Experiment-6: Molecular Dynamics Simulation : Protein Dynamics using Gromacs

- **AIM:** To analyze the dynamic behavior of a molecule using simulation technique using Gromacs.
- **TOOL:**

PDB: http://www.rcsb.org

Gromacs 5.1.1

Xmgrace

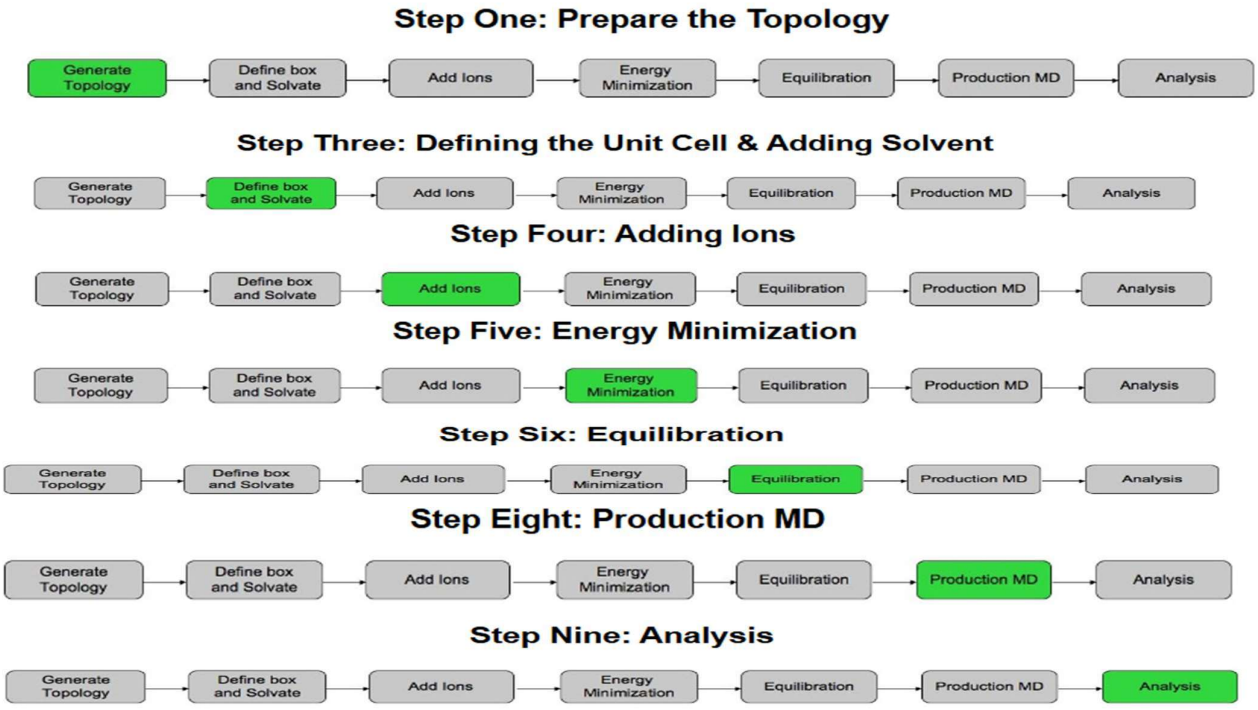
Files: ions.mdp, mininm.mdp, nvt.mdp, npt.mdp

GROMACS (GROningen MAchine for Chemical Simulations) is a molecular dynamics simulation package originally developed in the University of Groningen, now maintained and extended at different places, including the University of Uppsala, University of Stockholm and the Max Planck Institute for Polymer Research. GROMACS is open source software released under the GPL.

