BT5420

Computer Simulations of Biomolecular Systems
Final Course Project

Shreeharsha G Bhat | BE21B037
Department of Biotechnology
Indian Institute of Technology Madras

Table of contents:

Computer Simulations of Biomolecular Systems

The paper :

Gosu, V., Shin, D., Song, K.-D., Heo, J., & Oh, J.-D. (2022). **Molecular modeling and dynamic simulation of chicken Mx protein with the S631N polymorphism**. *Journal of Biomolecular Structure & Dynamics*, *40*(2), 612–621. https://doi.org/10.1080/07391102.2020.1819419

Objective :

Myxovirus are any of a group of viruses of the families Orthomyxoviridae (agents of influenza) and Paramyxoviridae, members of which can cause the common cold, mumps, and measles in humans, canine distemper & rinderpest in cattle, and Newcastle disease in fowl. Mx or Myxovirus resistant proteins are antiviral GTPases induced by type I interferons.

S631N is a Mx protein variant found in chickens and is suggested to possess antiviral activity. This paper aims to analyze the impact of this variant on chicken Mx(chMx) protein structure and conformation. Using computational methods such as molecular modeling , MD simulations and interdomain motion and residue network, the paper examines the structure and dynamic behavior of wild-type and mutant chMx

The Reason I choose this article -

Given the nature of the article, in which the molecular analysis of the dynamics of the wild type protein is compared with the dynamics of the mutated protein would show me the application computer simulations in biological systems. The article sheds light on the conformational changes induced by the mutation in the chMx protein. By providing useful information on the intrinsic dynamics and structural alterations between the mutant and wild type, the article aids in better understanding the antiviral activity chMx that might be associated with the S631N polymorphism.

The paper required the generation of a complete theoretical 3D model for the protein and its corresponding mutation. This would imply the use of new softwares and web servers like I-TASSER, ModRefine and Discovery Studio Visualizer which piqued my interest. In this article the protein is put in a PBC box that is not the conventional cubic shape, but it is of a dodecahedron.

The MD run for the final simulation was very similar to that of simulating lysosomes in water. This provided me with a range of new software and tools to be learnt, while still re-learning the basics.

Results Reproduced :

Trajectory analysis was performed on both chmx_wt(wild type) and chmx_S631N(mutant). In the paper, Root-mean-square deviation (RMSD), Radius of Gyration(Rg), solvent accessible surface area(SASA), root-mean-square fluctuation(RMSF) for the last 100 ns of the MD trajectory analysis and Interdomain analysis were performed for both the protein and mutant.

After performing the MD run for both we need to remove the PBC box from the trajectory before analyzing any of the above function, to avoid erroneous results when molecules move out into adjacent boxes

The command used to do this is

gmx trjconv – s md.tpr – f md. xtc – o md_noPBC. xtc – pbc mol – center

a) -s md.tpr provides the structure file for the reference structure
b) -f md_nopbc.xtc has the trajectory of the protein at all the time frames recorded at 100ps each
c) -o md_nopbc.xtc is the output trajectory file without the PBC box
d) -pbc mol sets the periodic boundary condition treatment, mol puts the center of mass of the molecule
e) -center centers the atom in the box

When prompted, select (1) - `Protein' as the group to be centered and (0) - `System' for output

Root Mean Square Deviation (RMSD) :

RMSD is used to analyze time dependent motion of a given structure and for determining its spatial convergence throughout the simulation. When prompted by GROMACS, select (4)- `Backbone' to compute the least square fit and (1)- `Protein' for calculating the RMSD function Throughout the run, a low RMSD value indicates the structure is very similar to the reference structure. It suggests that the system is relatively stable and hasn't undergone significant structural conformational changes. While a high RMSD value indicates otherwise

The command used to do this is

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

a) -f md_nopbc.xtc has the trajectory of the protein at all time frames recorded at 100ps
b) -s md.tpr provides the structure file for the reference structure, against which the RMSD is calculated.
c) -o rmsd.xvg is where the output file is stored

Radius of Gyration (Rg) :

Radius of gyration is a structure metric used in MD simulation. It provides information about the compactness or the spread of a molecular structure.The command in GROMACS gives the length of the protein i.e. the radial measure of the protein from the center of mass(COM). While prompted select (1)- `Protein' to compute the Rg (all protein atoms taken into account)

