BT5420: Computer Simulations of Biomolecular Systems
July – Nov 2023 Semester
GROMACS Simulation Project-Report

Sapna R
BS20B032

All the simulation files used for the purpose of the assignment are in this folder

RESEARCH PAPER CHOSEN :
Title: Structural analysis of Ca2+ dependent and Ca2+ independent type II antifreeze proteins: A comparative molecular dynamics simulation study

Link::https://doi.org/10.1016/j.jmgm.2012.05.004

OBJECTIVE:

- The aim of this paper is to elucidate a possible role played by Ca2+ ion in transferring stability to type II antifreeze proteins.
- For this purpose they have chosen 2 proteins: One is a Ca2+ dependent psychrophilic type II antifreeze protein derived from herring (*Clupea harengus*) refereed to in the paper as hAFP(PDB Code:2PY2).Another is a Ca2+ independent type II antifreeze protein derived from ong snout poacher (Brachyopsis rostratus) referred to in the paper as lpAFP(PDB Code 2ZIB)
- Thus the aim of the paper is to leverage the power of molecular dynamic simulations to make a comparative study on the aspects of secondary structure content, molecular flexibility, solvent accessibility, intra molecular hydrogen bonds and protein–solvent interaction at 5 different temperatures (277K,298K, 377K, 423K, 473K).
- The study also aims to monitor the thermal unfolding pathways of the protein.
- The results obtained from molecular dynamic simulations can be used to understand the molecular basis of thermostability for 2 structurally similar proteins which function in the presence and absence of Ca2+.

REASONS FOR CHOOSING THIS PAPER:

- I find the class of antifreeze proteins a very interesting species to study.It is expressed by psychrophilic organisms that live in extremely cold environments.
- It is said that AntiFreeze Proteins or the SFPs depress the freezing point of blood and body fluids to that below the freezing point of sea by binding and inhibiting the growth of ice crystals.
- AntiFreeze proteins have enormous potential for commercial applications.
- Their potential applications range from protecting economically important fish like Atlantic Salmon and goldfish from frost in low temperatures via genetic engineering, to preserving and improving the quality of frozen food etc,

cryoprotective agents for organ and cell cryopreservation, as chemical adjuvants to cancer cryosurgery.

- Since molecular dynamic simulations is a rapidly growing powerful tool in understanding the molecular basis for thermostability, it has been used in this paper to perform a comparative analysis of the thermostability and flexibility of 2 different Antifreeze proteins.
- I found this paper a very fascinating read with a detailed description of the MD simulations making it easier for me to reproduce a few of the results.
- I had earlier chosen another paper titled “Molecular Dynamics Simulations of Protein-Tyrosine Phosphatase 1B. II. Substrate-Enzyme Interactions and Dynamics”. However, since the details for MD simulations were not too explicit in this paper, I could not go ahead with this paper.

RESULTS CHOSEN TO REPRODUCE:

- I have chosen to reproduce the following analyses: RMSF,Distance between Ca2+ and Asp111O,RMSD vs Rg.
- RMSF is a very important metric used to evaluate and compare the relative root mean square fluctuations experienced by each atom/residue over the period of production run.
- This gives insight into which residue underwent maximum fluctuation relative to the starting structure or about the averaged structure and which residue remained relatively stationary over the course of the simulation.
- Studying the RMSF plot gives some insight on what residues make the active site of the protein.
- It is observed that regions corresponding to maximum flexibility is that of the poorly organized loop region or the termini of the protein. Residues belonging to secondary structures exhibit low mobility.
- Since I have carried out the simulation at 473 K ,most of the residues exhibit elevated flexibility as it is given in the paper that this is usually the unfolding temperature of the protein.
- It has been found through sequence analysis that Ca2+ ions in hAFP forms coordinate bond interaction with Gln89 Oε1, Asp91 Oδ2, Glu96 Oε1, Asn110 Oδ1 and Asp111 O and Oδ1. However, this coordinate bond does not remain intact above 373k.
- This has been verified by analyzing the distance between Ca2+ and Asp111 O at different temperatures in the paper.
- It was distinctly observed that the curves corresponding to temperatures above 373k had significantly greater values for distance between Ca2+ ion and Asp 111O.
- Hence, I have carried out the analysis too for my 1 ns production run at 473k to verify the same.
- Radius of gyration characterizes the overall compactness of the structure based on the overall distribution of the protein mass. With increase in Rg, we can understand that the protein is uncoiling from its native form.
- RMSD analysis gives some insight on how the coordinates of the atoms in the backbone or side chains vary with the starting structure at every instant. The smaller the RMSD, the more similar the two structures .

