5-2022

Spike Protein Structural Dynamics of SARS-CoV-2 Coronaviruses Studied Using Molecular Dynamics

William Strickland
University of Arkansas, Fayetteville

Spike Protein Structural Dynamics of SARS-CoV-2 Coronaviruses Studied Using Molecular Dynamics

An Honors Thesis submitted in partial fulfillment of the requirement for Honors Studies in Biology

By

William Strickland

Spring 2022

Biology

J. William Fulbright College of Arts and Science

The University of Arkansas

Acknowledgements

I would first like to thank Dr. Mahmoud Moradi for his guidance and leadership in serving as my Honors Thesis Mentor. He has invested a great amount of time in my work, opened new opportunities for me, and has allowed me to gain new skills and expertise in topics that have transformed my perception of lab research. I have no doubt that the experience and techniques I have learned in the lab will help me later in my journey towards my professional career.

I would like to thank Dr. Mortaza Derakhshani Molayousefi for serving as another fantastic mentor for me during my time in the Moradi lab. Dr. Mortaza's investments and sacrifices assisted me in overcoming multiple obstacles that arose in the lab, and his insight and direction were crucial towards having a wonderful research experience. Confronting challenges such as oral and written computational research presentations can be particularly daunting, but Dr. Mortaza provided me with the necessary guidance to ensure that I was thoroughly prepared.

I would also like to thank the other members of the Moradi Research Lab for all their help and support in my research. Shadi Badiee, Matthew Brownd, and Adithya Polasa all held significant roles in helping me in the lab, displayed through training me and giving me helpful suggestions. I would also like to thank the rest of the group for all their contributions as well.

I would finally like to thank the University of Arkansas Honors College for their funding and support of my project and the department of chemistry and biochemistry for providing the resources and facilities to make it possible. I also want to thank the members of my thesis defense committee for taking the time to serve on my committee: Dr. Moradi, Dr. Beaulieu, Dr. Naithani, and Dr. Dowdle.

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health under grant 1R15GM139140-01.

Further Ackowledgements:

Texas Advanced Computing Center

XSEDE Extreme Science and Engineering Discovery Environment

National Science Foundation and the National Institutes of Health

AMD COVID-19 HPC Fund

Arkansas High Performing Computing Center

Pittsburgh Supercomputing Center

Table of Contents

Acknowledgements **1**

I. Abstract **3**

II. Introduction **4**

III. Methods **11**

IV. Results **14**

V. Concluding Remarks **21**

VI. References **24**

I. Abstract

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has overwhelmingly impacted the global population, accounting for millions of confirmed infections and deaths over the last year. The virus's influence on the health and safety of individuals, the economy, and daily life has been disruptive and devastating. While SARS-CoV-2 and SARS-CoV-1, two closely related members of the SARS coronaviruses, have shown the ability to cross the species barrier and infect humans, SARS-CoV-2 has predominantly been the virus responsible for the number of infections presently known. SARS-CoV-2 has also proven to be volatile, as many variants have recently materialized based on amino acid structure mutations. Understanding the differential behavior of the SARS coronaviruses and the many SARS-CoV-2 variants may provide insight into interpreting how the spreading of COVID-19 occurs and could lead to further intuition and discovery. Specifically, studying the structural dynamics of spike proteins that play a crucial role in host cell receptor recognition could expedite the development of vaccines and antivirals that identify sites as potential drug targets.

All variants of SARS-CoV-2 recognize the same receptor in humans, yet oftentimes the variants themselves exhibit varying degrees of characteristics such as transmissibility and infectivity. It is implied that the spike proteins, which are the most variable region in the entire genome, may potentially be a source of the different traits these variants present. Specifically, in the lab, we aimed to investigate the activation process of the spike protein and the conformational changes that must occur for the receptor-binding motif (RBM) to be made available for binding to the human receptor (ACE2). We analyzed and targeted the D614G mutation present in many of the SARS-CoV-2 variants and compared it to the differential

characteristics present in the wild-type form of the virus. To visualize a detailed account of prefusion spike protein binding to ACE2, we used an extensive set of equilibrium microsecond-level all-atom molecular dynamics simulations. These models are both atomistic and dynamic, allowing us to visualize differences in protein conformation over time at remarkable degrees. The differential behaviors analyzed aided in determining the dynamical changes of the spike proteins and not just their inactive and active states. We determined that the D614G mutation altered sets of interactions throughout the spike protein, potentially resulting in different structural conformations. We also concluded that the D614G variant favored an active state due to increased relative stability, while the original Wild Type variant preferred an inactive state. These results suggest that the D614G mutation may cause variability in the activation mechanisms and stability of virus variants, potentially playing a crucial role in determining the differential characteristics that the viruses possess.

II. Introduction

Over the last two years, coronavirus disease 2019 (COVID-19) has swept the globe, infecting millions worldwide. Originally appearing in Wuhan, China, it has affected individuals in many ways as it has disturbed daily life, put health in jeopardy, and disrupted the economy (Harapan et al., 2020). On January 20, 2020, the World Health Organization declared a Public Health Emergency of International Concern due to the COVID-19 pandemic, a disease caused by novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Harapan et al., 2020). SARS-CoV-2, along with other pathogenic species of coronaviruses such as SARS-CoV-1 and MERS-CoV, is classified as a Betacoronavirus among the four genera: Alphacoronavirus, Betacoronavirus, Gammacoronavirus, and Deltacoronavirus (Ciotti et al., 2020). SARS-CoV-2,

SARS-CoV-1, and MERS-COV are all originally zoonotic in origin and have the potential to cross the species barrier and infect humans, causing severe respiratory illnesses and fatalities (Hasoksuz et al., 2020). Despite striking similarities between the SARS-CoV-1 and SARS-CoV-2 coronaviruses, the pandemic caused by the SARS-CoV-2 virus has been significantly more widespread than the 2003 SARS epidemic caused by SARS-CoV-1. As of March 2022, infections of COVID-19 due to SARS-CoV-2 have expanded to more than 450,000,000 cumulative cases and over 6,000,000 cumulative deaths (World Health Organization, 2022).

Recognition Mechanism of Coronavirus via Spike Protein Binding

Undoubtedly, the impact that SARS-CoV-2 has had on the population has been extensive. Therefore, it is necessary to analyze the mechanism by which coronaviruses attach to and gain entry into human cells. Coronaviruses enter the host cell through numerous transmembrane spike proteins that protrude from their lipid bilayer (Ciotti et al., 2020; Tortorici & Veesler, 2019; Duan et al., 2020). These transmembrane spike proteins are composed of three protomers which constitute a protein trimer (Ke et al., 2020). The spike protein trimers are how SARS-CoV-2 can bind to receptors on the surface of target cells; in this case, these trimers attach to angiotensin-converting enzyme 2 (ACE2) in humans and mediate viral uptake and fusion (Ke et al., 2020; Walls et al., 2020; Hoffman et al., 2020; Shang et al., 2020; Wang et al., 2020; Wrapp et al., 2020).

Structurally, the spike protein is comprised of two subunits (S1 and S2) that together are responsible for viral entry into the host cell (Ciotti et al., 2020). However, each subunit possesses its respective role in the infection process: the S1 subunit is culpable for host cell receptor binding, while the S2 subunit is responsible for cellular membrane fusion (Kirchdoerfer et al.,

2016; Millet & Whittaker, 2015). In order for the spike proteins of the coronaviruses to bind, conformational changes must occur for the receptor-binding domain (RBD) to be made available to attach to the human receptor ACE2. Figure 1 below provides a visual of the coronavirus binding process to host cells. Figure 2 depicts a linear model of a spike protein that highlights its different regions and the residue ranges that govern them.

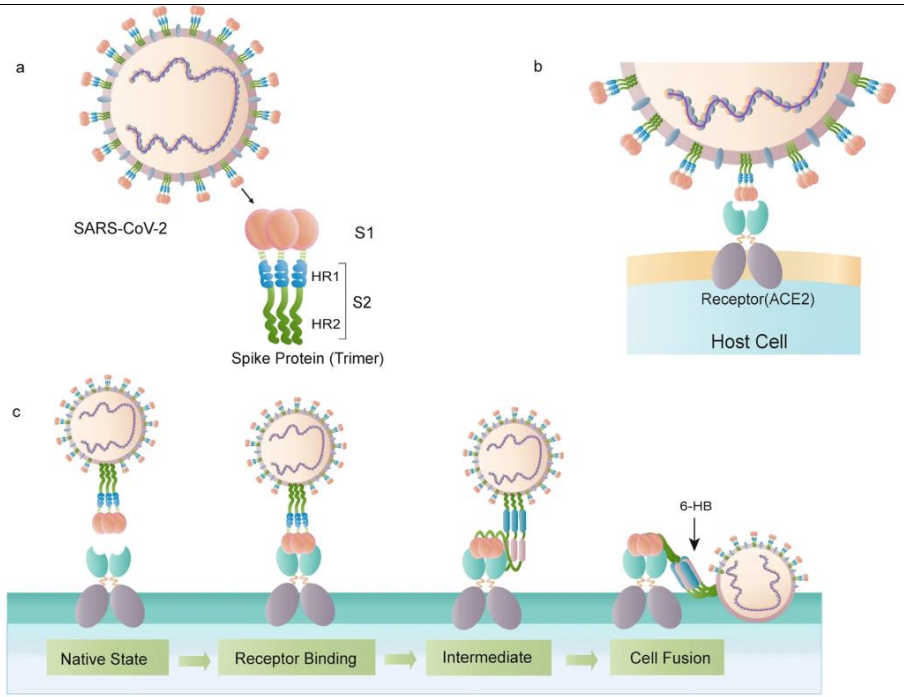


Figure 1: Schematic of the SARS-CoV-2 Spike Protein (Huang et al., 2020)

Shown above is the binding process between the SARS-CoV-2 spike protein and human ACE2 receptor. Image A details the structure of the spike proteins that are located on the outer membrane of the SARS-CoV-2 virus and provides an overview of the overall appearance of the virus itself. The three protomers that comprise the spike protein are shown, as well as the S1 and S2 subunits that assist in receptor binding and cellular membrane fusion. Image B depicts the binding between the spike protein and host cell receptor, while Image C portrays the process of virus-cell fusion that is mediated by the spike protein (Huang et al., 2020).

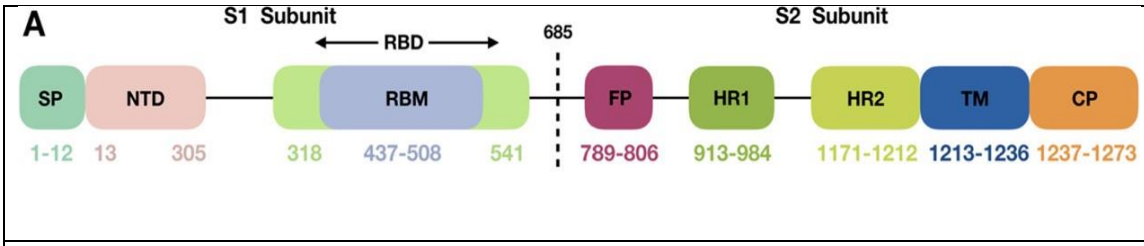


Figure 2: Diagram of functional domains in the SARS-CoV-2 Spike Protein (Woo et al., 2020)

Shown above is a diagram of the SARS-CoV-2 spike protein that illustrates its different domains. The S1 and S2 subunits are visualized along with their respective constituents. Here, the RBD region is shown to extend from residue 318 to residue 541, and the Receptor Binding Motif (RBM) is detailed to encompass residues from 437 to 508. The RBM is contained in the RBD and aids in mediating contact with the ACE2 human receptor.

It has previously been exhibited that the respective receptor binding domains of both SARS-CoV-1 and SARS-CoV-2 interact with the human ACE2 receptor (Fang et al., 2005; Xing et al., 2013; