Anirudh Rao
BE21B004

BT4110 – Computational Biology Lab

Practical 3

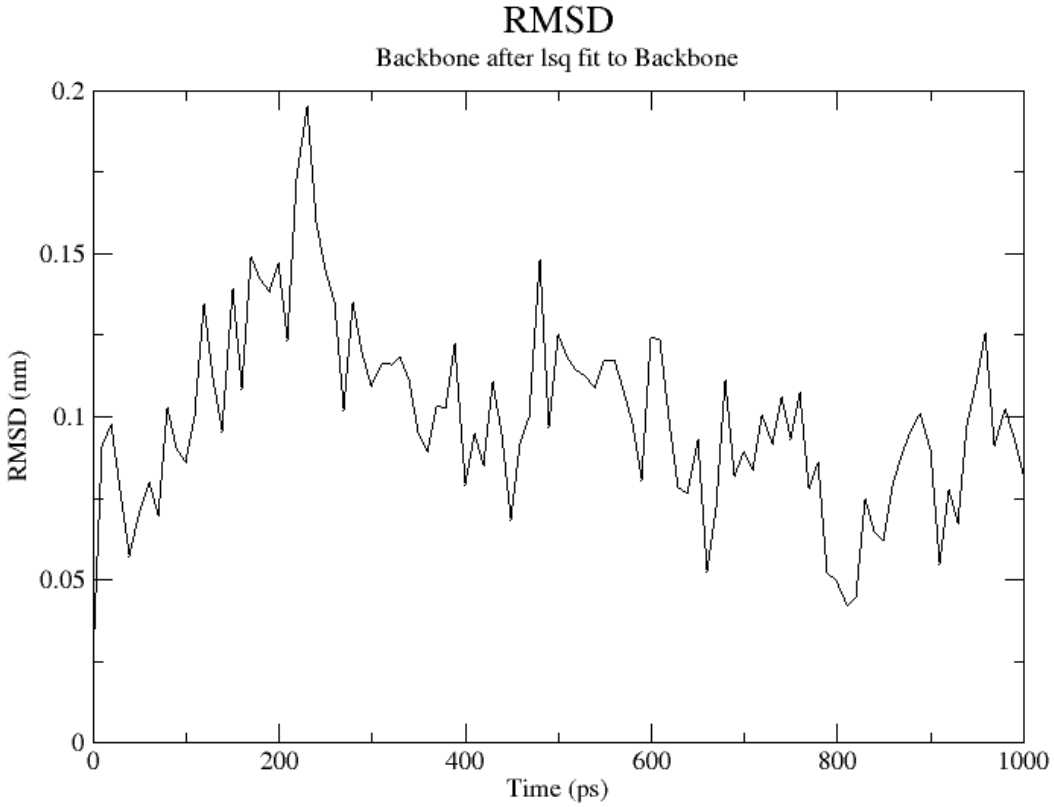
1

RMSD stands for 'root mean square deviation'. It is used to measure the average difference between the structures of a protein at two different timepoints of an MD simulation. It is an indirect measure of the stability of the protein. A high RMSD implies that the initial protein structure is unstable under the given conditions. A low RMSD implies that the initial structure is relatively stable. Usually, the RMSD is calculated on the basis of the backbone atoms. The two structures are overlayed on each other by least squares fitting and the RMSD is computed as

$$\mathrm{RMSD}(t_1, t_2) = \sqrt{\frac{\sum_i m_i \|r_i(t_1) - r_i(t_2)\|^2}{\sum_i m_i}}$$

Here $r_i(t)$ is the position / coordinates of atom i at time t.

The RMSD of the protein backbone during the simulation was computed with reference to the energy minimised crystal structure by using `gmx rms`. The plot was then visualised in Grace.



