Universiteit van Amsterdam

Biomolecular Simulations

Exploring the Conformational Space of SARS-CoV-2 Spike Protein

Ignas Krikštaponis (13250868)[a]

[a]ignas.krikstaponis@student.uva.nl

30 May 2021

Abstract

blah.

1 Introduction

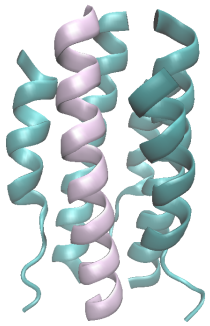


Figure 1: Protein. Pink - chain A.

The project is layed out in two parts: molecular dynamics and metadynamics simulations.

2 Methods

2.1 Molecular dynamics

Molecular dynamics (MD) is a simulation technique that allows conformational space exploration of a molecular system. The main principle behind the simulation is integration of Newton's laws of motion equations (Leach [2007]). To investigate the stability of spike protein complex two simulations were carried out at a temperature of 310 K and one at 400 K. Two simulations at 310 K were conducted to account for the stochastic nature of the molecular dynamics simulations with each simulation drawing a different sample of velocities assigned to the atoms based on the Boltzmann distribution at 310 K. All simulation were carried out using `gromacs` software package.

System preparation The protein complex was placed into a box filled with 6407 water molecules. The resulting size of the box - 216.52 nm^3. This was then followed by energy minimisation to mitigate unwanted interactions between atoms. 43

Table 1: Parameters used for molecular dynamics simulations.

Simulation	310 K	400 K
Force field	AMBER99SB-ILDN	AMBER99SB-ILDN
Water model	TIP3P	TIP3P
Reference pressure (bar)	1	1
Barostat	Parrinello-Rahman	None
Thermostat	V-rescale	V-rescale
Time (ns)	100	100
Integration algorithm	leap-frog	leap-frog
Coulomb method	PME	PME
Van der Waals method	cut-off (1.1 nm)	cut-off (1.1 nm)

water molecules were replaced with 23 NA and 20 CL ions and again followed by energy minimisation. Finally, an equilibration step was applied on the system. After this, the molecular dynamics simulations were started. Parameters used for system preparation and simulations are displayed in Table 1.

2.2 Metadynamics

Metadynamics (MetaD) simulations were used to investigate the free energy profile of the spike protein complex. By adding bias potential to make the most visited states unfavourable, MetaD allows better sampling of rare events and thus better exploration of the conformational space (Liao [2020]). The simulations were performed using the `plumed` plugin for `gromacs`. The same `gromacs` files as prepared with the procedure laid out in Section 2.1 were used for metadynamics simulations (at a temperature of 310 K). In total, 3 simulations were performed - specifications can be found in Table 2.

Simulations 1 and 3 (Table 2) were set up to explore the free energy profile of the unfolding process of chain A. The main difference between the two simulations lay in the hill height parameter - it is 2 times higher for simulation 3. This was done to investigate whether increasing the hill height accelerates the exploration of the free energy surface. Simulation 2 explores the free energy profile of the chain A detachment from the rest of the complex.

Table 2: Parameters and collective variables used for metadynamics simulations

Simulation	1	2	3
Collective variable (CV)	Distance between atoms 1-19 and atoms 369-388	Distance between atoms 1-388 and atoms 389-2046	Distance between atoms 1-19 and atoms 369-388
CV description	Distance between mass centers of N and C-terminus in chain A	Distance between mass center of chain A and mass center of the rest of the complex	Distance between mass centers of N and C-terminus in chain A
Hill height (kJ/mol)	0.01	0.01	0.02
Hill width (nm)	0.35	0.35	0.35
Hill step size	500	500	500

2.3 Analysis

A number of measurements were extracted from the simulations to quantify the processes observed in simulations. Visual investigation was done using `VMD` software.

Physical properties Temperature, pressure and volume progressions during MD simulations were extracted via `gromacs`' `energy` module.

RMSD Root mean square deviation of the whole complex was calculated over time via `gromacs`' `rms` module.

