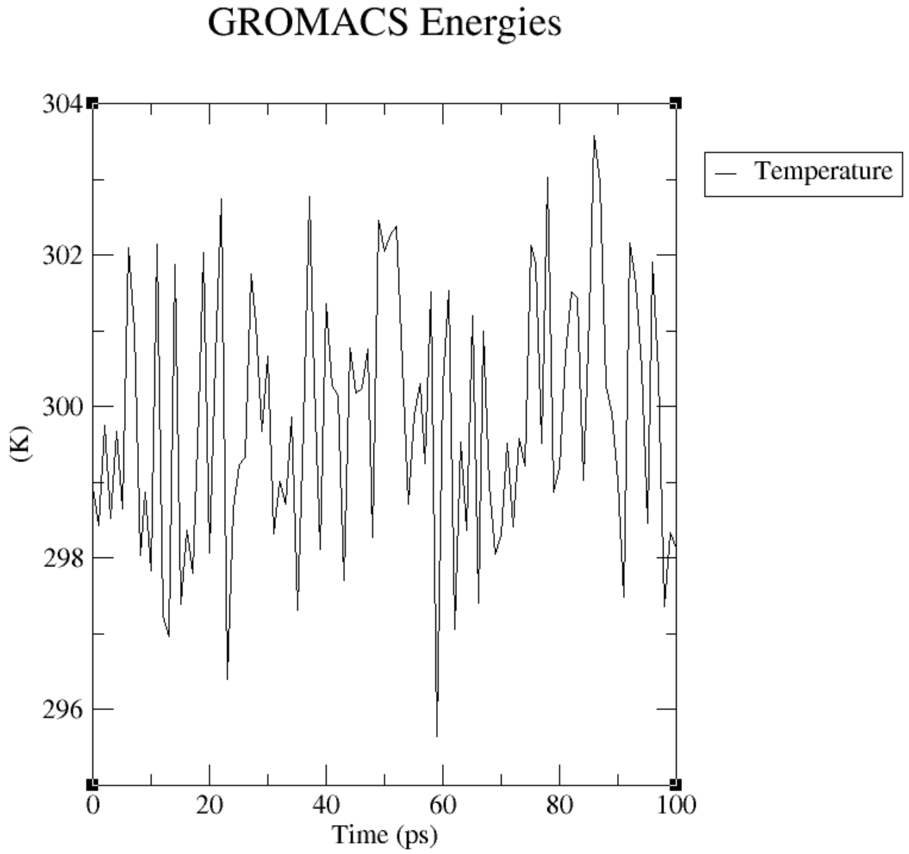


Figure: Temperature Profile During NVT Equilibration

During the NVT equilibration phase, the system's temperature was stabilized to the target of 300 K, as shown in Figure (Temperature Profile). The plot demonstrates a stable temperature achieved after an initial fluctuation period, confirming that the system reached thermal equilibrium under constant volume conditions.