The command used to do this is

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

Solvent accessible surface area (SASA) :

Solvent Accessible Surface Area (SASA) is a measure of the surface area of a biomolecule that is accessible to solvent molecules, typically water, during a molecular dynamics simulation. Hence

a SASA analysis will be helpful in understanding the protein's ability to interact with solvents and other molecules. When prompted after typing the command below, select (1)- `Protein' to compute SASA(for every protein atom including the hydrogen atoms)

The command used to do this is

gmx sasa -f md_noPBC.xtc -s md.tpr -o sasa.xvg -n input.ndx

input.ndx will contain the atoms of interest for SASA for a given residue

These are the **three analyses** that will be performed as part of the report, for both the wild type and Mutant protein.

These are the analysis I am determined to perform due to their ease of replication and direct analysis,

RMSF is not replicated since the paper analyzes RMSF for the last 100ns, and since we are performing the analysis for just the first 1ns, we cannot analyze it.

Along with these three, a self written code for the **Ramachandran Plot** for chmx_wt and chmx_S631N will also be plotted as part of this project

Simulation Details:

1) The Force Field used is AMBER99-sb-ILDN
2) Both the native and mutant protein are solvated with tip3p water molecules
3) PBC box - Dodecahedron with a periodic distance of 1.4nm
4) Since Protein shows no net charge, no extra ions included to neutralize the system
5) Energy Minimization performed using steepest gradient method with 1000 KJ/nm
6) Coulomb and Van der Waals interactions truncated at 1.0nm
7) NVT simulations performed for 100ps
8) NPT simulations performed for 500ps using positional restraints
9) V-rescale thermostat and a Parrinello-Rahman barostat used to maintain temperature at 300K and pressure at 1.0bar

Finally a production run for 200ns(1ns for this simulation) was performed in the absence of positional restraints

The analysis for RMSD, RMSF, Rg and SASA was performed using GROMACS analysis modules gmx, rms, gmx & gyrate. The graphs were prepared using Excel

*.mdp file modifications :

1) ions.mdp
2) minim.mdp
3) nvt.mdp
4) md.mdp (for 1ns)

Had no modifications done to them.

npt.mdp - modified to make the npt simulation run for 500ps
Nsteps changed to 250000 ; 2*250000 = 500ps for both the mutant and wild type

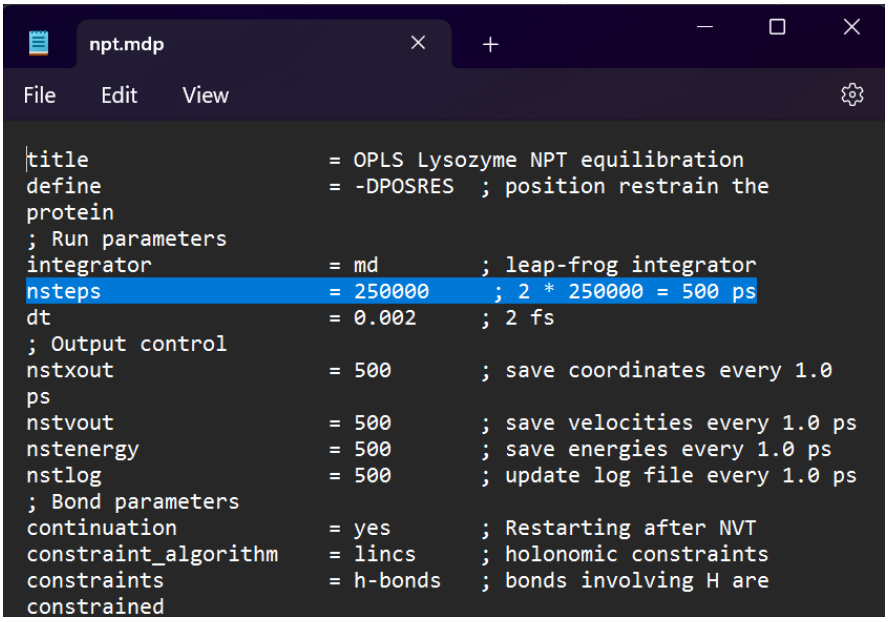

Fig 1: changes made to npt.mdp highlighted in blue

These modifications were made in regards to the paper, so that the results would could be replicated accurately

VMD movie:

The video and the screenshots for both the chmx_wt & chmx_S631N are attached as hyperlinks.

To convert the screenshots(bmp format) to video(mp4) format, the following python script was implemented.

Description of Computation:

Protein Modeling:

- The amino acid sequence of the chmx protein was taken from UniProt Database(ID: Q90597)
- The sequence was then submitted to the I-TASSER web server(recommended by the paper), which is the most popular online tool for automated structural prediction and annotation.
- The server predicted 5 models and the one with the highest c-score[model 2](prediction quality) was picked.
- The model 2 was then sent to the ModRefine server to remove steric bumps and clashes in the structural model. (pdb file - chmx_wt)
- To construct the mutant model, the serine at position 631 was changed to asparagine using discovery studio Visualizer (pdb file - chmx_S631N)

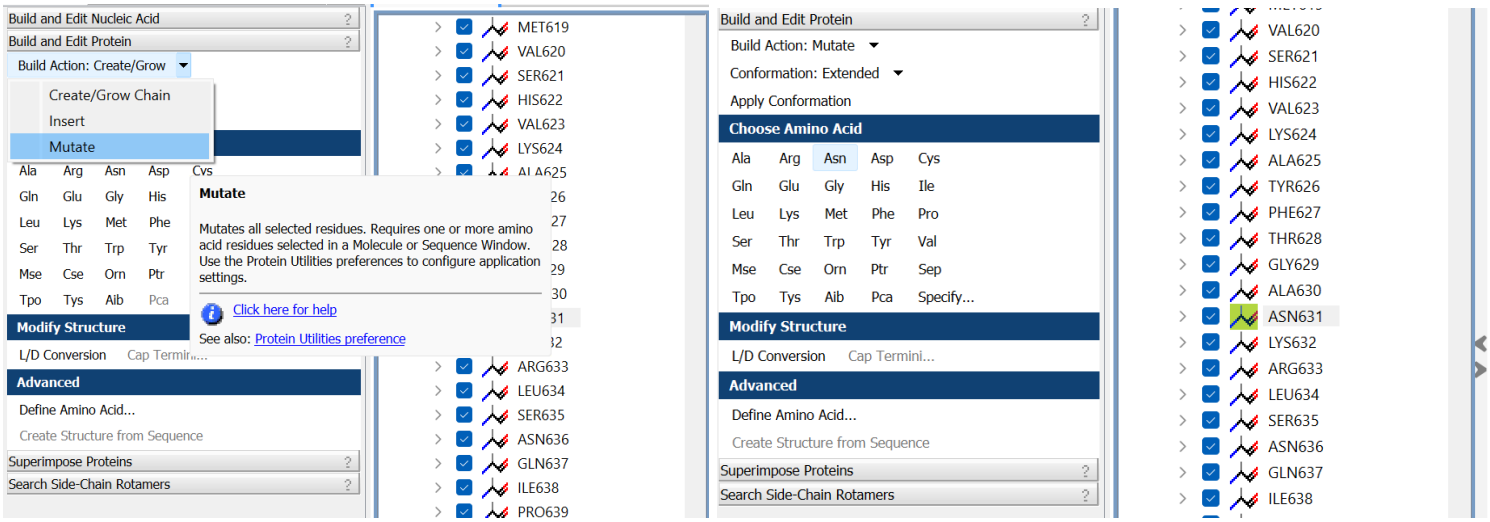

Fig 2: Mutating the protein by replacing the SER at 631 to ASN in Discovery studio Visualizer

Protein Simulation:

- Since there are no water molecules in both the PDB files, there is no need to remove anything.
- We now need to perform energy minimization on both chmx_wt & chmx_S631N
- **.pdb to .gro conversion**(-ignh flag added to remove H atoms from the pdb, so that GROMACS can add them automatically when it converts from pdb to gro)

gmx pdb2gmx -f chmx_wt.pdb -to chmx_wt.gro -ignh for chmx_wt
gmx pdb2gmx -f chmx_S631N.pdb -to chmx_S631N.gro -ignh for chmx_S631N