INPUT PROTEIIN:-

PDB ID : 1EFW
Protein Name : Crystal structure of aspartyl-tRNA synthetase
Classification:
LIGASE/RNA

- **METHODOLOGY:**
  - **STEPS:**



1. Retrieve of Protein:

v. Go to PDB database and search for 1EFW protein structure
vi. Download the structure in .pdb file format
vii. Open PyMOL and upload the protein file
viii. Extract the Chain A and save this as .pdb file

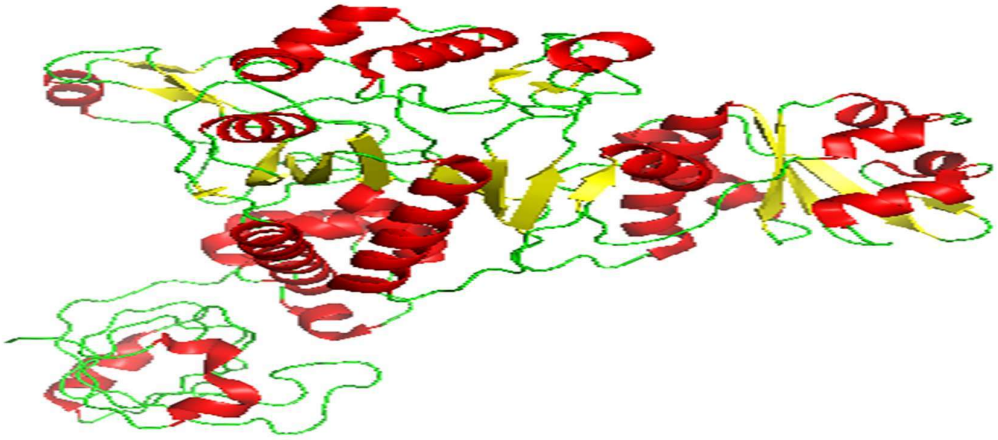


Figure 40: 1EFW

2. Dowanload .mdp file:

i. Go to MD simulation tutorial and download **ions.mdp, mininm.mdp, nvt.mdp, npt.mdp** files.
ii. Create an folder in desktop and save these files into the folder.

➢ COMAND:

3. Generate topology:

i. **grep -v HOH Protein.pdb > clean.pdb**
ii. **grep -v HETATM 1efwA.pdb > ab.pdb**
iii. **gmx pdb2gmx -f clean.pdb -o output.gro -water spce -ignh**

Then, it will be prompted to select the force field: OPLS-AA/L

```
 1: AMBER03 protein, nucleic AMBER94 (Duan et al., J. Comp. Chem. 24, 1999-2012, 2003)
 2: AMBER94 force field (Cornell et al., JACS 117, 5179-5197, 1995)
 3: AMBER96 protein, nucleic AMBER94 (Kollman et al., Acc. Chem. Res. 29, 461-469, 1996)
 4: AMBER99 protein, nucleic AMBER94 (Wang et al., J. Comp. Chem. 21, 1049-1074, 2000)
 5: AMBER99SB protein, nucleic AMBER94 (Hornak et al., Proteins 65, 712-725, 2006)
 6: AMBER99SB-ILDN protein, nucleic AMBER94 (Lindorff-Larsen et al., Proteins 78, 1950-58, 2010)
 7: AMBERGS force field (Garcia & Sanbonmatsu, PNAS 99, 2782-2787, 2002)
 8: CHARMM27 all-atom force field (CHARM22 plus CMAP for proteins)
 9: GROMOS96 43a1 force field
10: GROMOS96 43a2 force field (improved alkane dihedrals)
11: GROMOS96 45a3 force field (Schuler JCC 2001 22 1205)
12: GROMOS96 53a5 force field (JCC 2004 vol 25 pag 1656)
13: GROMOS96 53a6 force field (JCC 2004 vol 25 pag 1656)
14: GROMOS96 54a7 force field (Eur. Biophys. J. (2011), 40,, 843-856, DOI: 10.1007/s00249-011-0700-9)
15: OPLS-AA/L all-atom force field (2001 aminoacid dihedrals)
15

Using the Oplsaa force field in directory oplsaa.ff
```

Figure 41: Force Fields

The purpose of pdb2gmx is to generate three files:

- The topology for the molecule: topol.top
- A position restraint file: posre.itp
- A post-processed structure file: protein.gro

4. Define box:

i. gmx editconf -f output.gro -o box.gro -c -d 1.0 -bt cubic

Define the box dimensions using the editconf module.