Distances Two distances were calculated from the MD simulations: between C and N-terminus of chain A and between chain A and the rest of the complex. The distances were calculated between the centers of mass - analogous to *Collective variable (CV)* in Table 2.

Free energy profile The bias potential every 500 deposited hills was calculated via `plumed`'s `sum_hills` module for the MetaD simulations.

3 Results and Discussion

3.1 Molecular dynamics

3.1.1 Physical properties

Physical properties were inspected to ensure that the simulations were carried out under the right conditions. Average temperature, pressure and volume of the three MD simulations can be found in Table 3. Values for 310 K simulations were as expected - average temperature, volume and pressure follow the values as explained in Section 2.1. For the simulation at 400 K, as there is no barometer to keep the pressure constant, the pressure is higher while the volume is lower. This is due to the fact that in order for water to maintain liquid properties at 400 K the pressure must increase.

Progression of selected physical properties can be found in Appendix A.

Table 3: Averaged selected physical properties of MD simulations

Simulation	310 K - 1	310 K - 2	400 K
Temperature (K)	310.01	310.02	400.01
Pressure (bar)	0.45	2.55	1500.27
Volume (nm^3)	215.09	215.14	17.93

3.1.2 Overall stability of the complex

In Figure 2 the RMSD of the whole protein complex is plotted over the simulation period. We can observe an obvious conformational state transition at ≈ 60,000 ns in the *310K - 1* simulation (from ≈ 0.2 nm to ≈ 0.38 nm). This is also clearly visible by the two hills in the density plot on the right. A similar transition happens in *310K - 2*, however it happens earlier and in multiple steps. As expected, the RMSD for the *400K* simulation is higher - this can be explained by the higher velocities linked to the higher temperature. *400K* simulation transition from ≈ 0.25 nm to ≈ 0.42 to ≈ 0.5. The RMSD stays relatavely stable for all 3 simulations after ≈ 70,000 ns.

Upon visual inspection with `VMD` we could conclude that most of the deviations from the original structure can be attributed to the unfolding of α-helices, therefore it is also useful to inspect hydrogen bonds in the system. The results can be found in Figure 3. Simulations at 310 K had on average ≈ 120 hydrogen bonds, while the simulation at 400 K had ≈ 106. No clear state transitions can be observed here.

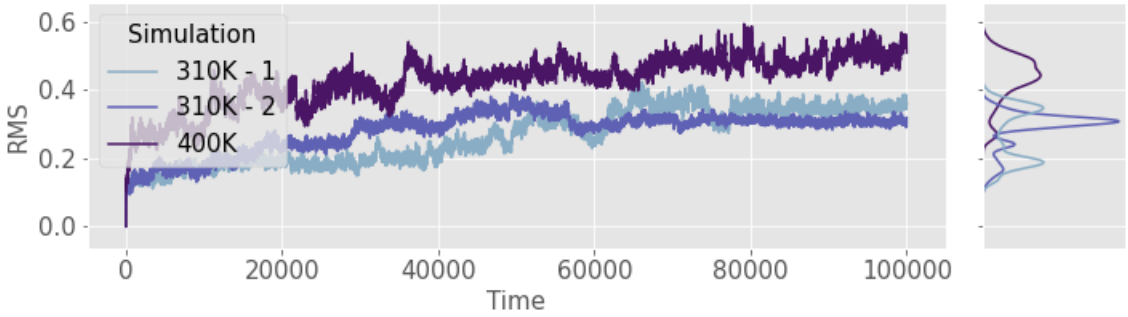


Figure 2: RMSD of the whole protein complex.