(6) - to pick AMBER99-sb-ILDN force field and (1) for Tip3P water molecule solvation

- **Adding Dodecahedron Box** and centering

gmx editconf -f chmx_wt.gro -o chmx_wt_box.gro -c -d 1.4 -bt dodecahedron for chmx_wt
gmx editconf -f chmx_S631N.gro -o chmx_S631N_box.gro -c -d 1.4 -bt dodecahedron for chmx_S631N

- **Neutralize,** Since there was -2e charge in the predicted model by I-Tasser, we add two NA+ ions to neutralize them.(both chmx_wt & chmx_S631N)

gmx grompp -f ions.mdp -c chmx_wt_solv.gro -p topol.top -o ions.tpr
gmx genion -s ions.tpr -to chmx_wt_solv.gro -p topol.top -pname NA -nname CL -neutral
(chmx_wt)

gmx grompp -f ions.mdp -c chmx_S631N_solv.gro -p topol.top -o ions.tpr
gmx genion -s ions.tpr -to chmx_S631N_solv.gro -p topol.top -pname NA -nname CL -neutral
(chmx_S631N)

When prompted type (13) to neutralize

- **Energy Minimization**

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

gmx mdrun -v -deffnm em for chmx_wt

gmx grompp -f minim.mdp -c chmx_S631N_solv.gro -p topol.top -o em.tpr

gmx mdrun -v -deffnm em for chmx_S631N

- **NVT Simulation** (100ps)

For both chmx_wt and chmx_S631N

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

gmx mdrun -v -deffnm nvt

- **NPT Simulation** (after modification mentioned in *.mdp modification)(500ps)

For both chmx_wt and chmx_S631N

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

gmx mdrun -v -deffnm npt

- **MD Simulation** (1ns)