The above command centres the protein in the box (-c), and places it at least 1.0 nm from the box edge (-d 1.0). The box type is defined as a cube (-bt cubic).

5. Add solvent:

i. gmx solvate -cp box.gro -cs spc216.gro -o solv.gro -p topol.top

Fill the box with water using the solvate module (formerly called genbox).

Here, it is using spc216.gro, which is a generic equilibrated 3-point solvent model.

The output is called solv.gro, and we tell solvate the name of the topology file (topol.top) so it can be modified.

6. Adding ions:

i. gmx grompp -f ions.mdp -c solv.gro -p topol.top -o ions.tpr

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

The tool “genion” reads through the topology and replace water molecules with the ions that the user specifies. To produce a .tpr file with grompp, it will need an additional input file, with the extension .mdp ; grompp will assemble the parameters specified in the .mdp file with the coordinates and topology information to generate a
.tpr file.
When prompted, choose group **13 "SOL"** for embedding ions.
Here, the structure/state file (-s) as input, generate a .gro file as output (-o), process the topology (-p) to reflect the removal of water molecules and addition of ions, define positive and negative ion names (-pname and -nname, respectively), and tell genion to add only the ions necessary to neutralize the net charge on the protein by adding the correct number of negative ions.

7. Energy minimization:

i. gmx grompp -f minim.mdp -c ion_gen.gro -p topol.top -o em.tpr

The structure is relaxed through a process called energy minimization (EM).

ii. gmx mdrun -v -deffnm em

This command invokes mdrun to carry out the EM.
It will generate the following files: em.edr, em.log, em.trr, em.gro

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

iv. xmgrace potential.xvg

The em.edr file contains all of the energy terms that GROMACS collects during EM. One can analyse any .edr file using the GROMACS energy module.

8. Equilibration step:

(a) The first phase is conducted under an NVT ensemble : "isothermal-isochoric" or "canonical" ensemble.

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

ii. gmx mdrun -deffnm nvt -v

Analyse the temperature progression, again using energy:

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

Type "16 0" at the prompt to select the temperature of the system and exit.

iv. xmgrace temperature.xvg

(b) The second phase is conducted under an NPT ensemble :"isothermal-isobaric" ensemble

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

Including the -t flag to include the checkpoint file from the NVT equilibration; this file contains all the necessary state variables to continue our simulation.

Invoke the mdrun:

vi. gmx mdrun -deffnm npt -v

Analyse the pressure progression:

vii. gmx energy -f npt.edr -o pressure.xvg

Check the graph:

Viii. xmgrace pressure.xvg

(c) Density graph:

ix. gmx energy -f npt.edr -o density.xvg

Enter "24 0" at the prompt.

X. xmgrace density.xvg

9. Production md:

Upon completion of the two equilibration phases, the system is now well-equilibrated at the desired temperature and pressure. It is now ready to release the position restraints and run production MD for data collection.

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

Execute mdrun:

ii. gmx mdrun -deffnm md -v

9. Analysis:

(a) The first is trjconv, which is used as a post-processing tool to strip out coordinates, correct for periodicity, or man alter the trajectory (time units, frame frequency, etc).

i. gmx trjconv -s md.tpr -f md.xtc -o md_noPBC.xtc -pbc mol -ur center

Select 1 ("Protein") as the group to be centred and 0 ("System") for output.

(b) **RMSD: Looks for structure stability**

i. gmx rms -s md.tpr -f md_noPBC.xtc -o rmsd.xvg -tu ns

Choose 4 ("Backbone") for both the least-squares fit and the group for RMSD calculation. The -tu flag will output the results in terms of ns, even though the trajectory was written in ps.