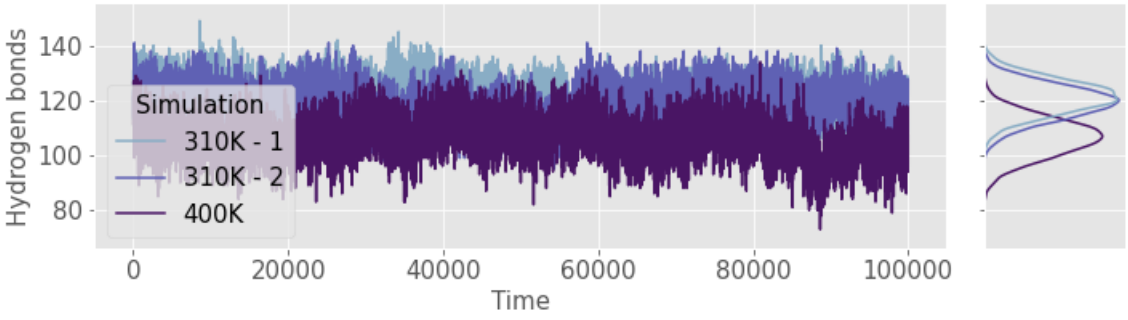


Figure 3: Number of hydrogen bonds in the system over time.

3.1.3 Stability of chain A

In Figure 2 the distance between the C and N terminus of chain A is displayed over simulation time. In *310K - 1* simulation the distance remain relatively stable - $\approx$ 3.80 nm. In *310K - 2* simulation we can see a transition from $\approx$ 4 nm to $\approx$ 3.43 nm. *400K* simulation is less stable - we can observe fluctuations around the value of $\approx$ 3.38 nm with sharp drops to $\approx$ 2 nm. Intuitively, during protein unfolding we would expect the distance between the terminus to increase, however we see the opposite - the distance shrinks. Interestingly, when observing the progression of hydrogen bonds in chain A (Figure 5) we observe a state transition in *400K* simulation from $\approx$ 20 hydrogen bonds to $\approx$ 15 at $\approx$ 40,000 ns. This signifies that once those bonds responsible for α-helix structure are lost they do not recover, therefore this cannot explain the distance drops seen in Figure 4. Furthermore, albeit the distance between C and N termini in *310K - 2* simulation was on average lower than in *310K - 1*, the

number of hydrogen bonds was roughly the same at ≈ 20.75. To understand the mechanism behind the shrunken distance a visual inspection with `VMD` is useful.

The results of the visual investigation of the *400K* simulation can be seen in Figure 10. In (b) we can see a state that explains the shrunken distance between the terminus as the protein is bent and the terminus are closer to each other. In (c) we can see a state that represents the peaks in Figure 4 as both termini have unfolded and moved away from each other. Another interesting thing to notice is that the C terminus unfolds before the N terminus.

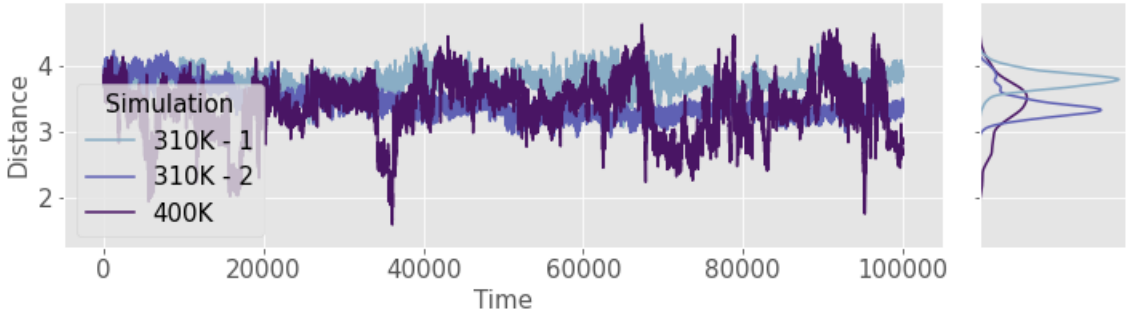


Figure 4: Distance between C-terminus and N-terminus in chain A over time.

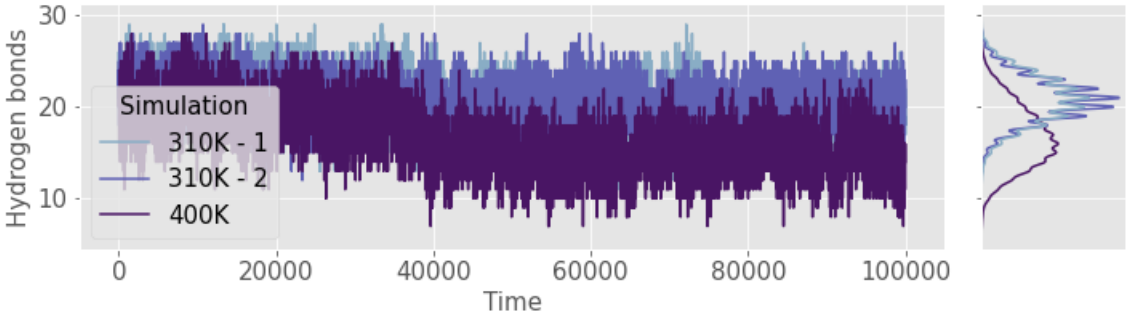
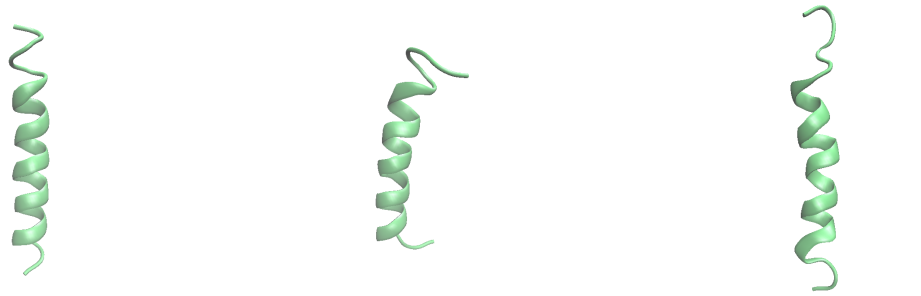


Figure 5: Number of hydrogen bonds in chain A.



(a) Initial state (b) "Bent" state (c) Unfolded state

Figure 6: Chain A during 400 K simulation.

3.1.4 Integrity of the complex

In Figure 7 the distance between chain A and the rest of the complex is displayed over simulation time. As in subsection 3.1.2, we can once again see a distinct state transition from $\approx$ 0.8 nm to $\approx$ 0.98 in *310K - 1* simulation - at around 60,000 ns. Interestingly in *310K - 2* simulation there is a gradual decrease of distance from $\approx$ 0.8 nm to $\approx$ 0.74 nm. As expected, the distance is bigger in *400K* simulation - it undergoes a transition from $\approx$ 0.8 to $\approx$ 0.88, however the values fluctuate substantially towards the end of the simulation. The small distance change signify the stability of the interactions that keep the 6 chains together.

As the distance changes were small, visual inspection did not provide any insight.

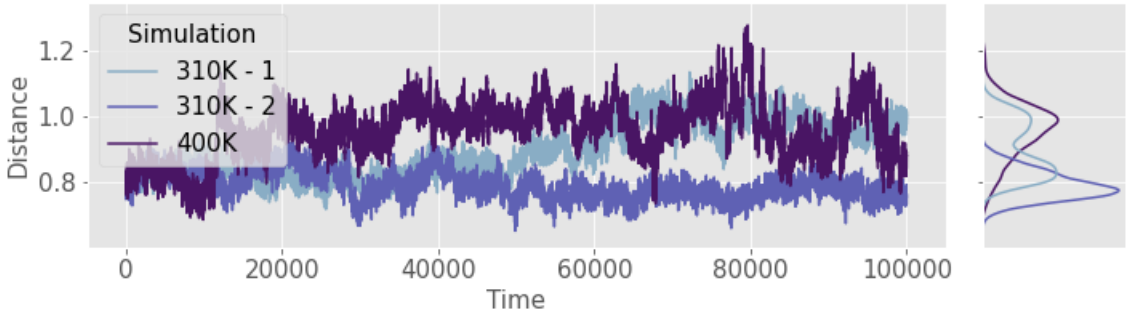
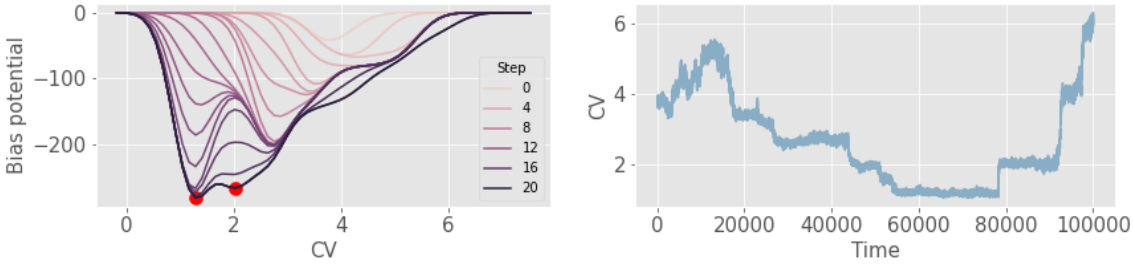