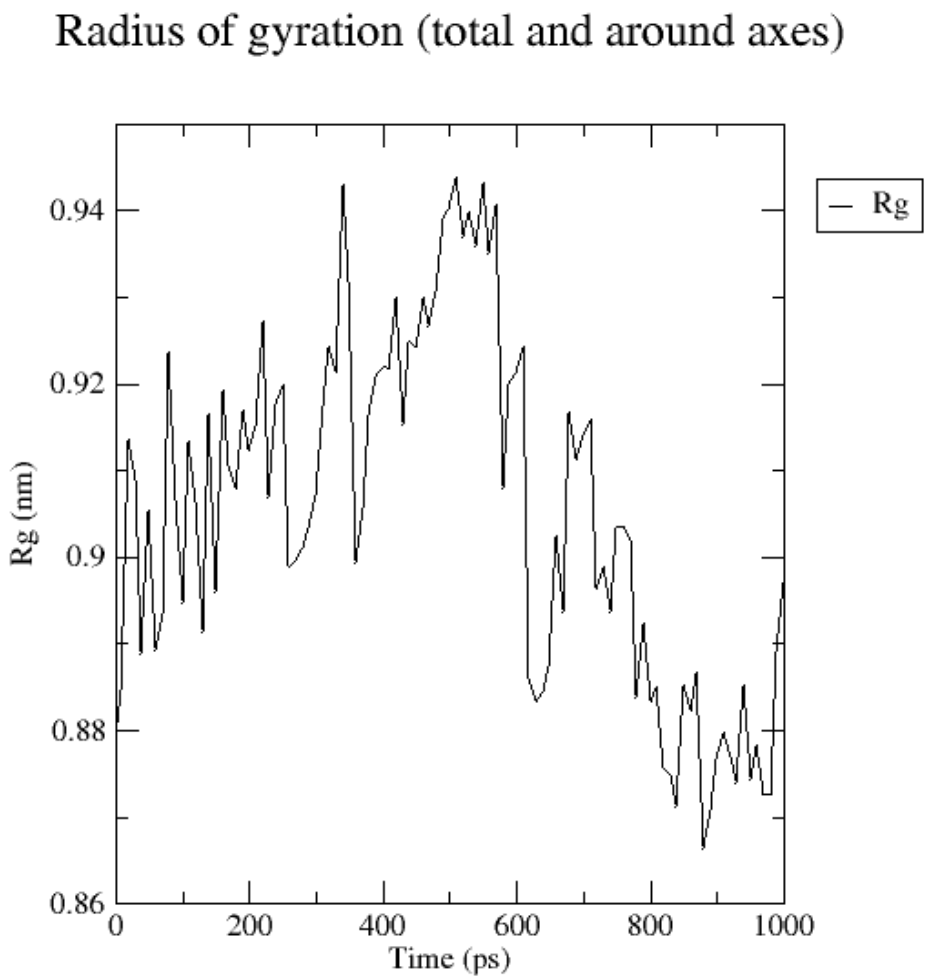
The RMSD initially increases but then stabilises towards the end of the production run. After 1 ns, the RMSD is **0.0820437 nm**.

Rg is the radius of gyration. It is the root mean square distance between all atoms of the protein and the centre of mass of the protein. A low Rg implies that the protein is compact while a high Rg implies that the protein is large in size.

$$R_g(t) = \sqrt{\frac{\sum_i m_i(\|r_i(t) - r_{\mathrm{COM}}(t)\|)^2}{\sum_i m_i}}$$

If the Rg of a protein increases during a simulation, it implies that the native structure is losing stability and is unfolding.