Plotted together, results look something like:

ii. **xmgrace rmsd.xvg rmsd1.xvg**

(c) **RMSF:**

iii. gmx rmsf -res -s md.tpr -f md_noPBC.xtc -o rmsf_res.xvg

‘ Choose 3 ("Cα") for both

iii. **xmgrace rmsf_res.xvg**

(d) Radius of gyration:

The radius of gyration of a protein is a measure of its compactness. If a protein is stably folded will likely maintain a relatively steady value of Rg. If a protein unfolds, its Rg will change over time.

i. gmx gyrate -s md.tpr -f md_noPBC.xtc -o gyrate.xvg

(Choose 1 for Protein)

ii. xmgrace gyrate.xvg

(e) H-Bond:

i. **gmx hbond -f md_noPBC.xtc -s md.tpr -num h.xvg -tu ns**

(Select 1(Protein) for both)

ii. **xmgrace h.xvg**

(f) SASA:

i. gmx sasa -f md_noPBC.xtc -s md.tpr -o sasa.xvg -tu ns

(Select 1(Protein) for both)

iii. xmgrace sasa.xvg

- # RESULT:

1.Energy Minimization: Steepest Descent

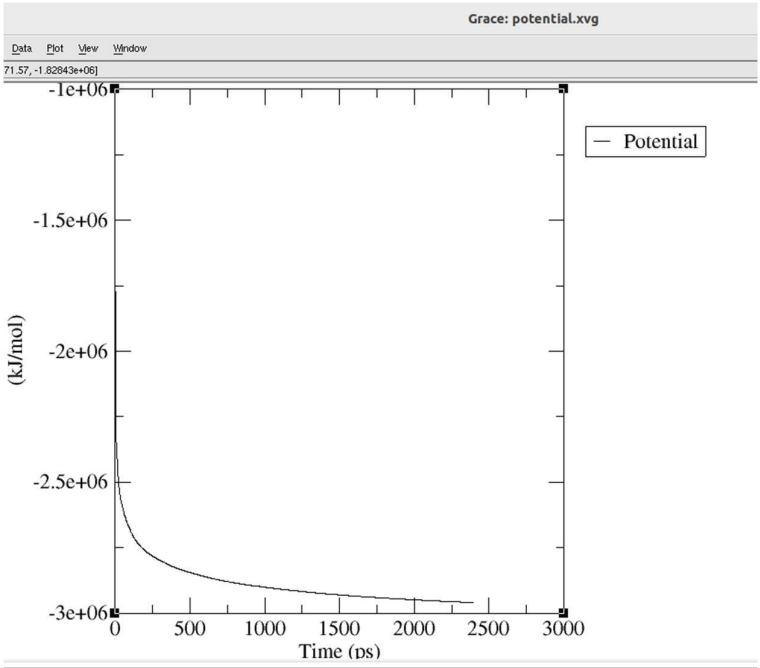


Figure 42: Potential Energy Plot : Potential Energy (KJ/mol) VS Energy Minimization Step

- **INTERPRETATION:**

x-axis: Potential Energy in KJoule/Mol

y-axis: Time in Pico second

- The Plot demonstrating the nice, steady convergence of Epot. The potential energy of the system exponentially decreases, **near 2500 ps the energy of the system is minimum about -3e+06**. The system settles into a stable configuration.

2. Temperature: NVT Equilibration

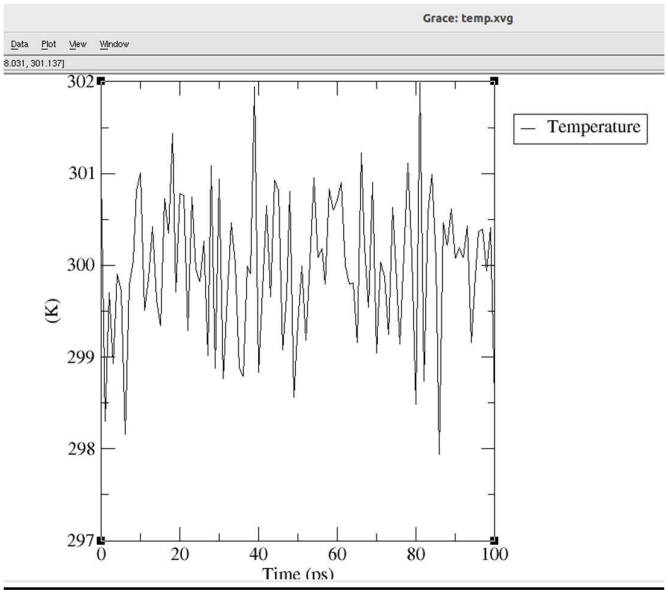


Figure 43: Temperature (K) VS Time (ps) Plot

- **INTERPRETATION:**

x-axis: Temperature in Kelvin

y-axis: Time in Pico second

- From the Plot, it is clear that the temperature of the system quickly reaches the **target value (300K),** remains stable over the remainder of the equilibration.

