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

6 June 2021

UNIVERSITY
OF AMSTERDAM

Abstract

Structural characterisation of SARS-CoV-2 spike protein is paramount in the fight against COVID-19. In this paper molecular dynamics simulations were utilised to explore the conformational space of the SARS-CoV-2 spike protein. Simulations at body temperature demonstrated the stability of the complex, however a simulation at an elevated temperature demonstrated the unfolding process of an α-helix. With metadynamics it was identified that the chain A is most stable in a compact "bent" state, where the termini have moved closer to each other. Furthermore, it was also identified that the protein is stable when chain A and X have detached from the complex. Lastly, it was shown that by increasing the hill height parameter in the metadynamics simulations we could explore a wider collective variable space and identify extra metastable states.

1 Introduction

Since the outbreak of COVID-19 caused by SARS-CoV-2 at the end of 2019 a swift and highly coordinated scientific research infrastructure has been put in place all over the world. A lot of this effort has been concentrated on the structural characterisation of the main components of the virus (Moreira et al. [2020]). As the spike protein is a key target for development of vaccines, therapeutic antibodies, and diagnostics - understanding the conformational space of the spike protein is of paramount importance in the fight against the pandemic (Wrapp et al. [2020]).

SARS-CoV-2 utilises the spike protein to gain entry into the host cells (Xia et al. [2020]). It is made up of six chains - A, B, C, X, Y and Z - that form a 6 α-helix complex. The analysis in this paper concentrates on a fragment of this spike protein - as displayed in Figure 1. A, B and C chains are made up of 26 residues, while X, Y and Z chains have 18 residues.

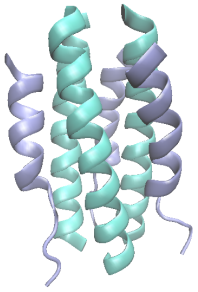


Figure 1: Fragment of SARS-CoV-2 spike protein fusion domain. Cyan - chain A, B and C; purple - chain X, Y, Z.

Two techniques - molecular dynamics and metadynamics - were applied in this paper to explore the conformational space of the SARS-CoV-2 spike protein. This was done by concentrating the analysis on chain A - firstly, by investigating specific properties (such as RMSD and distance between termini) at body temperature and at elevated temperature. The most stable confirmations of the protein were identified via metadynamics simulations by following the evolution of free energy profile.

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 the spike protein complex two simulations were carried out at a temperature of 310 K (body temperature) and one at 400 K. Two simulations at 310 K were conducted to account for the stochastic nature of the MD simulations with each simulation drawing a different sample of velocities assigned to the atoms based on the Boltzmann distribution at 310 K. All simulations were carried out using `gromacs` software package.

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
Integration time-step (fs)	2	2
Coulomb method	PME	PME
Van der Waals method	cut-off (1.1 nm)	cut-off (1.1 nm)

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 fol-

lowed by energy minimisation to mitigate unwanted interactions between atoms. 43 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 MD 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.

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

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

conformational space and thus provides more detailed insight. Simulation 2 explores the free energy profile of the chain A detachment from the rest of the complex.

2.3 Analysis

A number of measurements were extracted from the simulations to quantify the processes observed. 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.

Hydrogen bonds Hydrogen bonds in the whole complex and chain A over time were extracted using `gromacs`' `hbond` 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 5000 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 ps 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 ps.

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.