- Thus a 2 dimensional plot between RMSD and Radius of gyration gives some insight on the correlation between the simultaneous collapse of the chain and the protein adopting a denatured conformation.
- The plot between RMSD vs Rg obtained in the paper shows a simultaneous increase in Rg and RMSD in hAFP at 473k representing 2 state kinetics.
- For these reasons , I chose to reproduce the result for RMSD vs Rg if such a correlation and diagonal movement can be observed for a 1 ns production run as well.

SIMULATION DETAILS AS GIVEN IN THE ARTICLE for hAFP at 473 K

- MD Simulations were carried out using GROMACS 4 and GROMOS96 43a1 force fields implemented on LINUX architecture.
- The starting structure of the Ca2+ dependent AFP derived from herring, C. *harengus*, was obtained from PDB (PDB Code:2PY2) and it was composed of 127 residues.
- The hetero atoms other than Ca2+ ions were removed.
- The starting structure was immersed in a Triclinic box of dimensions 5.74 x 5.6x 6.21 nm (473 k simulation).
- The system composed of 1265 protein atoms and 1 Ca2+ ion was solvated with SPC water molecules (5803).
- 5 Na+ ions were added to replace water molecules and neutralize the charge of the system.
- All protein atoms were placed at least 1.0 nm from the box edges (Periodic Boundary Conditions)
- The system was minimized by energy minimization using the steepest descent algorithm using 2000 steps.
- The minimized systems for hAFP were equilibrated for 50 ps inorder to relax the solvent by position restrained molecular dynamics (LINCS scheme)
- Electrostatic interactions were calculated using Particle Mesh Ewald Algorithm.
- The cutoff for short range electrostatics and vanderwaal forces were set to 0.9nm
- Periodic Boundary conditions combined with minimum image convention were used under isothermal, isobaric conditions using Berendson Algorithm.
- The relaxation times for isothermal and isobaric conditions were ser 0.1 and 0.5 ps respectively.
- The non-bonded pair list was updated every 10 steps and conformations were stored every 2 ps.
- The trajectory analysis was performed for 10ns in the paper.

MODIFICATIONS MADE IN THE MDP FILES

ions - Notepad

File Edit Format View Help

```
; ions.mdp - used as input into grompp to generate ions.tpr
; Parameters describing what to do, when to stop and what to save
integrator  = steep         ; Algorithm (steep = steepest descent minimization)
emtol       = 100       ; Stop minimization when the maximum force < 100 kJ/mol/nm
emstep      = 0.01          ; Minimization step size
nsteps      = 2000      ; Maximum number of (minimization) steps to perform

; Parameters describing how to find the neighbors of each atom and how to calculate the interactions
nstlist         = 1         ; Frequency to update the neighbor list and long range forces
cutoff-scheme   = Verlet    ; Buffered neighbor searching
ns_type         = grid      ; Method to determine neighbor list (simple, grid)
coulombtype     = cutoff    ; Treatment of long range electrostatic interactions
vdwtype         = cutoff    ; treatment of long range vanderwaal interactions
rcoulomb        = 0.9       ; Short-range electrostatic cut-off
rvdw            = 0.9       ; Short-range Van der Waals cut-off
pbc             = xyz       ; Periodic Boundary Conditions in all 3 dimension
```

ions.MDP

← The number of steps was changed to 2000 as given in the paper. The cutoff for energy difference was changed to 100kJ/mol to ensure minimisation does not complete before 2000 steps

← The cutoff for Short range interactions for electrostatic and van der waal forces were changed to 0.9 nm

minim - Notepad

File Edit Format View Help

```
; minim.mdp - used as input into grompp to generate em.tpr
; Parameters describing what to do, when to stop and what to save
integrator  = steep         ; Algorithm (steep = steepest descent minimization)
emtol       = 100.0        ; Stop minimization when the maximum force < 100.0 kJ/mol/nm
emstep      = 0.01          ; Minimization step size
nsteps      = 2000          ; Maximum number of (minimization) steps to perform

; Parameters describing how to find the neighbors of each atom and how to calculate the interactions
nstlist         = 1         ; Frequency to update the neighbor list and long range forces
cutoff-scheme   = Verlet    ; Buffered neighbor searching
ns_type         = grid      ; Method to determine neighbor list (simple, grid)
coulombtype     = PME       ; Treatment of long range electrostatic interactions
rcoulomb        = 0.9       ; Short-range electrostatic cut-off
rvdw            = 0.9       ; Short-range Van der Waals cut-off
pbc             = xyz       ; Periodic Boundary Conditions in all 3 dimensions
```

minim.MDP

← The number of steps was changed to 2000 as given in the paper. The cutoff for energy difference was changed to 100kJ/mol to ensure minimisation does not complete before 2000 steps

← The cutoff for Short range interactions for electrostatic and van der waal forces were changed to 0.9 nm

nvt - Notepad

File Edit Format View Help

```
title                    = OPLS Lysozyme NVT equilibration
define                   = -DPOSRES  ; position restrain the protein
; Run parameters
integrator               = md        ; leap-frog integrator
nsteps                   = 25000     ; 2 * 25000 = 50 ps
dt                       = 0.002     ; 2 fs
; Output control
nstxout                  = 1000       ; save coordinates every 2.0 ps
nstvout                  = 1000       ; save velocities every 2.0 ps
nstenergy                = 1000       ; save energies every 2.0 ps
nstlog                   = 1000       ; update log file every 2.0 ps
```

```
; Nonbonded settings
cutoff-scheme           = Verlet    ; Buffered neighbor searching
ns_type                 = grid      ; search neighboring grid cells
nstlist                 = 10        ; 20 fs, largely irrelevant with Verlet
rcoulomb                = 0.9       ; short-range electrostatic cutoff (in nm)
rvdw                    = 0.9       ; short-range van der Waals cutoff (in nm)
DispCorr                = EnerPres  ; account for cut-off vdW scheme
```

```
; Temperature coupling is on
tcoupl                  = V-rescale             ; modified Berendsen thermostat
tc-grps                 = Protein Non-Protein   ; two coupling groups - more accurate
tau_t                   = 0.1     0.1           ; time constant, in ps
ref_t                   = 473     473           ; reference temperature, one for each group, in K
```

```
; Velocity generation
gen_vel                 = yes       ; assign velocities from Maxwell distribution
gen_temp                = 473       ; temperature for Maxwell distribution
gen_seed                = -1        ; generate a random seed
```

nvt.MDP

← The number of steps was changed to 25000 as given in the paper for equilibrating the system for 50 ps as given in the paper .

← The following parameters under Output control was changed to 1000 to save conformations ever 2 ps as given in the paper

← The cutoff for Short range interactions for electrostatic and van der waal forces were changed to 0.9 nm

← The tau_t for temperature coupling is 0.1 ps. The reference temperature for each group was changed to 473k.

← The gen_temp for maxwell distribution was changed to 473k.

npt - Notepad

File Edit Format View Help

```
title                    = OPLS Lysozyme NPT equilibration
define                   = -DPOSRES  ; position restrain the protein
; Run parameters
integrator               = md        ; leap-frog integrator
nsteps                   = 25000     ; 2 * 25000 = 50 ps
dt                       = 0.002     ; 2 fs
; Output control
nstxout                  = 1000       ; save coordinates every 2.0 ps
nstvout                  = 1000       ; save velocities every 2.0 ps
nstenergy                = 1000      ; save energies every 2.0 ps
nstlog                   = 1000      ; update log file every 2.0 ps
```

npt.MDP

← The number of steps was changed to 25000 as given in the paper for equilibrating the system for 50 ps as given in the paper .

← The following parameters under Output control was changed to 1000 to save conformations ever 2 p s as given in the paper

```
; Nonbonded settings
cutoff-scheme            = Verlet    ; Buffered neighbor searching
ns_type                  = grid      ; search neighboring grid cells
nstlist                  = 10        ; 20 fs, largely irrelevant with Verlet scheme
rcoulomb                 = 0.9       ; short-range electrostatic cutoff (in nm)
rvdw                     = 0.9       ; short-range van der Waals cutoff (in nm)
DispCorr                 = EnerPres  ; account for cut-off vdW scheme
```

← The cutoff for Short range interactions for electrostatic and van der waal forces were changed to 0.9 nm

```
; Temperature coupling is on
tcoupl                   = V-rescale             ; modified Berendsen thermostat
tc-grps                  = Protein Non-Protein   ; two coupling groups - more accurate
tau_t                    = 0.1     0.1           ; time constant, in ps
ref_t                    = 473     473           ; reference temperature, one for each group, in K
; Pressure coupling is on
pcoupl                   = Parrinello-Rahman     ; Pressure coupling on in NPT
pcoupltype               = isotropic             ; uniform scaling of box vectors
tau_p                    = 0.5                   ; time constant, in ps
ref_p                    = 1.0                   ; reference pressure, in bar
compressibility          = 4.5e-5                ; isothermal compressibility of water, bar^-1
```

← The tau_t for temperature coupling is 0.1 ps. The reference temperature for each group was changed to 473k.

← The tau_t for pressure coupling is changed to 0.5 ps as given in the paper.

```
title                    = OPLS Lysozyme NPT equilibration
; Run parameters
integrator               = md        ; leap-frog integrator
nsteps                   = 500000    ; 2 * 500000 = 1000 ps (1 ns)
dt                       = 0.002     ; 2 fs
; Output control
nstxout                  = 0         ; suppress bulky .trr file by specifying
nstvout                  = 0         ; 0 for output frequency of nstxout,
nstfout                  = 0         ; nstvout, and nstfout
nstenergy                = 1000      ; save energies every 2.0 ps
nstlog                   = 1000      ; update log file every 2.0 ps
nstxout-compressed       = 1000      ; save compressed coordinates every 2.0 ps
compressed-x-grps        = System    ; save the whole system
```

md.MDP

← The number of steps was changed to 500000 for carrying out the production run for 1ns

← The following parameters under Output control was changed to 1000 to save conformations ever 2 p s as given in the paper

```
; Neighborsearching
cutoff-scheme            = Verlet    ; Buffered neighbor searching
ns_type                  = grid      ; search neighboring grid cells
nstlist                  = 10        ; 20 fs, largely irrelevant with Verlet scheme
rcoulomb                 = 0.9       ; short-range electrostatic cutoff (in nm)
rvdw                     = 0.9       ; short-range van der Waals cutoff (in nm)
```

← The cutoff for Short range interactions for electrostatic and van der waal forces were changed to 0.9 nm

```
; Temperature coupling is on
tcoupl                   = V-rescale             ; modified Berendsen thermostat
tc-grps                  = Protein Non-Protein   ; two coupling groups - more accurate
tau_t                    = 0.1     0.1           ; time constant, in ps
ref_t                    = 473     473           ; reference temperature, one for each group, in K
; Pressure coupling is on
pcoupl                   = Parrinello-Rahman     ; Pressure coupling on in NPT
pcoupltype               = isotropic             ; uniform scaling of box vectors
tau_p                    = 0.5                   ; time constant, in ps
ref_p                    = 1.0                   ; reference pressure, in bar
compressibility          = 4.5e-5                ; isothermal compressibility of water, bar^-1
```

← The tau_t for temperature coupling is 0.1 ps. The reference temperature for each group was changed to 473k.

← The tau_t for pressure coupling is changed to 0.5 ps as given in the paper.

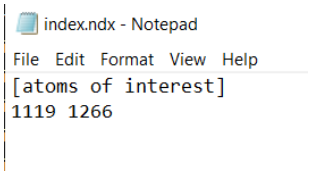
VMD MOVIE OF THE 1 NS PRODUCTION RUN

The link for the vmd movie of the 1nm production run including only frames every 100 ps.

COMPLETE DESCRIPTION OF COMPUTATION OF THE PROPERTIES OF INTEREST

- I have obtained the analyses for the following properties: RMSF, Radius of gyration, RMSD vs Rg for hAFP(Ca2+ dependent type II AntiFreeze protein derived from *Clupea harengus* at 473k.
- I had followed the simulation details I have described above and made the following modifications in mdp files I have pasted above.
- I obtained the starting structure of hAFP from PDB. PDB Code is 2PY2.
- The PDB file had 6 chains each containing 127 residues and one Ca2+ ions and few water molecules.
- I had removed all the chains except chain A along with their corresponding Ca2+ ions and the water molecules present using pymol.
- I then used the pdb2gmx tool in linux terminal to create the topology file ,position restraint file and the gro file and added hydrogen atoms.I specified the water model as spc. After this stage the system had 1266 atoms: 1265 protein and 1 Ca2+ ions.
    - **gmx pdb2gmx -f 2py2edit.pdb -o 2py2_processed.gro -ignh -water spc**
    - I chose option 9 Gromos96 43a1 force field
- I then defined a triclinc box of dimensions 5.74x5.6x6.21 nm^3 to immerse the system using the following command and defined the PBC conditions as 1.0 nm from the edges of the box.
    - **gmx editconf -f 2py2_processed.gro -o 2py2_newbox.gro -c -d 1.0 -bt triclinic -box 5.74,5.6.6.21**
- I solvated the protein using 5781 water molecules.. Here ,there is a slight deviation from the paper. While the paper had used 5803 water molecules, I could not add more than 5781 water molecules because of solvent-solvent overlap.
    - **gmx solvate -cp 2py2_newbox.gro -cs spc216.gro -o 2py2_solv.gro -p topol.top -maxsol 5803**
- I neutralized the system by adding 5 Na+ counter ions using the genion tool.
    - **gmx grompp -f ions.mdp -c 2py2_solv.gro -p topol.top -o ions.tpr**
    - **gmx genion -s ions.tpr -o protein_solv_ions.gro -p topol.top -pname NA -nname CL -neutral**
    - I selected water as the system to be replaced with na+ ions.
- I minimized the energy of the system after making the indicated changes in minim.mdp file using the below command
    - **gmx grompp -f minim.mdp -c protein_solv_ions.gro -p topol.top -o em.tpr**
    - **gmx mdrun -v -deffnm em**
- I then carried out NVT equilibration after making the indicated changes in the nvt.MDP file.
    - **gmx grompp -f nvt.mdp -c em.gro -r em.gro -p topol.top -o nvt.tpr**
    - **gmx mdrun -deffnm nvt**
- I then carried out NPT equilibration after making the indicated changes in the npt.MDP file.
    - **gmx grompp -f npt.mdp -c nvt.gro -r nvt.gro -t nvt.cpt -p topol.top -o npt.tpr**
    - **gmx mdrun -deffnm npt**