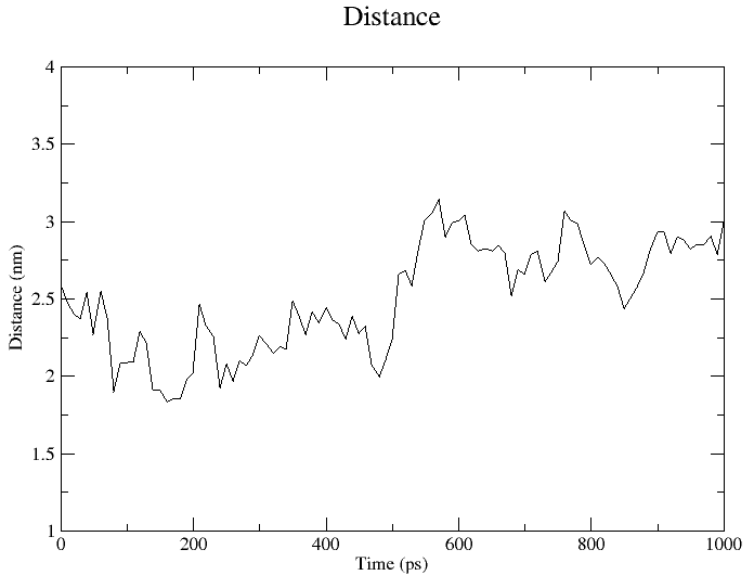
The Rg of the protein during the simulation was computed by using `gmx gyrate`. The plot was then visualised in Grace.



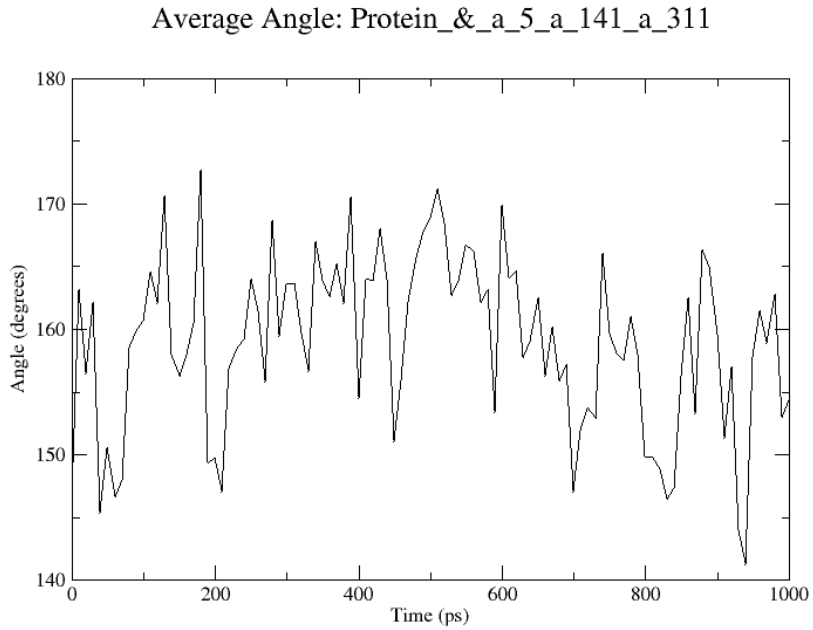
The Rg initially increases but decreases as the simulation ends. At the end of the production run, the Rg was **0.897243 nm**.

3

From the .gro files, we can see that the first and last atoms have indices 1 and 328 respectively. These indices were put into an .ndx file using `gmx make_ndx`. The distance between them as a function of time was computed using `gmx distance`. The plot was then visualised in Grace. The distance between the two atoms gradually increases over time after some fluctuations.