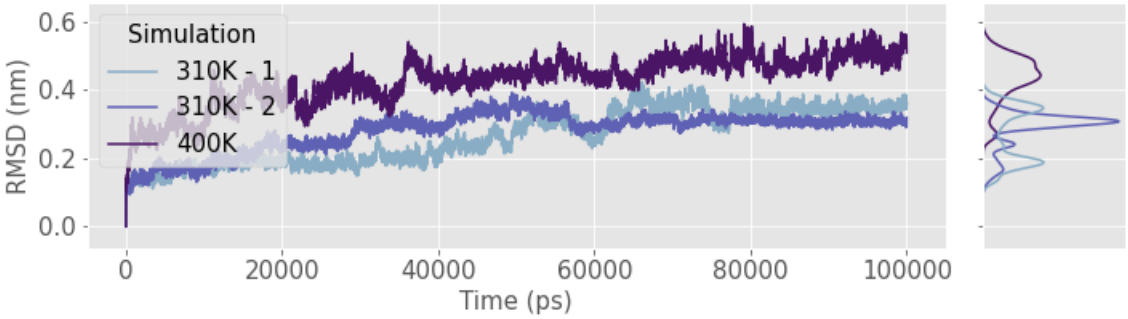


Figure 2: RMSD of the whole protein complex.

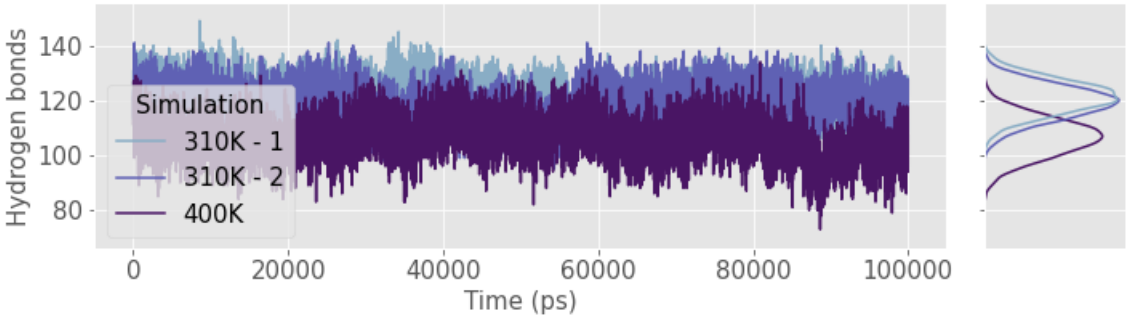


Figure 3: Number of hydrogen bonds in the system over time.

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 remains 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 ps. 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 average number of hydrogen bonds was roughly the same at $\approx$ 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 6. In (b) we can see a state that explains the shrunken distance between the termini 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 aspect is that the C terminus unfolds before the N terminus.

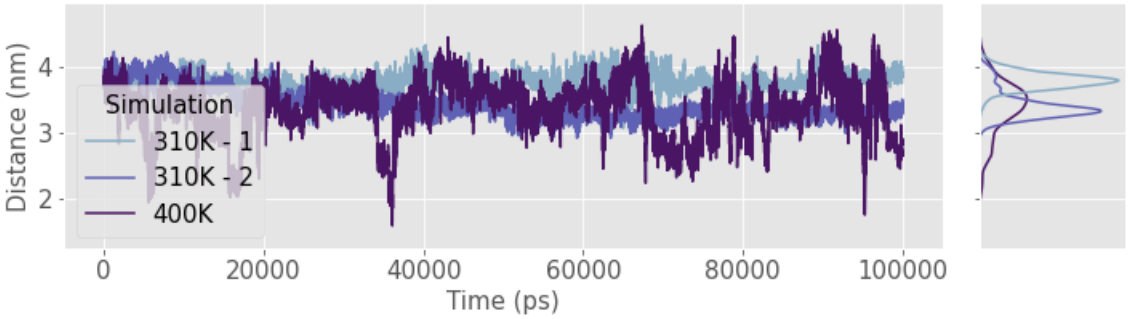


Figure 4: Distance between C-terminus and N-terminus in chain A over time.