3.2 NPT Equilibration (Constant Pressure, Temperature)

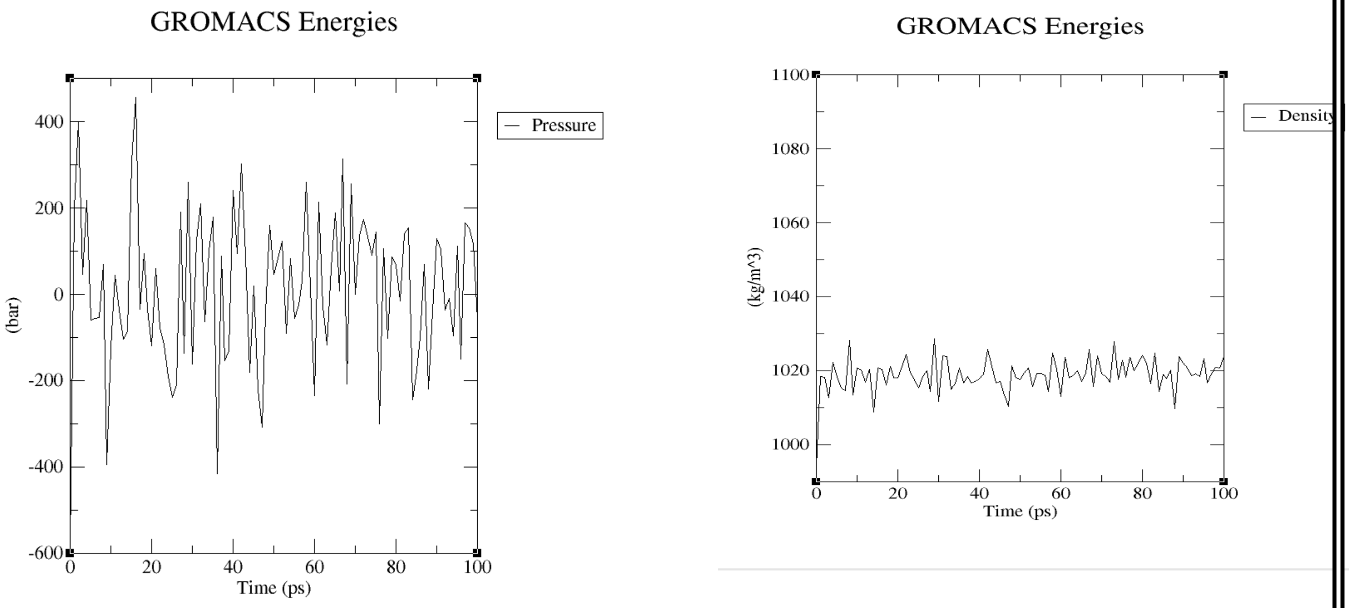


Figure: Pressure and Density Profiles During NPT Equilibration

In the NPT phase, the system was equilibrated under constant pressure to reach a density close to experimental water density. Figure (Pressure Profile) and Figure (Density Profile). The density profile confirms successful equilibration, ensuring system stability before the production MD run.

4. Production MD Simulation

4.1 Structural Stability: RMSD Analysis

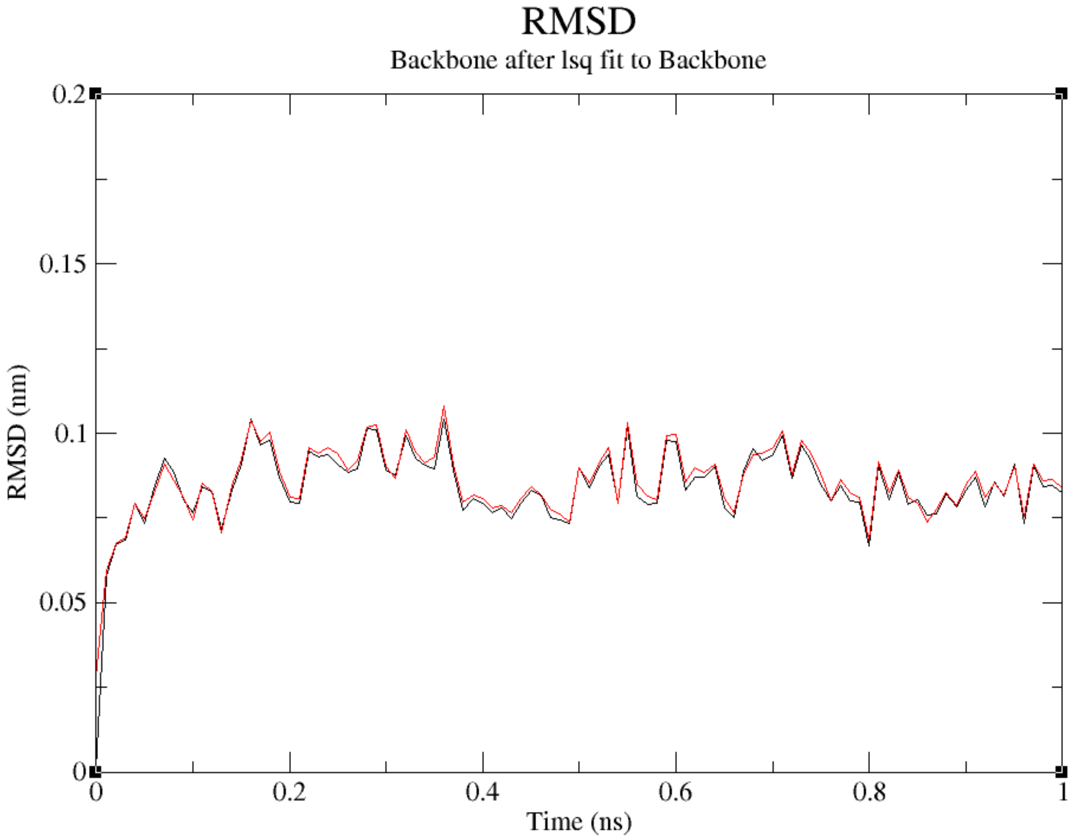


Figure: RMSD of Protein Backbone Over Time

To assess the structural stability of the protein, we calculated the RMSD of the backbone atoms throughout the MD simulation, as shown in Figure (RMSD Profile). The RMSD plot indicates initial structural adaptation followed by stabilization, suggesting that the protein maintains a stable conformation during the simulation, with fluctuations within an acceptable range.