For both chmx_wt and chmx_S631N

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

gmx mdrun -v -deffnm md

- **Removal of PBC box and trajectory run**

For both chmx_wt and chmx_S631N

gmx trjconv – s md.tpr – f md. xtc – o md_noPBC. xtc – pbc mol – center

- **RMSD, Rg Calculation**

For both chmx_wt and chmx_S631N

RMSD : gmx rms -f md_noPBC.xtc -s md.tpr -o rmsd.xvg (4)- Backbone, (1)- Protein

Rg : gmx gyrate -f md_noPBC.xtc -s md.tpr -o gyrate.xvg (1)- Protein

- **SASA Calculation for 631 position**

First we need to create an ndx file with the atoms of interest for position 631

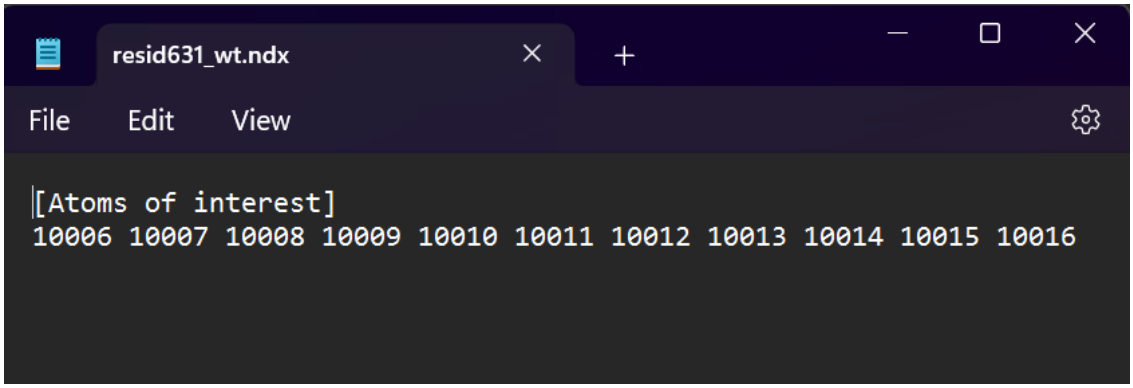

Fig 3: Ndx file for position 631 in wild type protein

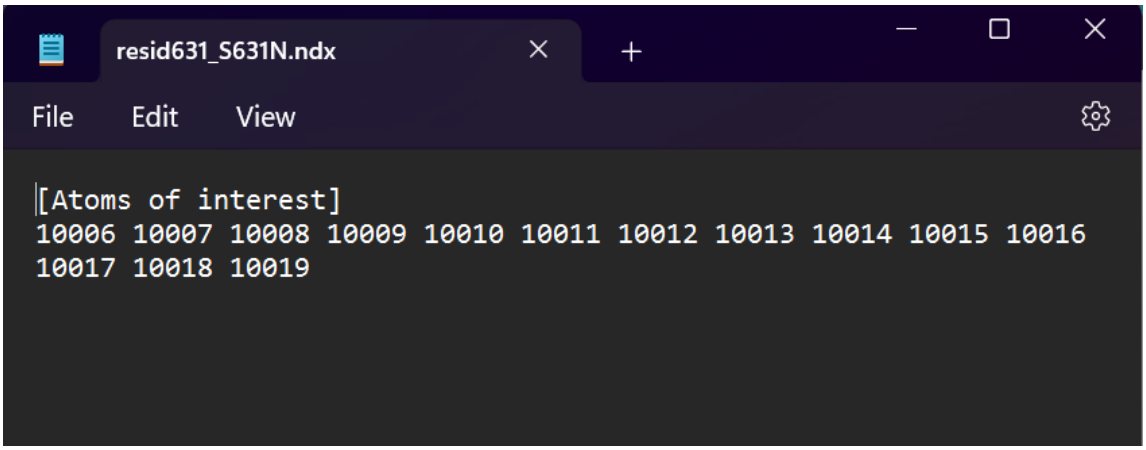

Fig 4: Ndx file for position 631 in mutant protein

Then we run the following command for chmx_wt and chmx_S631N respectively

gmx sasa -f md_noPBC.xtc -s md.tpr -o area1.xvg -n resid631_wt.ndx

gmx sasa -f md_noPBC.xtc -s md.tpr -o area2.xvg -n resid631_S631N.ndx

Results Comparison with article:

The Paper performed three simulation for each of RMSD, Rg & SASA, we will compare our results with simulations with the most overlap