3.Pressure: NPT Equilibration

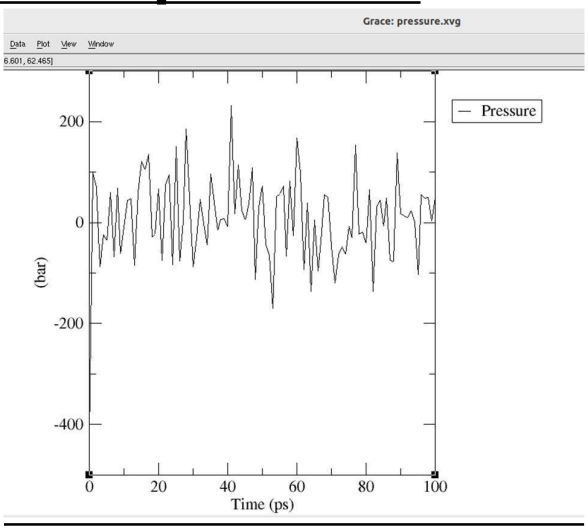


Figure 44: Pressure (bar) VS Time (ps) Plot

- **INTERPRETATION:**

x-axis: Pressure in bar

y-axis: Time in Pico second

- The pressure value fluctuates widely over the course of the 100-ps equilibration phase. It indicates a **well equilibrate system**. Fluctuation reflects thermal motion and system dynamics

4.Density: NPT Equilibration

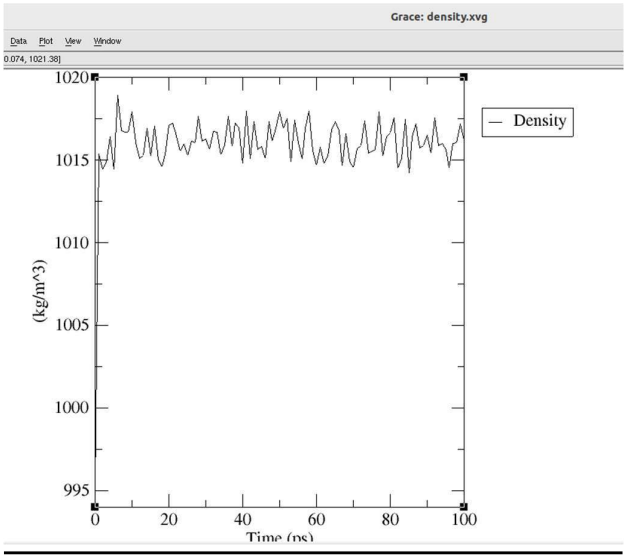


Figure 45: Density (Kg/ m³) VS Time (ps) Plot

- **INTERPRETATION:**

x-axis: Density in Kg/m³

y-axis: Time in Pico second

- The **average value** over the course of **100ps is 1016Kg/m³.** The parameters for the **SPC/E water model** closely replicate experimental values of water. The density values are very stable over the time, indicating that the system is well-equilibrated now with respect to pressure and density.