4.2 Compactness: Radius of Gyration

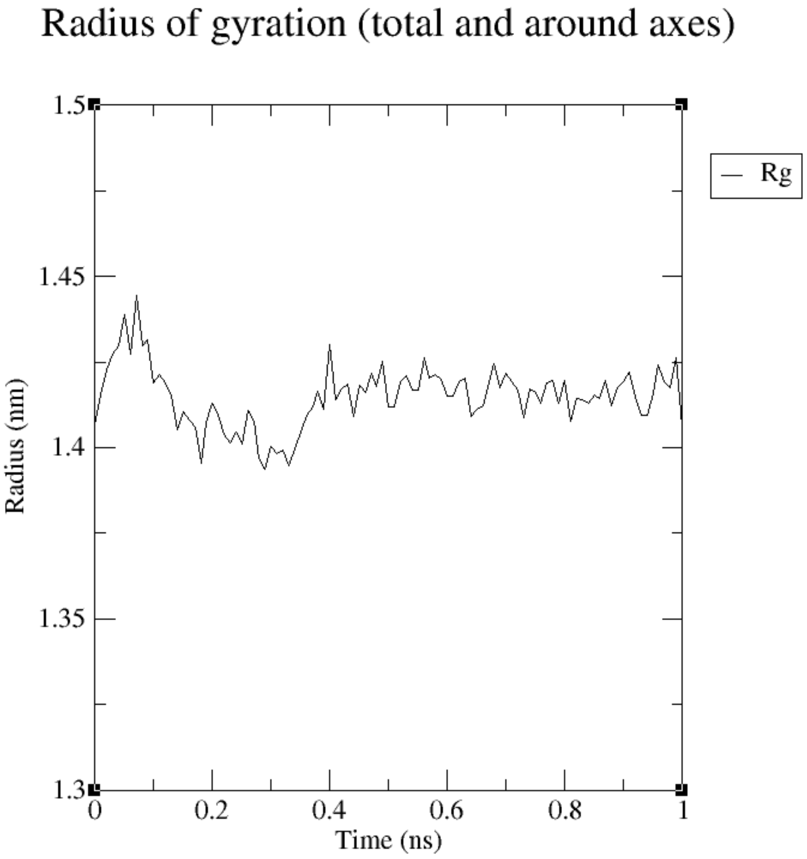
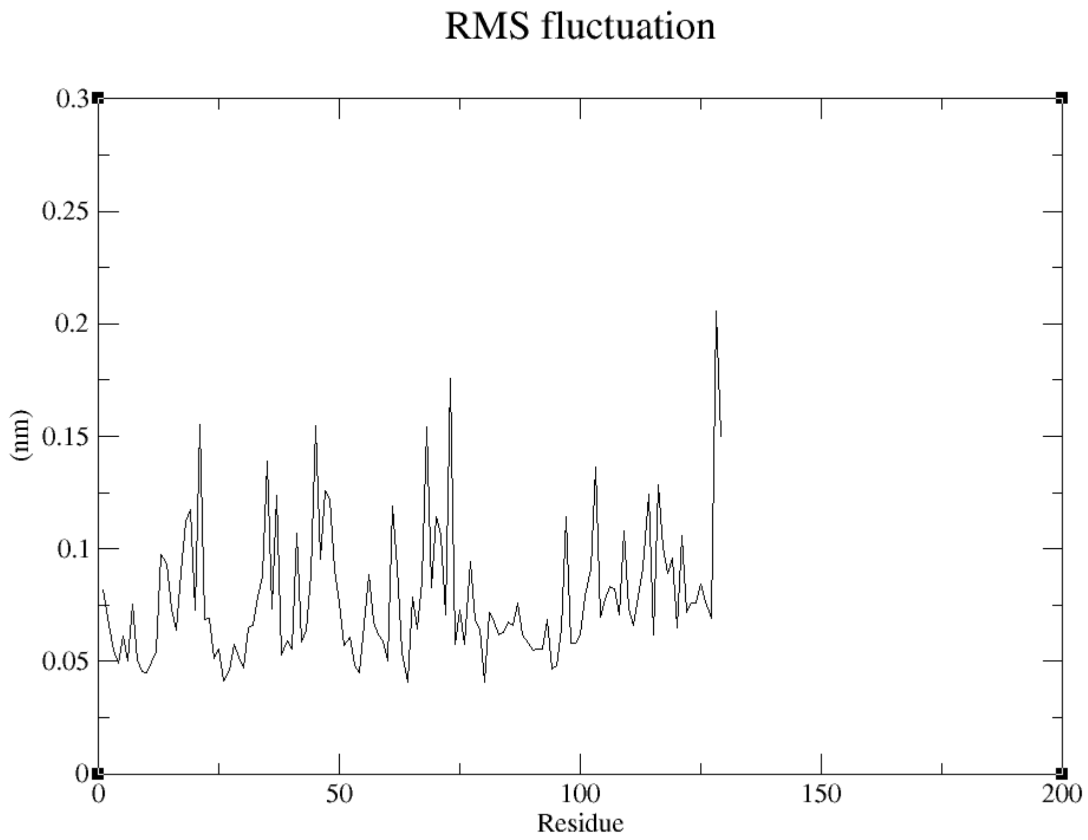