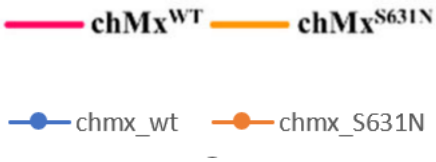

Fig 3: legends for the graphs given below

1) Root Mean Squared Deviation

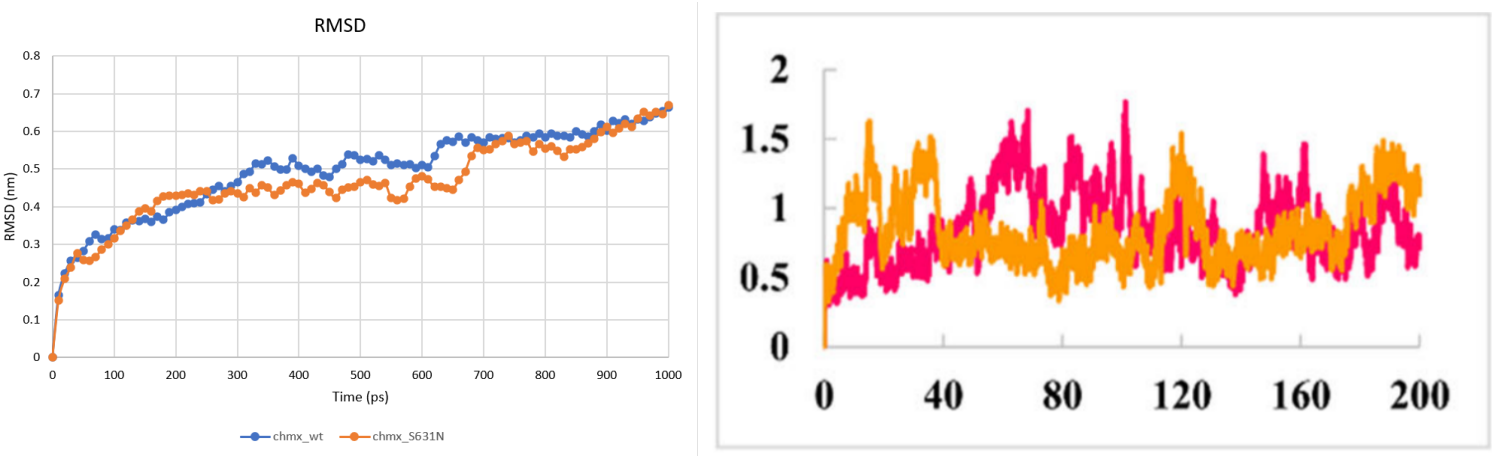

Fig 5: RMSD of chmx_wt and chmx_S631N (i) this project (ii) the article

Reasoning: This project runs the simulation only for 1ns, while the article has the MD run for 200ns. The shorter the simulation time the more will be the difference between the two, however if you zoom-in, we can see that the RMSD plot is very similar to that of simulation 2(article), ensure that the behavior of the results is as expected. We see that in both the graphs on the whole the RMSD value keeps increasing, and the replicated results show the same for both the protein and its mutant

2) Radius of gyration

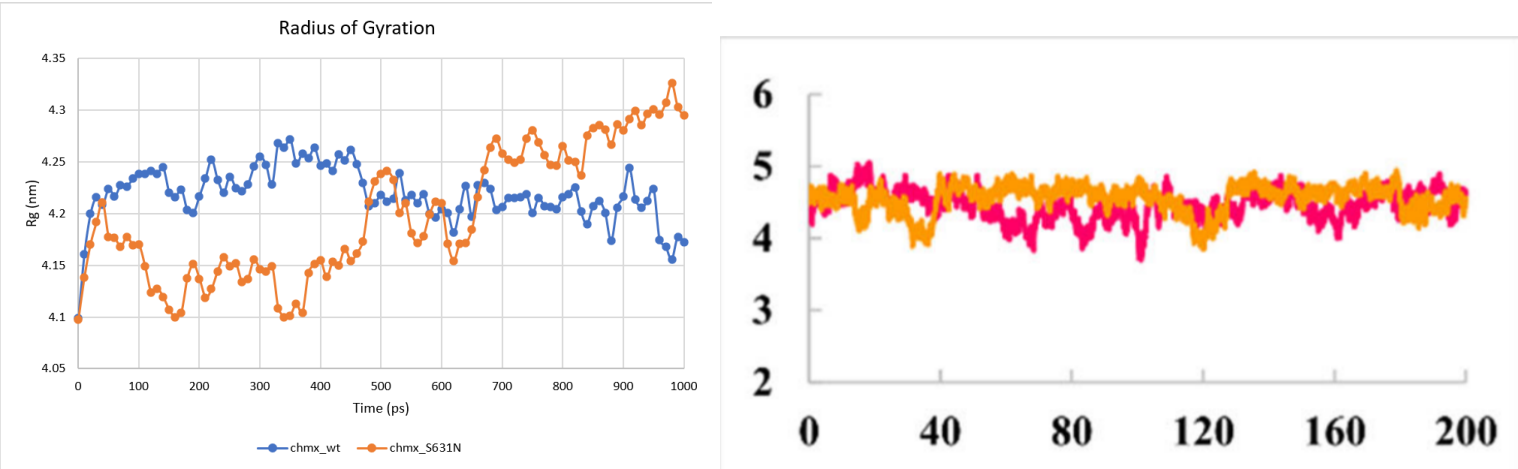

Fig 6: Rg of chmx_wt and chmx_S631N (i) this project (ii) the article

Reasoning: This project runs the simulation only for 1ns, while the article has the MD run for 200ns. The shorter the simulation time the more will be the difference between the two, however if you zoom-in and see the trends for the first nanosecond we see there is a minor variation between the mutant and the wild-type, this could be due to the 2 Na+ ions that were added only in this project and not in the article. The two Na+ ions were added since the I-TASSER prediction came with a negative charge of 2 on the whole.