5.Root Mean Square Deviation: Backbone

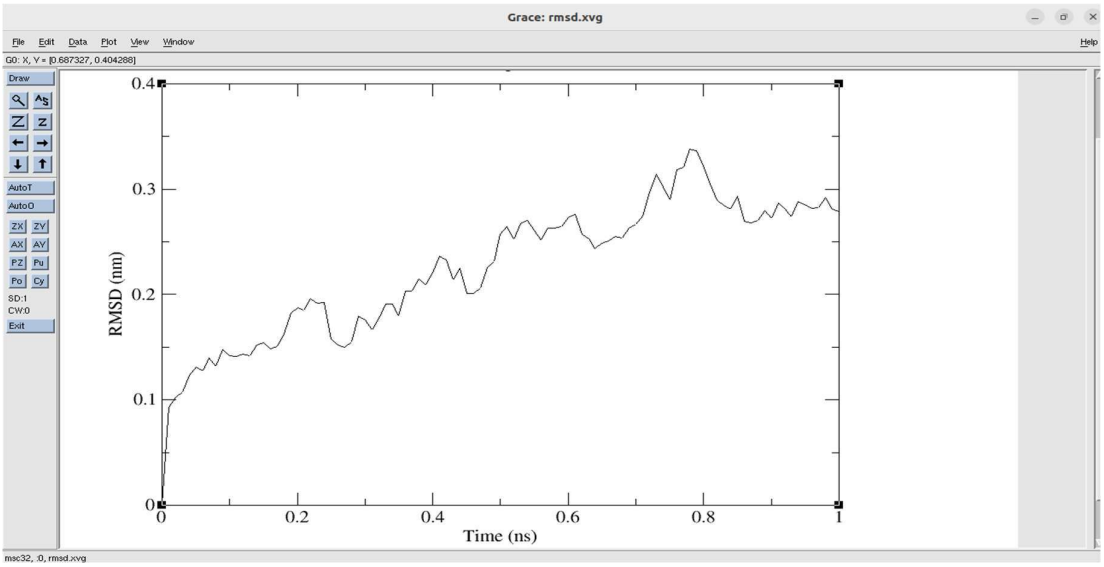


Figure 46: RMSD (nm) VS Time (ns) Plot

- **INTERPRETATION:**

x-axis: Root Mean Square Deviation in nm

y-axis: Time in nano second

- **RMSD:** The RMSD of atomic positions is the measure of the average distance between the atoms of superimposed proteins.
- The smaller the differences, the more spatially equivalents the compared structure, whereas more distantly related structures are defined comparatively greater RMSD value.

The RMSD levels off to ~0.1 nm, **indicating that the structure is more deviated from the reference structure.**