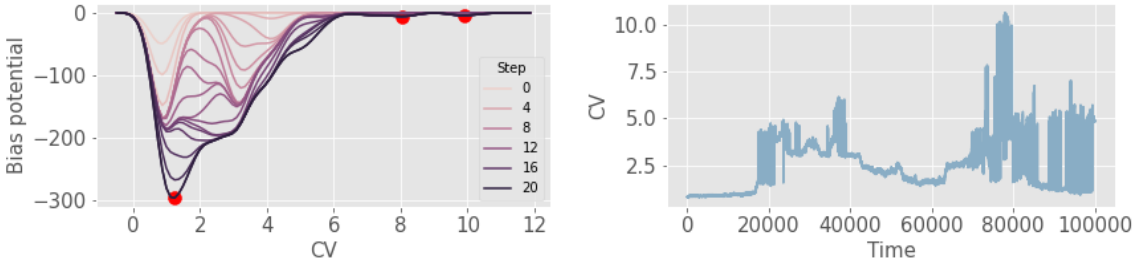
Figure 7: Distance between chain A and rest of the complex over time.

3.2 Metadynamics



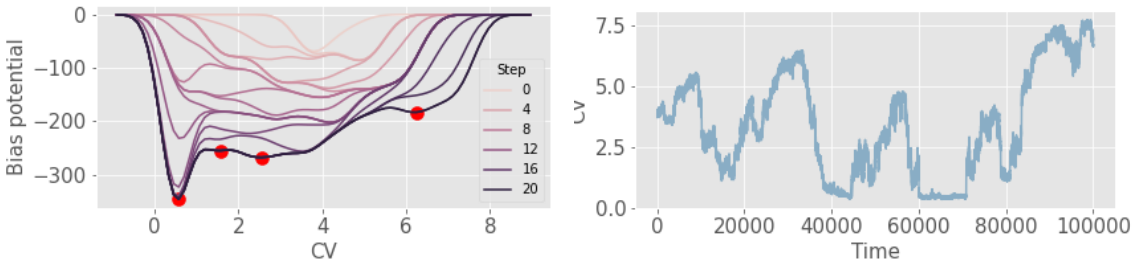
(a) Free energy profile (b) Progression of CV

Figure 8: Metadynamics results. Collective variable - distance between C and N termini.



(a) Free energy profile (b) Progression of CV

Figure 9: Metadynamics results. Collective variable - distance between chain A and rest of the complex.



(a) Free energy profile

(b) Progression of CV

Figure 10: Metadynamics results (increased hill height). Collective variable - distance between C and N termini.

4 Conclusions

blah.

References

A.R. Leach. 4.05 - ligand-based approaches: Core molecular modeling. In John B. Taylor and David J. Triggle, editors, *Comprehensive Medicinal Chemistry II*, pages 87–118. Elsevier, Oxford, 2007. ISBN 978-0-08-045044-5. doi: https://doi.org/10.1016/B0-08-045044-X/00246-7. URL `https://www.sciencedirect.com/science/article/pii/B008045044X002467`.

Qinghua Liao. Chapter four - enhanced sampling and free energy calculations for protein simulations. In Birgit Strodel and Bogdan Barz, editors, *Computational Approaches for Understanding Dynamical Systems: Protein Folding and Assembly*, volume 170 of *Progress in Molecular Biology and Translational Science*, pages 177–213. Academic Press, 2020. doi: https://doi.org/10.1016/bs.pmbts.2020.01.006. URL `https://www.sciencedirect.com/science/article/pii/S187711732030017X`.

A Physical properties

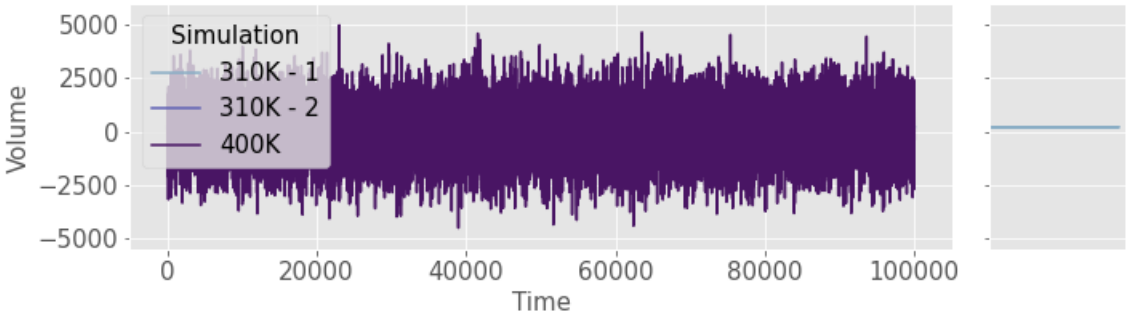


Figure 11: Volume over time.

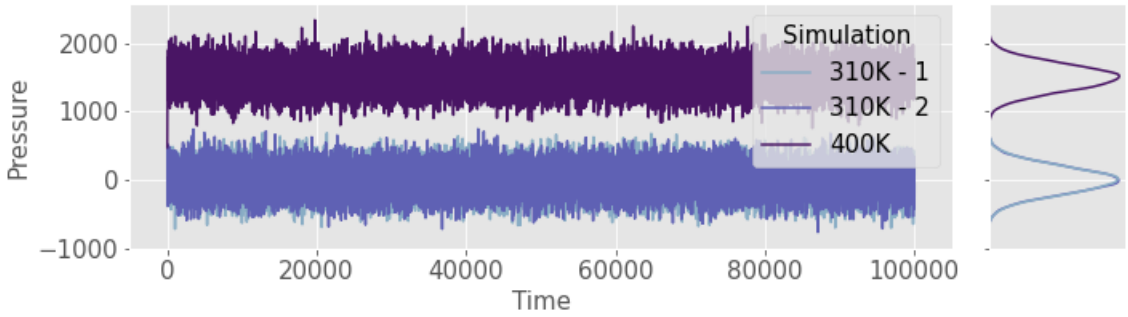


Figure 12: Pressure over time.

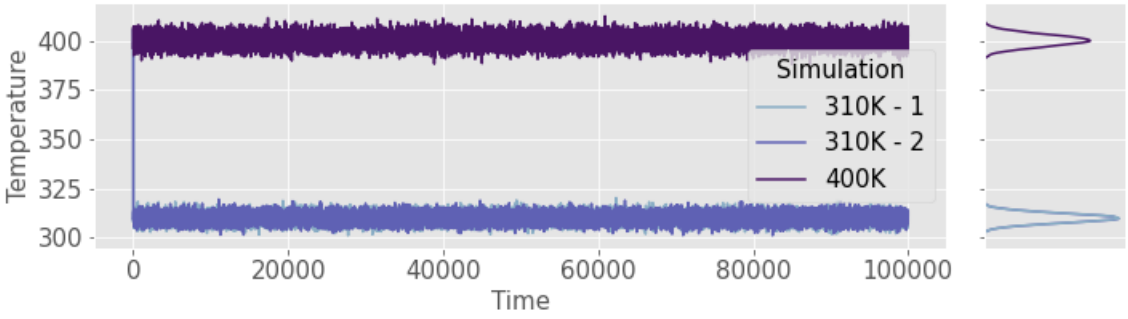


Figure 13: Temperature over time.