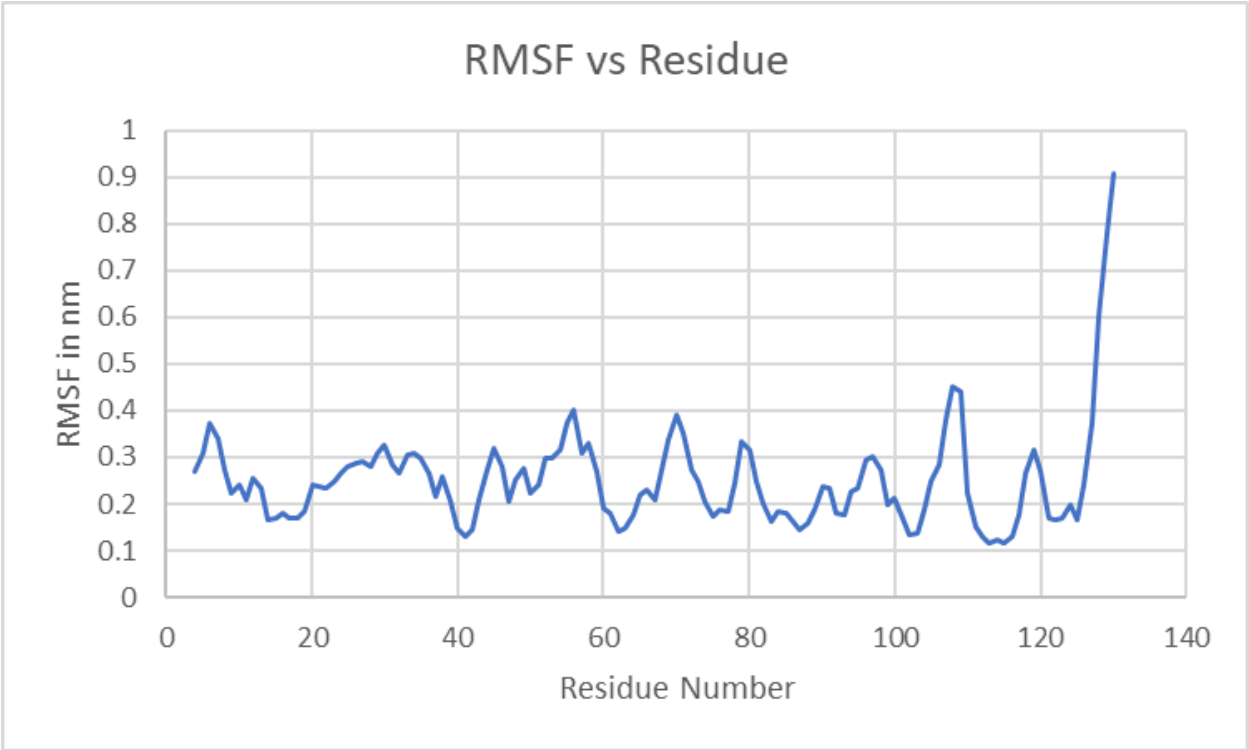
- I then carried out the 1ns production run for further analysis after making the indicated changes in md.mdp.
    - **gmx grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -o md.tpr**
    - **gmx mdrun -v -deffnm md**
- To do trajectory analysis for clear visualization it is necessary to remove the PBC box.
    - **gmx trjconv -s md.tpr -f md.xtc -o md_noPBC.xtc -pbc mol -center**
    - Here, I chosen 1 for the centering the protein and 0 for entire system as output.
- RMSF analysis was carried out using the following command and specifying the -res flag for residue wise RMSF.
    - **gmx rmsf -s md.tpr -f md_noPBC.xtc -o rmsf.xvg -res**
    - Protein was selected as the system when prompted
- Distance between Ca2+ ion and Asp1110
-