3) SASA

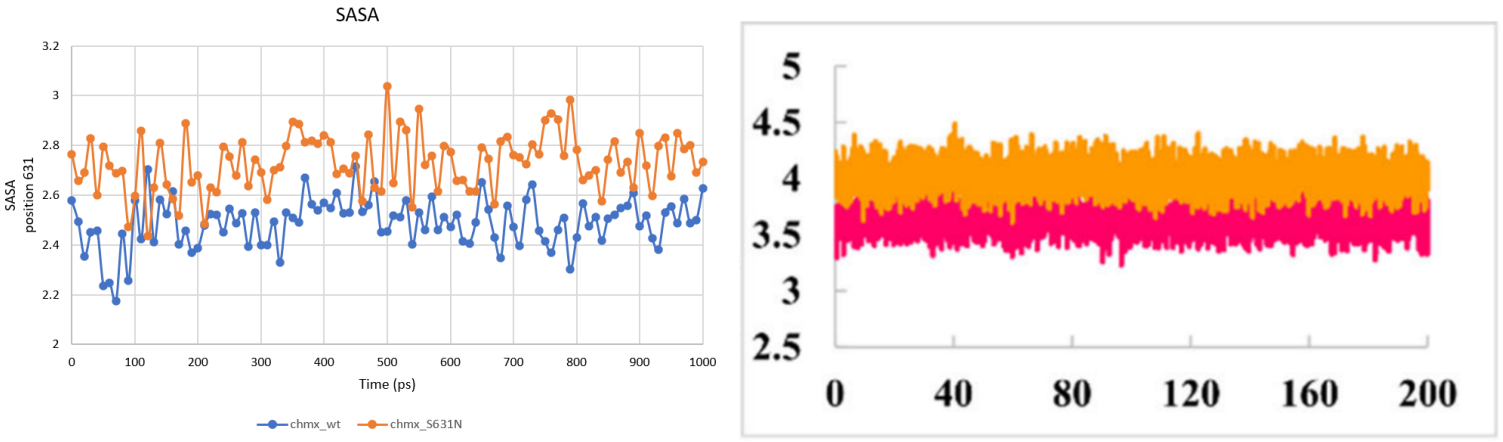

Fig 7: SASA for 631 position of chmx_wt and chmx_S631N (i) this project (ii) the article

Reasoning: This project runs the simulation only for 1ns, while the article has the MD run for 200ns. The shorter the simulation time the more will be the difference between the two, however if you zoom-in and see the trends for the first nanosecond are very similar when it comes to qualitative aspects. We see that the solvent accessible surface area is higher for the mutant than the wildtype which is the same in the article as well. The change in the value of SASA could be due to the 2 Na+ ions that were added only in this project and not in the article.

Ramachandran Plot of the given PDB file

We now Plot the Ramachandran plot for chmx_wt using the following Python script.