Figure: Radius of Gyration Over Time

The Radius of Gyration (Rg) provides insight into the compactness and folding stability of the protein

4.3RMSF Analysis

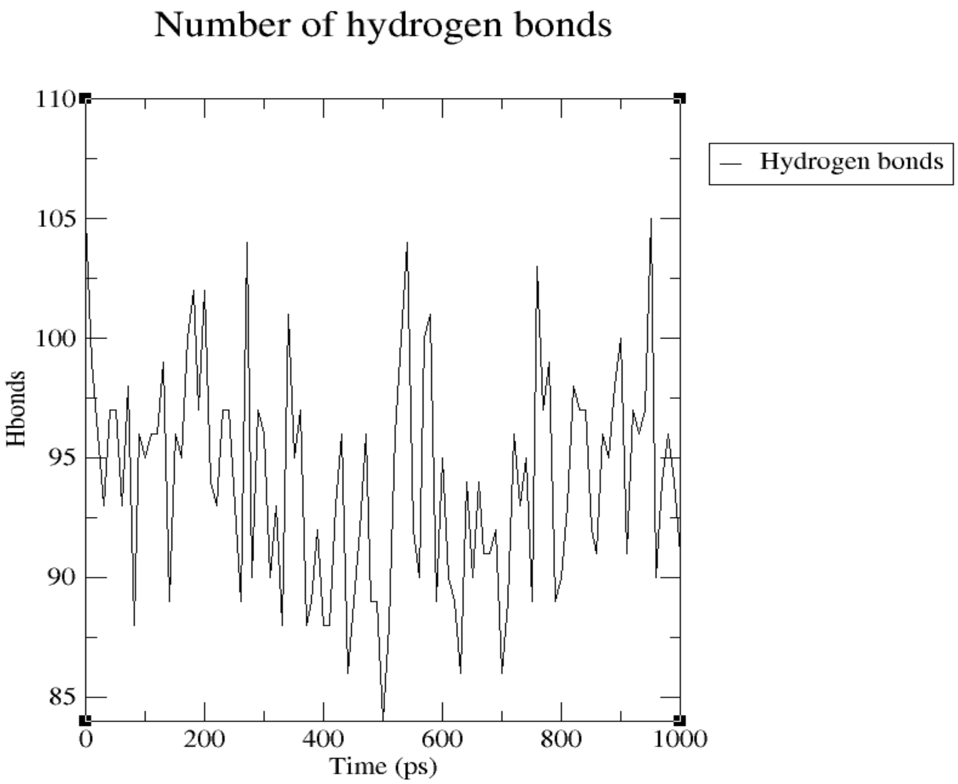
Root Mean Square Fluctuation (RMSF) helps in analyzing the flexibility of residues in a protein by showing their average deviation from a reference position. RMSF values are useful for identifying regions of the protein that are more flexible or stable throughout the simulation.



4.4. Hydrogen Bond Analysis

Hydrogen bonds are key in stabilizing secondary and tertiary structures in proteins. This analysis helps in tracking the number and strength of hydrogen bonds over time in your system.

gmx hbond computes and analyzes hydrogen bonds. Hydrogen bonds are determined based on cutoffs for the angle Hydrogen - Donor- Acceptor (zero is extended) and the distance Donor-Acceptor (or Hydrogen - Acceptor using -noda) OH and NH groups are regarded as donors, O is an acceptor always, N is an acceptor by default, but this can be switched using nitace. Dummy hydrogen atoms are assumed to be connected to the first preceding non-hydrogen atom.