6. Root Mean Square Fluctuation: Cα Residue

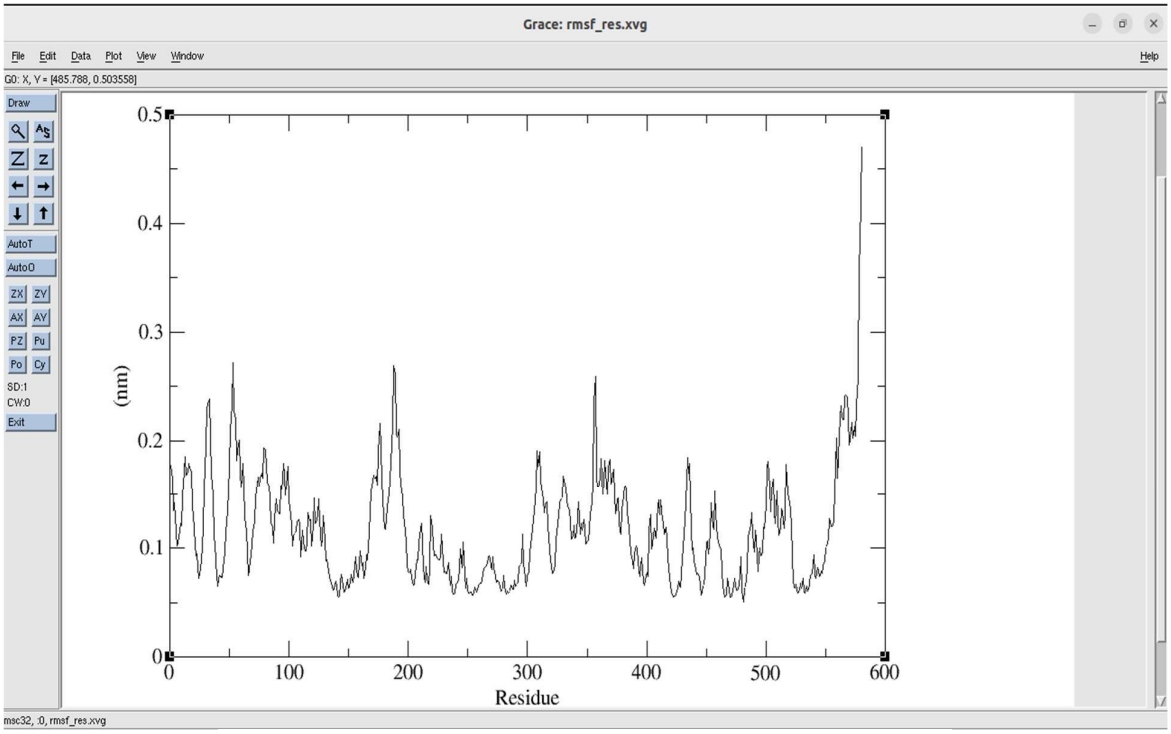


Figure 47: RMSF (nm) VS Residue Plot

- **INTERPRETATION:**

x-axis: Root Mean Square Deviation in nm

y-axis: Number of Residues

- **RMSF:** The RMSF calculate the fluctuation of Cα atoms in residues of a protein in comparison with the respective average structure throughout the simulation.
- Increased residual RMSF value indicate the instability in protein backbone that refers to unfolded structure.
- The C and N terminal end of the protein have the highest RMSF value .

50 to 60 numbered residues are more fluctuating that has **RMSF value near 0.3 nm**, these residues may be **present in an unfolded region.**

7. Radius of Gyration (Rg): Unrestrained MD

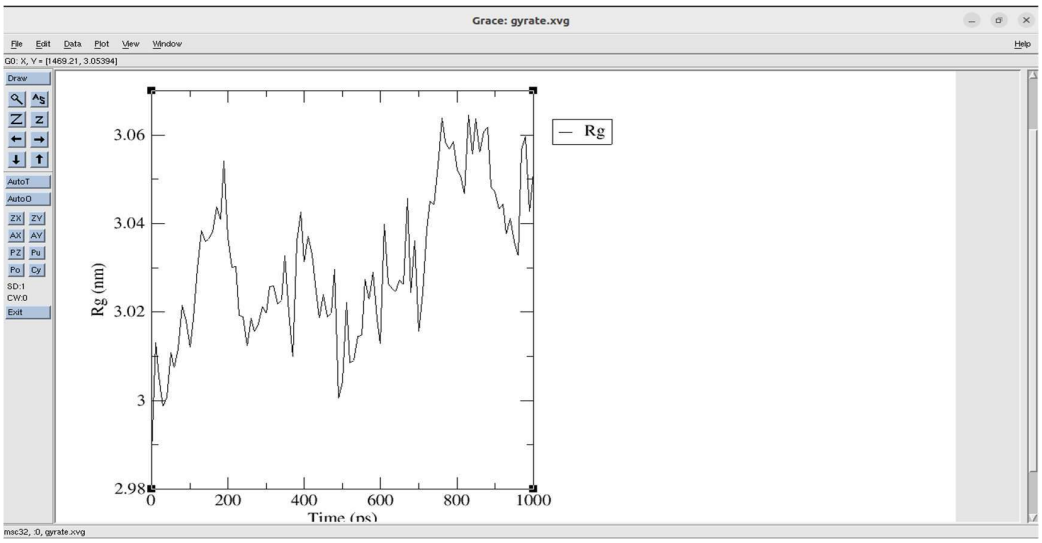


Figure 48: Radius of Gyration (nm) VS Time (ps) Plot

- **INTERPRETATION:**

x-axis: Radius of Gyration in nm

y-axis: Time in Pico second

- **Radius of Gyration:** The Rg of a protein is a measure of its compactness.
- If a protein is folded it maintains a steady value of Rg, on the other hand if a protein is present in unfolded conformation its Rg will change over the time.