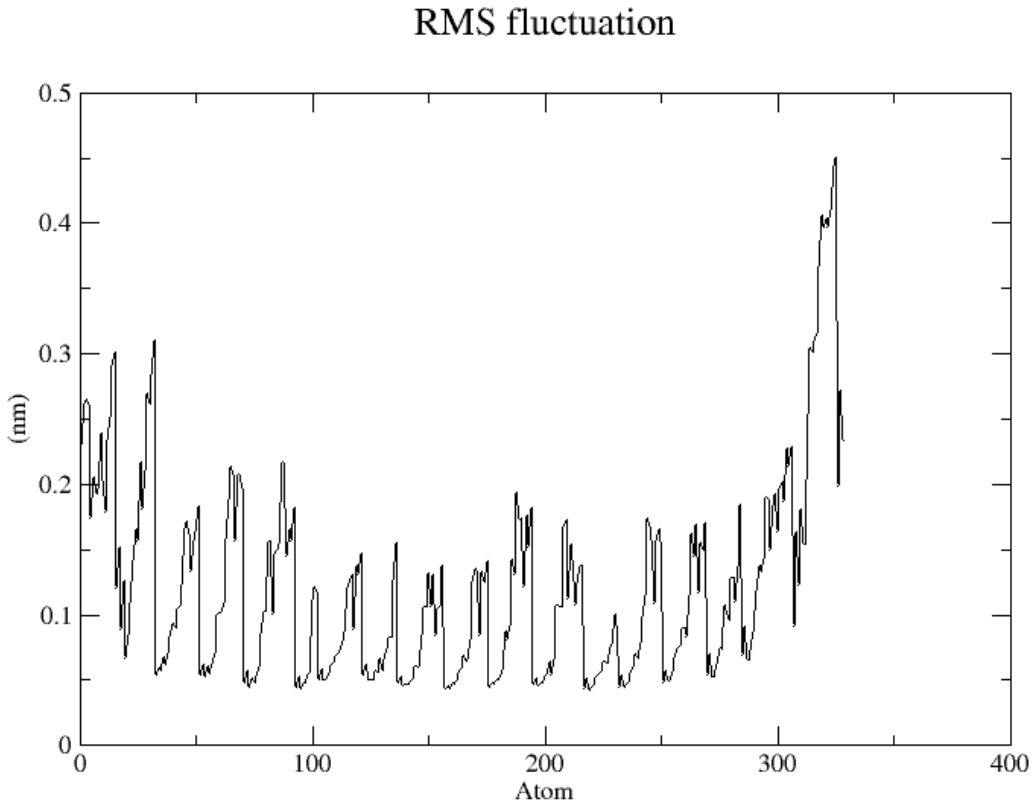
Next, the indices of the C_α atoms of Glu1, Phe9, and Leu18 (indices 5, 141, and 311) were put in an .ndx file. The angle between them as a function of time was computed using `gmx angle`. This was then visualised in Grace. The angle between them fluctuates throughout the production run. It varies from 140 degrees to 174 degrees, with an average of 159 degrees.



4

RMSF stands for 'root mean square fluctuation'. It is used to measure how each atom / residue fluctuate in their position during the simulation. Regions with high fluctuation are the ones that show the highest conformational change during the simulation. Regions that do not fluctuate show structural stability. Studying both these kinds of regions may be important to understand the function of the protein under the conditions being simulated.

The RMSF of the protein's atoms was computed using `gmx rmsf`. This was then plotted in Grace.