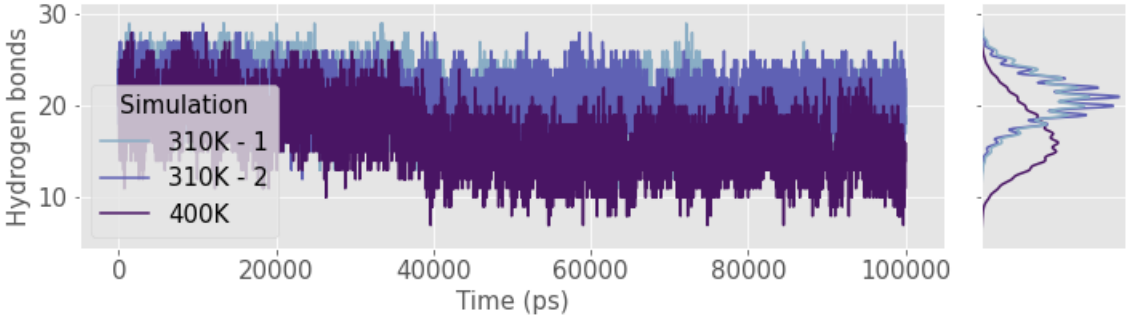
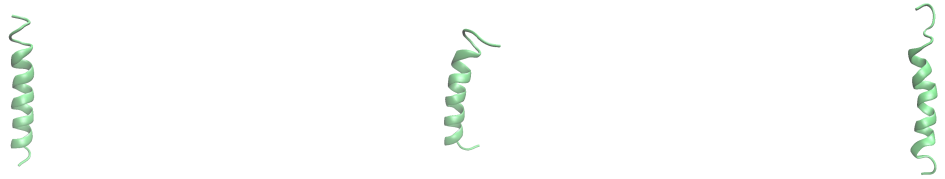


Figure 5: Number of hydrogen bonds in chain A.



(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 the simulation time. As in subsection 3.1.2, we can once again see a distinct

state transition from ≈ 0.8 nm to ≈ 0.98 nm in *310K - 1* simulation - at around 60,000 ps. Interestingly in *310K - 2* simulation there is a gradual decrease of distance from ≈ 0.8 nm to ≈ 0.74 nm. As expected, the distance is bigger in *400K* simulation - it undergoes a transition from ≈ 0.8 nm to ≈ 1 nm, 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.

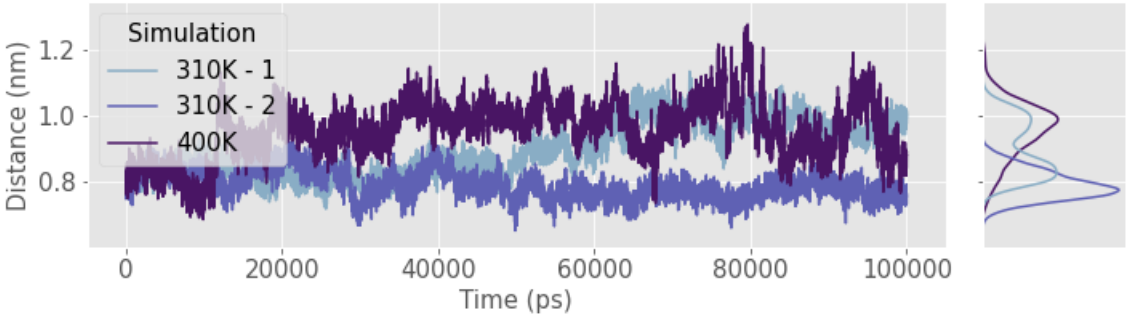


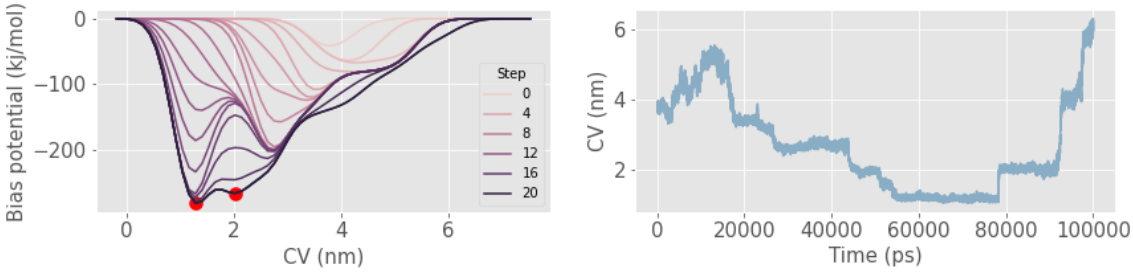
Figure 7: Distance between chain A and rest of the complex over time.