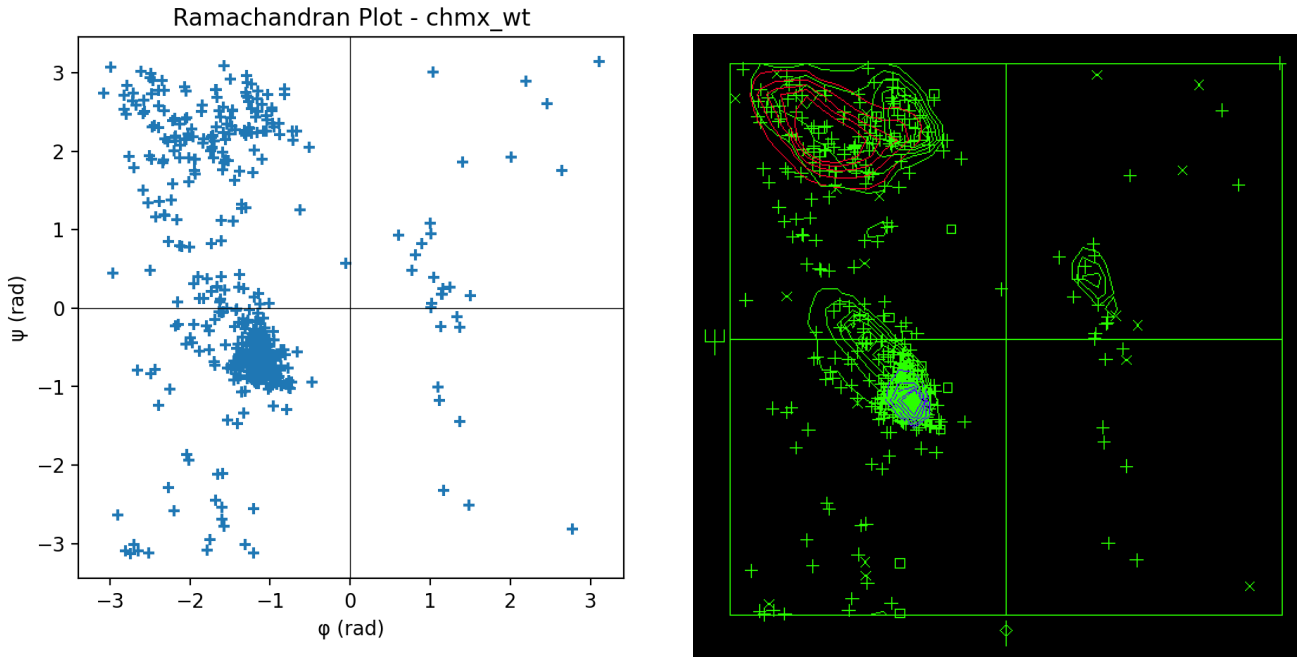

Fig 8: Ramachandran Plot comparison between the python script and online calculator for chmx_wt

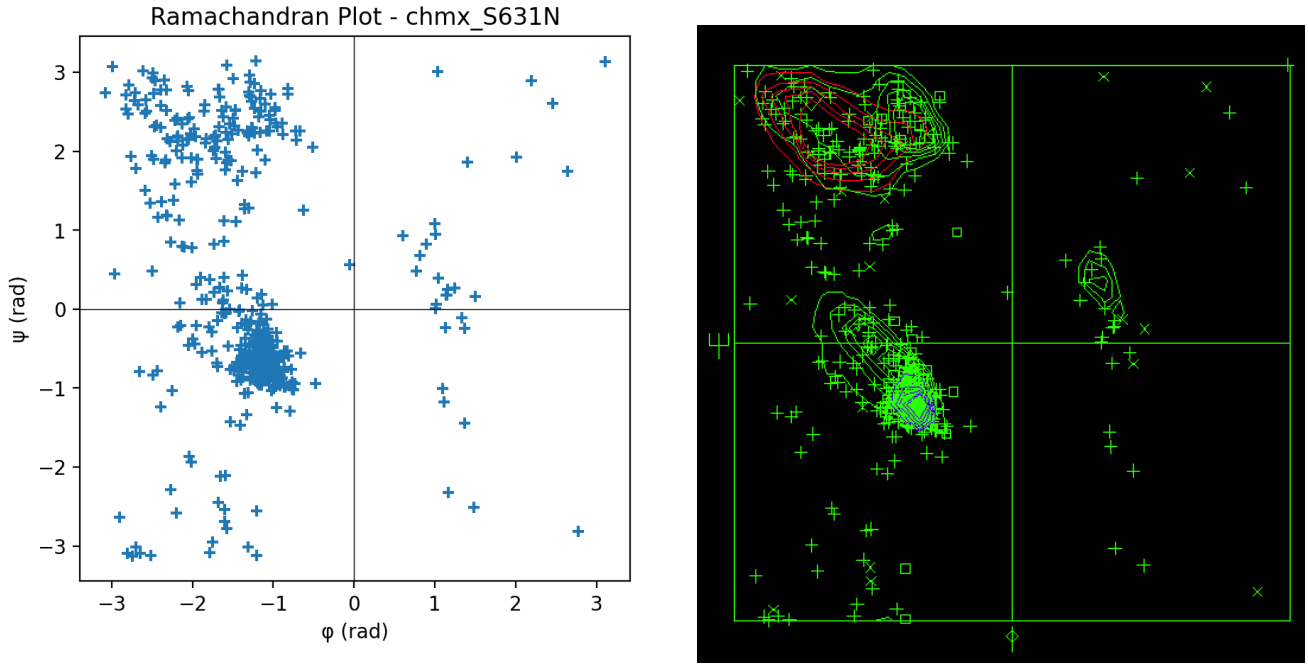

Fig 8: Ramachandran Plot comparison between the python script and online calculator for chmx_S631N

MDP files

The following are the mdp files used for this project - [mdp files]

Acknowledgements

I thank the course faculty Dr. Hamsa Priya Mohanasundaram, for her valuable inputs throughout the course.

****