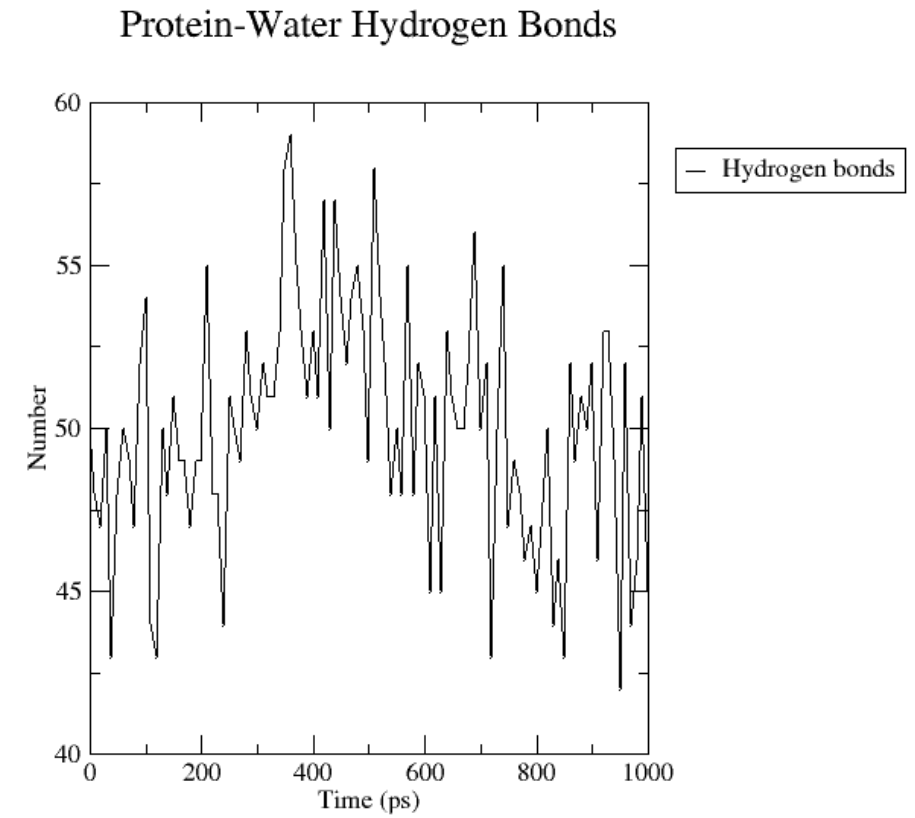


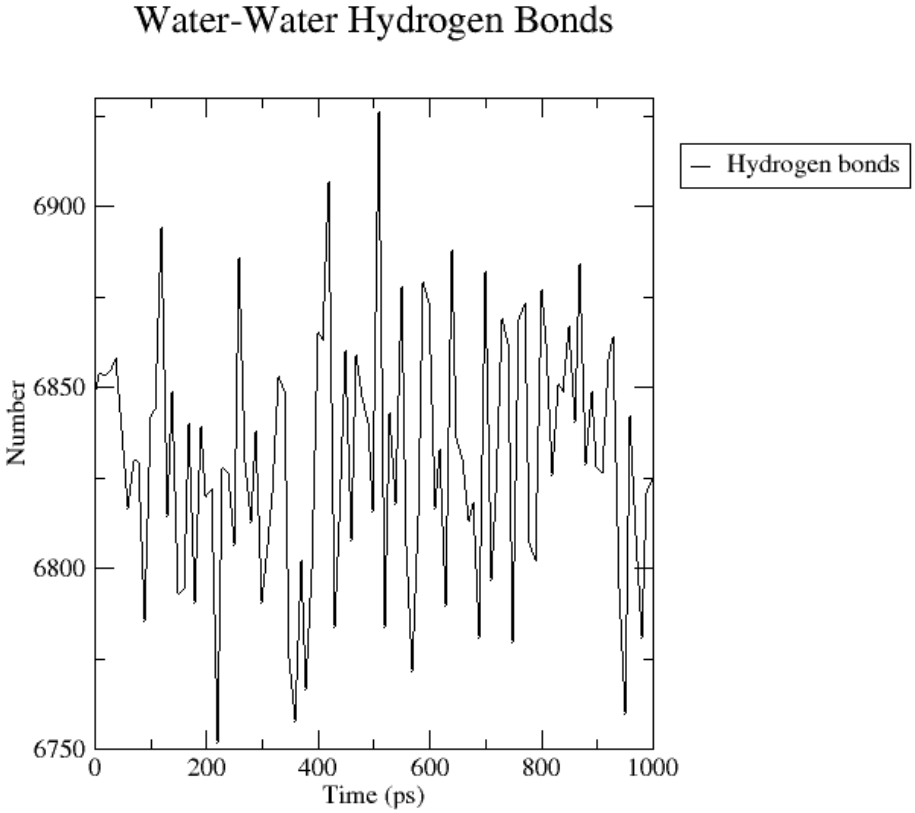
The atom that fluctuates the most is the **atom with index 325**, which has a fluctuation of 0.4506 nm. This atom is one of the hydrogen atoms (**HD23**) of the last residue of the protein – **Leu18**. It is attached to one of the C_δ atoms of Leu18.

5

Using `gmx hbond` and setting the two groups to be analysed as (System, System), we find that the average number of hydrogen bonds per time frame is **6895.564**.

We then obtain the number of protein-water hydrogen bonds and water-water hydrogen bonds as a function of time using `gmx hbond` and setting the groups to be analysed as (Protein, SOL), and (SOL, SOL) respectively. These plots were then visualised in Grace.





Both of these fluctuate throughout the simulation.