3.2 Metadynamics

3.2.1 Distance between C and N termini in chain A

The free energy profile progression and CV over simulation time is plotted in Figure 8. It can be observed that the free energy profile starts by sampling areas close to the starting structure of the chain A (at a distance of ≈ 4 nm). Towards the end of the simulation we can observe that two minima (step = 14) are emerging. In the final step we have two minima - metastable state at 2.02 nm and global minima at 1.28 nm. This a substantially reduced distance in comparison to the starting distance of 3.77 nm. The bias potential difference between the two minima is small - 15.07 kj/mol. As the shape of the free energy profile during last steps (16-20) goes through a significant change - the metastable minima becomes smaller - we can not draw a conclusion that the free energy profile has converged.

Visual representation of chain A at a state representing global minima and metastable state can be found in Figure 9. It can be seen that tails of the protein tend to move towards each other, however in the metastable state N terminus points away - hence the higher distance.



(a) Free energy profile

(b) Progression of CV

Figure 8: Metadynamics results. CV - distance between C and N termini. Red circles - (local) minima. Step - 5000 deposited hills.

(a) Initial state

(b) Metastable state

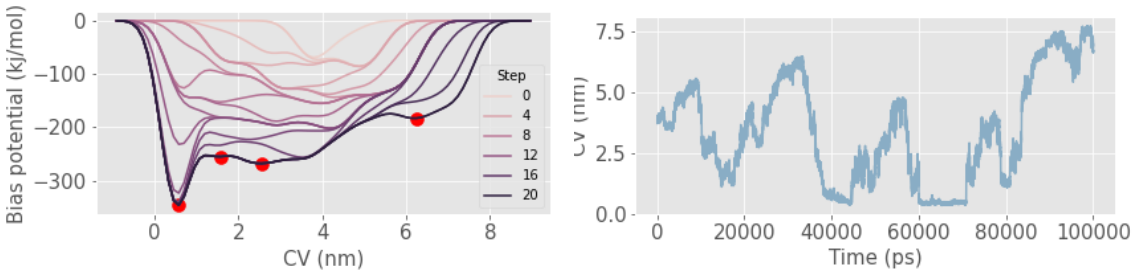
(c) Global minima state

Figure 9: Chain A in initial, global minima and metastable states during MetaD simulation.

Increased hill height The results of the metadynamics simulation with a doubled hill height is displayed in Figure 10. In the beginning the free energy profile progressing similarly to the one observed in Figure 8 - by sampling areas close to the starting structure of the chain A (at a distance of ≈ 4 nm). However, the profile progresses into a wide profile with no clear minimas. Around step = 16 a distinct minima is identified at 0.58 nm. In the final step there are four minimas: 0.58 nm (global), 1.57 nm (metastable), 2.55 nm (metastable) and 6.25 nm (metastable). Increasing hill size identified more minima and explored a larger CV-space, therefore, as expected, the increased hill size was beneficial for identifying a higher number of stable states. As the free energy profile during last steps (15-20) was still identifying significantly different energy minimas - we can not draw a conclusion that the free energy profile has converged.