In this plot, **Rg value is more** that indicating that **the protein is less compact (unfolded) and less stable.**

8. Hydrogen Bond:

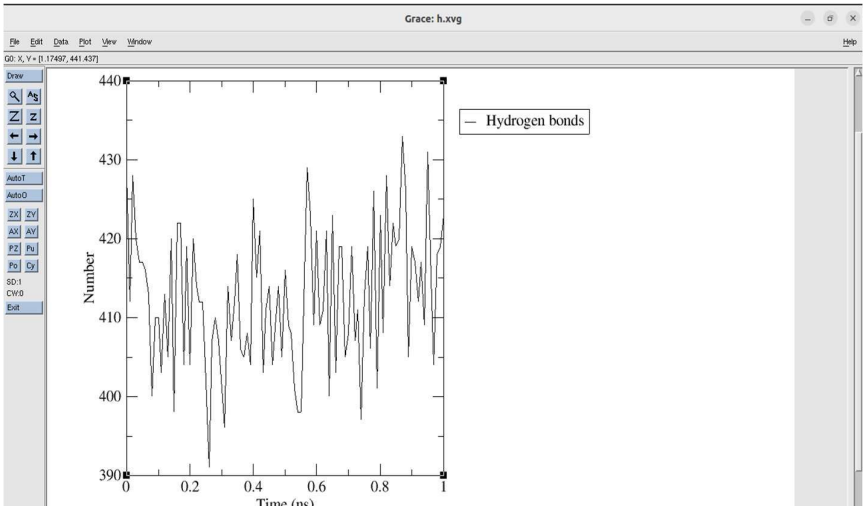


Figure 49: Number of Hydrogen Bond VS Time (ns) Plot

- **INTERPRETATION:**

x-axis: Number of Hydrogen Bond

y-axis: Time in nano second

- **H-Bond:** Maintain the functional 3D conformation of the protein

High H-bond indicating the stable molecular interaction.

9. Solvent Accessible Surface Area:

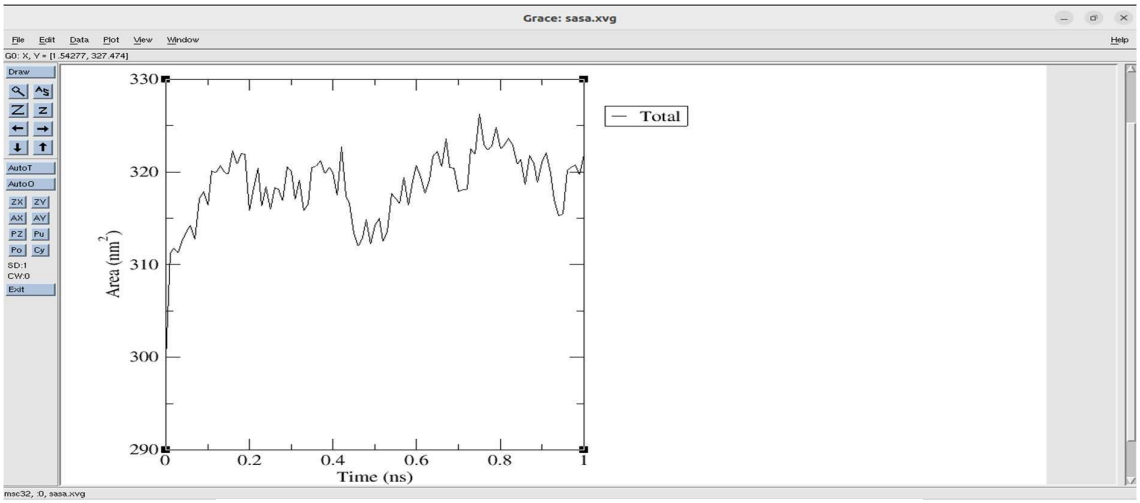


Figure 50: SASA(nm²) VS Time (ns) Plot

- **INTERPRETATION:**

x-axis: Number of Hydrogen Bond

y-axis: Time in nano second

- **Solvent Accessible Surface Area:** The area of protein that is exposed enough to make interactions with the neighbouring solvent molecules .

Increased SASA indicates that the protein is present in unfolded conformation.