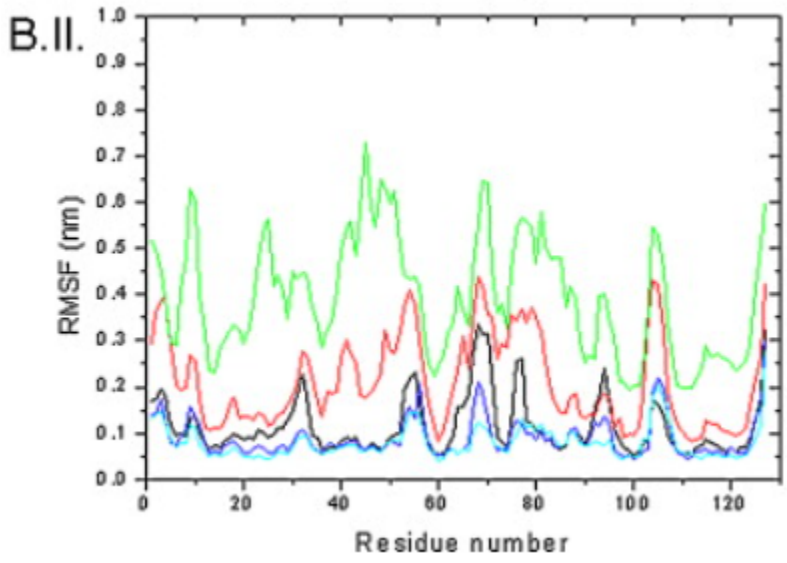


- The atom numbers corresponding to ASP111O and Ca2+ ions were found in npt.gro file.
- Once the index file was created the following command was run:
    - **gmx distance -f md_noPBC.xtc -s md.gro -oall dist.xvg -n index.ndx**
    - Then Group 0 was selected and Ctrl+D was pressed to terminate the command
- RMSD analysis was carried out using the following command:
    - **gmx rms -s md.tpr -f md_noPBC.xtc -o rmsd.xvg**
    - The option backbone was chosen for both RMSD calculation and least squares fit.
- Radius of gyration analysis was carried out using the following command
    - **gmx gyrate -s md.tpr -f md_noPBC.xtc -o gyrate.xvg**
    - Protein was chosen as the group of interest.
- All the plots were obtained using Microsoft Excel for the above mentioned Analyses because xmgrace had installation issues.
- For the final analysis RMSD vs Rg, the RMSD values and Radius of Gyration were extracted from rmsd.xvg and gyrate.xvg for plotting.
- Movie: First a structure file md.gro was loaded in vmd. Then a trajectory file was loaded onto the same molecule.
- Then frames corresponding to every one out of 50 frames , totalling to 10 frames corresponding to every 100 ps and the initial structure was kept deleting the other frames.
- The Background color was turned white and axes were turned off.
- The representation was changed to newcartoon and modified to keep only the proteins.
- From the extensions>Analysis>RMSD Visualizer tool , backbone was selected and I clicked align to arrest the rotational and translational motion.
- From the Extensions> Visualization> Movie Maker , I selected 2s as the duration of the movie. Then I clicked on Make Movie and chose the working directory as Downloads.
- The movie was obtained in mp4 format using Video|Mach.exe .

RESULTS OBTAINED AND COMPARISON WITH PAPER:

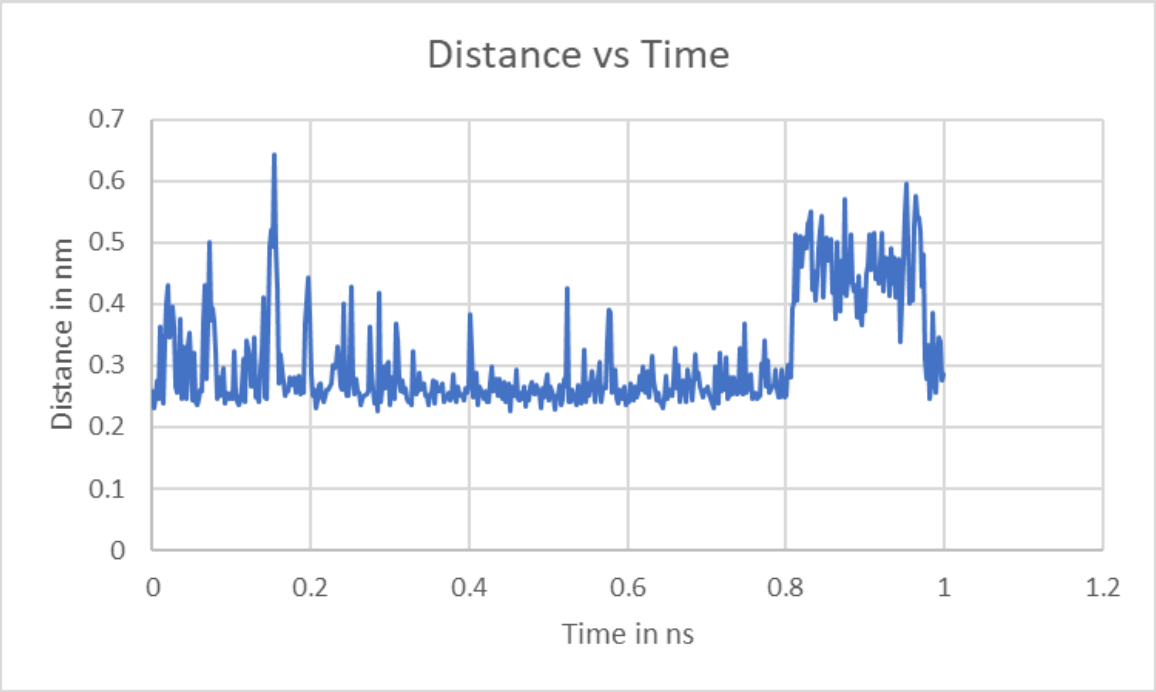
RMSF

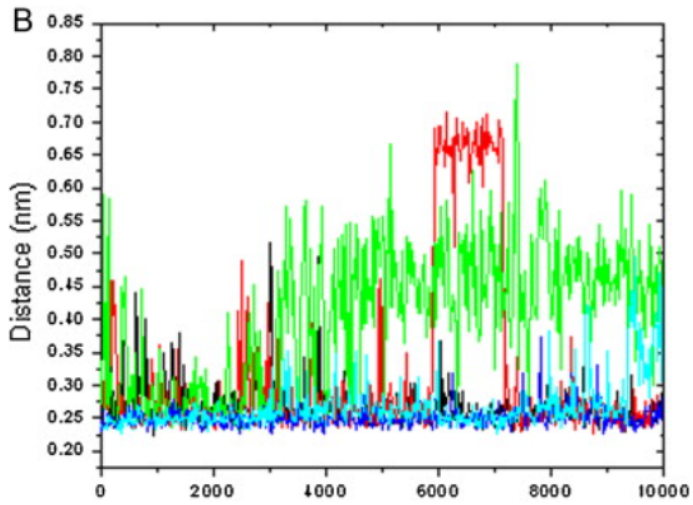




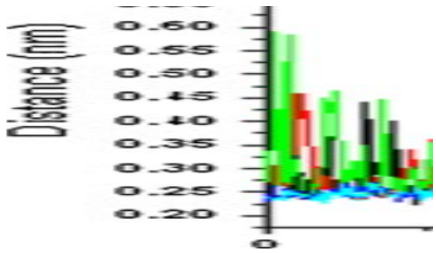
The **RMSF** plot I obtained for my 1ns production run for 2PY2 vs RMSF obtained in the paper for 10ns at 4743(green curve) is quite different from each other due to the difference in the duration of the run. Since , the major change in RMSD happened after 2.5 ns, the curves are quite different.

Distance between Ca2+ ion and Asp 111 O



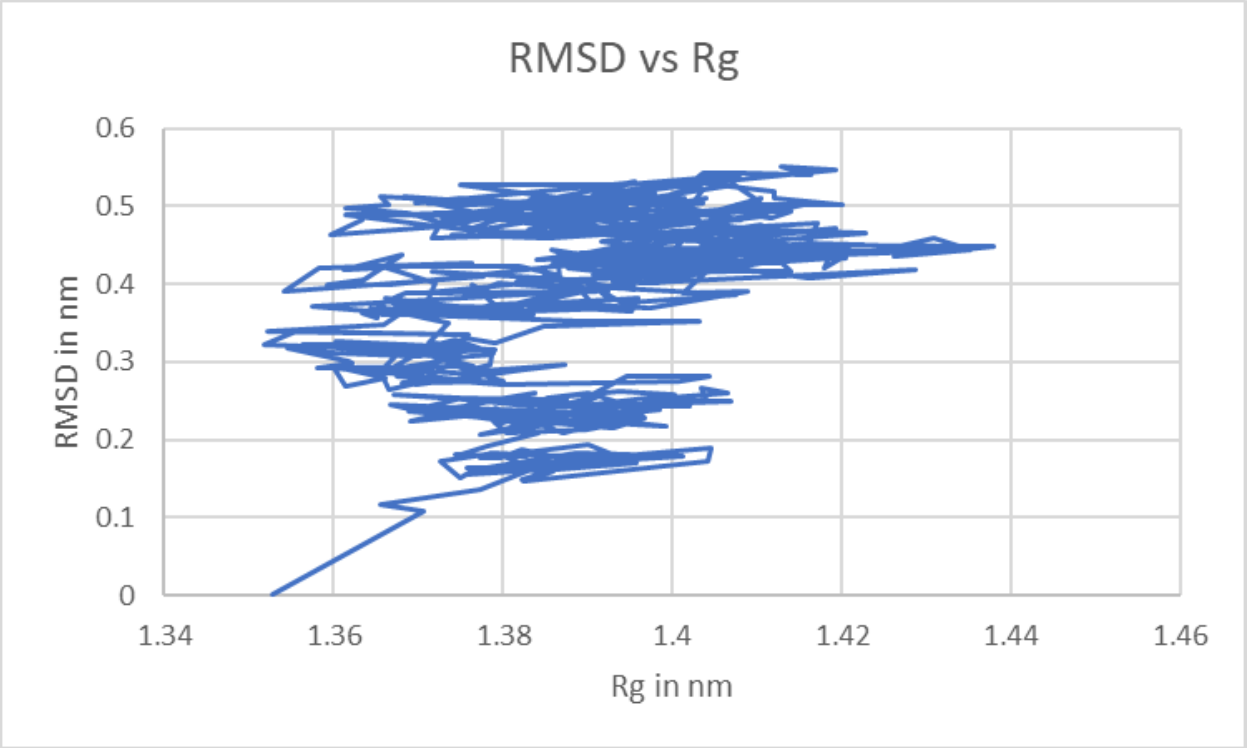


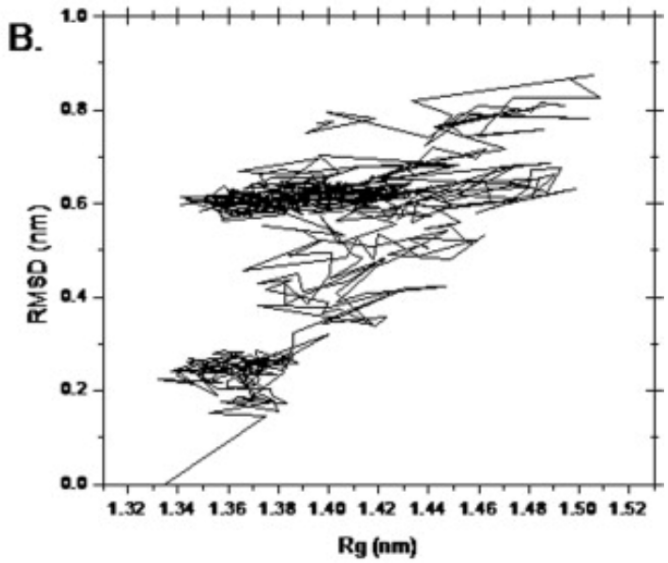
The distance plot I obtained for 1ns run at 473k for 2PY2 protein shows good resemblance in the range of values obtained and similar characteristics with the first ns of the distance plot obtained for 10ns run(green curve). However, it is not too evident due to the presence of other curves corresponding to different temperatures.



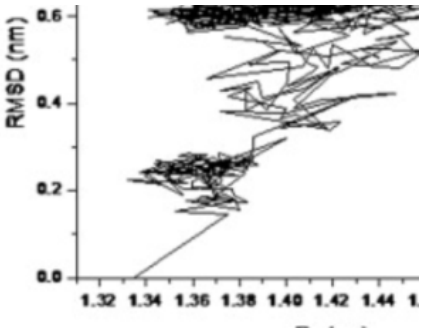
A zoomed out picture of the above plot featuring the plots only in the first nanosecond.

RMSD vs Rg





The RMSD vs Rg plot I obtained shows good resemblance to the one given in the paper
Below is a zoomed out picture of the plot on the left corresponding to the RMSD and Rg values in the first 1ns .



The dihedral angle analysis is carried out in this PDF.