Visual representation of chain A at a state representing global minima and metastable states can be found in Figure 11. The identified global minima state

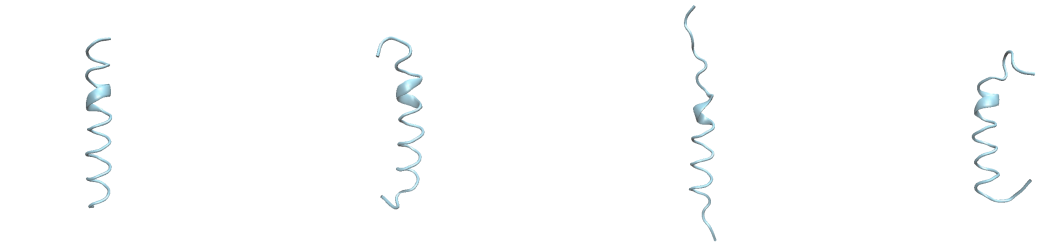
is very similar to the one found in Figure 9. However, an extra "stretched" state is identified where both termini have moved further away from each other (c).



(a) Free energy profile

(b) Progression of CV

Figure 10: Metadynamics results (increased hill height). Collective variable - distance between C and N termini. Red circles - (local) minima. Step - 5000 deposited hills.



(a) Initial state

(b) Metastable state (CV = 2.55 nm)

(c) Metastable state (CV = 6.25 nm)

(d) Global minima state

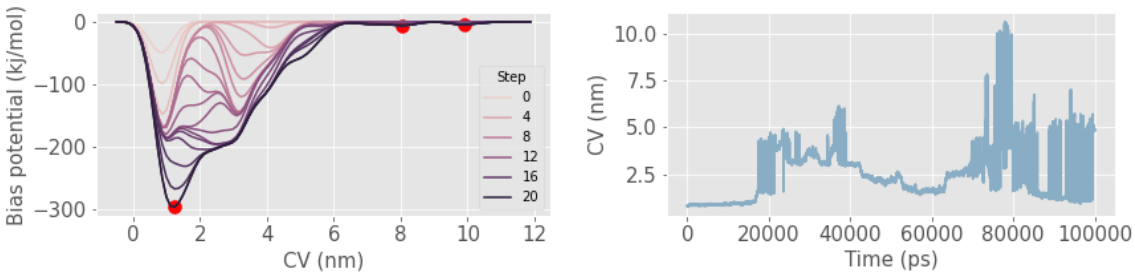
Figure 11: Chain A in initial, global minima and metastable states during MetaD simulation with increased hill height.

3.2.2 Distance between chain A and the rest of the complex

The results of the metadynamics simulation with CV as distance between chain A and the rest of the complex is displayed in Figure 12. It can be observed that the free energy profile starts by sampling areas close to the starting structure (at a distance of $\approx$ 1 nm). In the middle of the simulation we can observe that two minima (step = 8) are emerging. In the final step we have one global minima at 1.25 nm. This is a 58% increase from the starting distance of 0.79 nm. In the plot there are also two local minima at 8.04 nm and 9.90 nm - however, as they are very unstable, they will

not be discussed. Towards the end of the simulation the free energy profile shape does not change substantially, therefore the system seems to be converging.

Visual representation of the complex at a state representing global minima and initial state can be found in Figure 13. It can be seen that chain A detaches from the complex, however chain A remains attached to chain X.



(a) Free energy profile

(b) Progression of CV

Figure 12: Metadynamics results. CV - distance between chain A and rest of the complex. Red circles - (local) minima. Step - 5000 deposited hills.

(a) Initial state

(b) Global minima state

Figure 13: Protein in initial and global minima states during MetaD simulation. Chain A in grey.

3.2.3 Comparison with Molecular Dynamics at 400 K

Distance between termini C and N in chain A Both MD and MetaD explored the "bent" state of chain A. However, in MD simulations this state was beraly explored, while on the other hand the "bent" state was identified as the most stable state of chain A with MetaD. Furthermore, in both simulation C terminus unfolded first - therefore, the C terminus can be regarded as one of the most unstable points of the protein.

Chain A detachment from the complex MD simulations did not find chain A to detach from the complex - the integrity of all 6 chains was preserved throughout the simulation. This is in contrast to what was found with MetaD where the most stable state (albeit the distance was relatively similar to the one in MD simulation) was chain A detached from the complex together with chain X.

4 Conclusions and outlook

In this paper molecular dynamics simulations were applied in order to investigate the stability of SARS-CoV-2 spike protein. Via MD simulations at body temperature it was observed that the protein preserves most of its' initial structural features signifying the stability of interactions between chains. In a simulation at 400 K, chain A was unfolding and exploring a "bent" state, however the integrity of the 6-chain complex was still preserved. Via MetaD simulations it was identified that chain A is most stable in a compact "bent" state. Doubling the hill height in MetaD simulation allowed identification of an extra metastable state where termini C and N of chain A are stretched away from each other. Furthemore, MetaD simulations identified that the protein is most stable when Chain A and X have detached from the complex. Longer simulations are needed to identify a converged free energy profile from MetaD simulations - albeit it seemed that the profiles were converging, a concrete conclusion can not be drawn that the free energy profiles have converged in any of the performed MetaD simulations.

To better understand the process of chain A unfolding it may be useful to choose a different CV - number of hydrogen bonds inside chain A could be a good choice to overcome the "bent" state. Other altered metaD techniques could be applied to further explore the conformational space of the spike protein - e.g. well-tempered metadynamics (WTMetaD). WTMetaD allows quicker convergence of the free energy profile by concentrating the simulation on physically relevant regions of the conformational space by applying adaptive bias (Barducci et al. [2008]). Another technique that could be applied here is transition path sampling (TPS). TPS concentrates on simulating rare transitional events - this way the number of transitions in the simulation increases, thus allowing calculation of rate constants and pathways (Chong et al. [2017]). E.g., we could apply TPS to quantify the unfolding process of chain A - generate multiple transitional pathways to better understand the mechanism of unfolding. Furthermore, alchemical sampling could be applied to calculate free energy difference between folded and unfolded states of chain A. A perturbation parameter could be coupled with the Hamiltonian to drive the spike protein to difference states, thus allowing us to estimate free energy differences (Seeliger and de Groot [2010]).

References

Rodrigo A. Moreira, Mateusz Chwastyk, Joseph L. Baker, Horacio V. Guzman, and Adolfo B. Poma. Quantitative determination of mechanical stability in the novel coronavirus spike protein. *Nanoscale*, 12(31):16409–16413, 2020. doi: 10.1039/d0nr03969a.

Daniel Wrapp, Nianshuang Wang, Kizzmekia S. Corbett, Jory A. Goldsmith, Ching-Lin Hsieh, Olubukola Abiona, Barney S. Graham, and Jason S. Mclellan. Cryo-em structure of the 2019-ncov spike in the prefusion conformation. 2020. doi: 10.1101/2020.02.11.944462.

Shuai Xia, Yun Zhu, Meiqin Liu, Qiaoshuai Lan, Wei Xu, Yanling Wu, Tianlei Ying, Shuwen Liu, Zhengli Shi, Shibo Jiang, and et al. Fusion mechanism of 2019-ncov and fusion inhibitors targeting hr1 domain in spike protein. *Cellular Molecular Immunology*, 17(7):765–767, 2020. doi: 10.1038/s41423-020-0374-2.

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`.

Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Physical Review Letters*, 100(2), 2008. doi: 10.1103/physrevlett.100.020603.

Lillian T Chong, Ali S Saglam, and Daniel M Zuckerman. Path-sampling strategies for simulating rare events in biomolecular systems. *Current Opinion in Structural Biology*, 43:88–94, 2017. ISSN 0959-440X. doi: https://doi.org/10.1016/j.sbi.2016.11.019. URL `https://www.sciencedirect.com/science/article/pii/S0959440X16302068`. Theory and simulation • Macromolecular assemblies.

Daniel Seeliger and Bert L. de Groot. Protein thermostability calculations using alchemical free energy simulations. *Biophysical Journal*, 98(10):2309–2316, 2010. ISSN 0006-3495. doi: https://doi.org/10.1016/j.bpj.2010.01.051. URL `https://www.sciencedirect.com/science/article/pii/S000634951000216X`.