Number of hydrogen bonds
Hydrogen bonds
Hbonds
110
105
100
95
90
85
0
200
400
600
800
1000
Time (ps)

5. Solvent and Ion Distribution

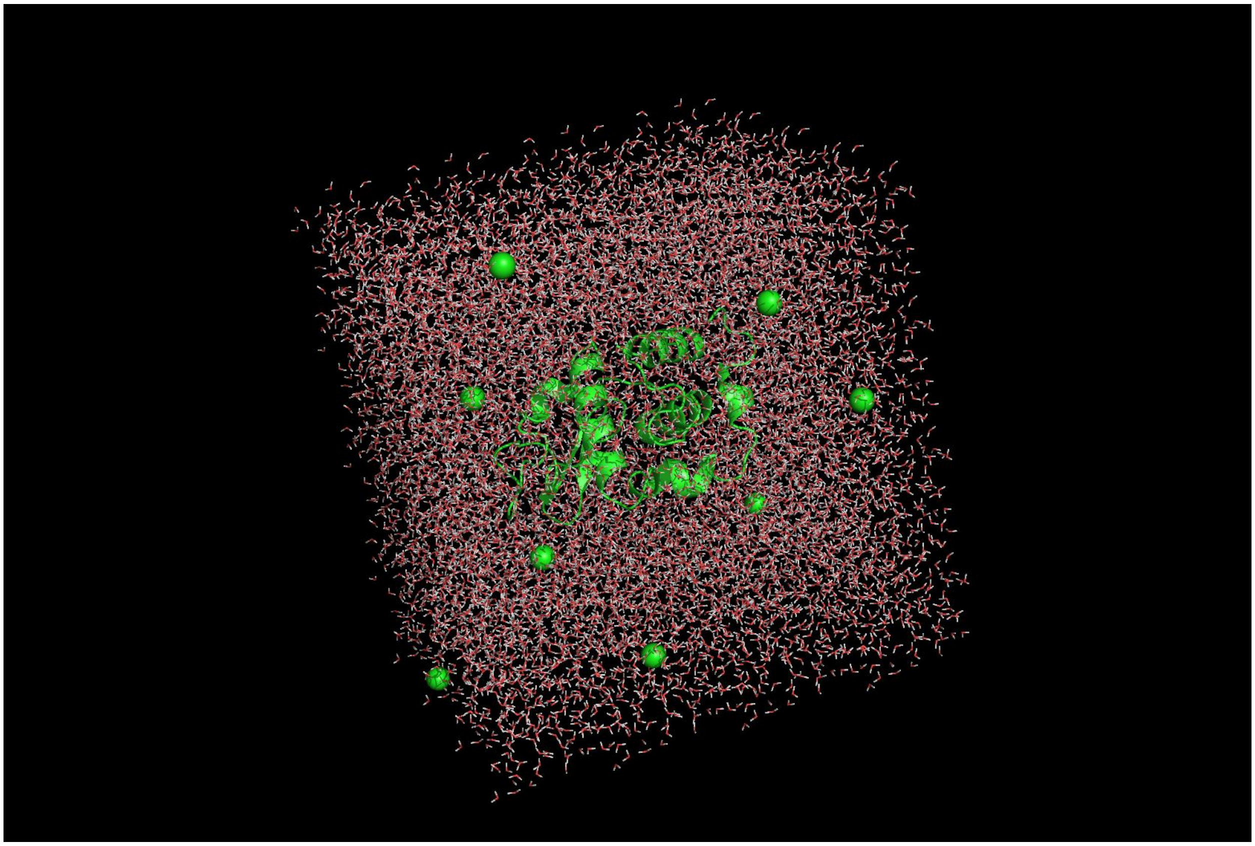


Figure: (Ion Distribution in Solvation Box) shows a uniform distribution of ions around the protein, ensuring the system's neutrality and stability. This step was essential for preventing artifacts in electrostatic interactions during MD simulation.

6. Conclusion

The MD simulation successfully captured the dynamic behavior and stability of the protein within a solvated and neutralized environment. The stable temperature, pressure, and density profiles confirm the adequate equilibration of the system. At the same time, RMSD and Radius of Gyration analyses indicate a stable and compact protein structure throughout the simulation. Future studies may delve deeper into the conformational transitions observed in this system to further understand their relevance to the protein’s biological functions.