A Physical properties

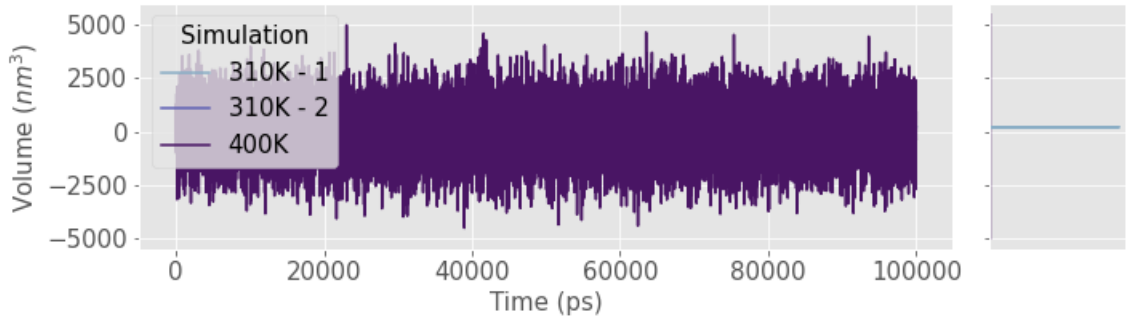


Figure 14: Volume over time.

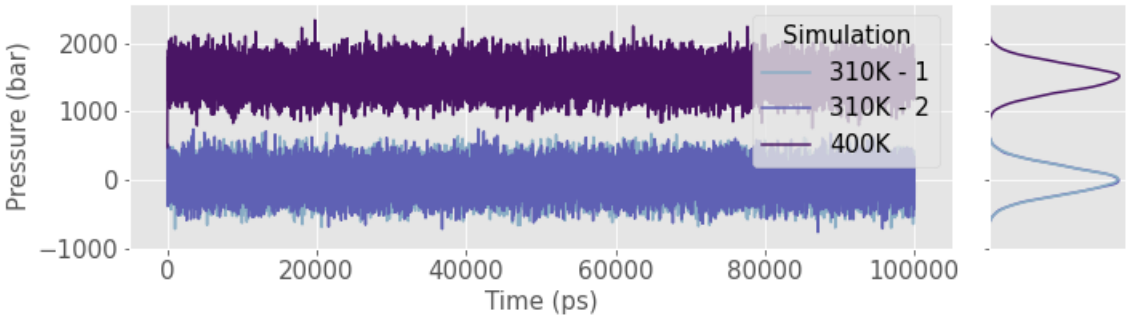


Figure 15: Pressure over time.

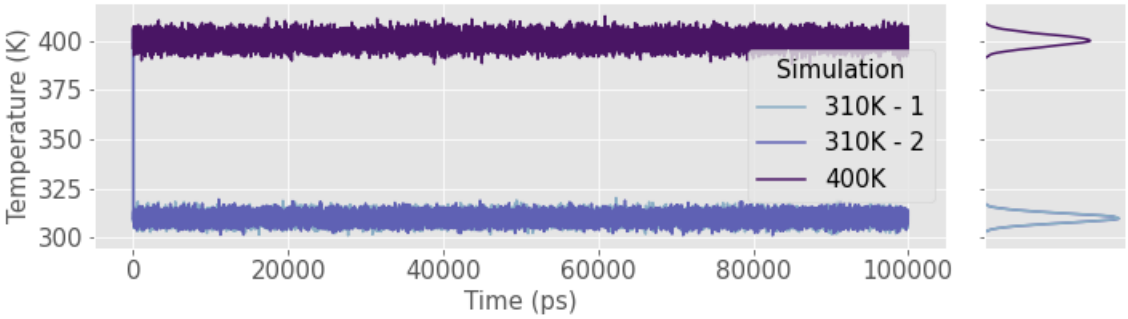


Figure 16: Temperature over time.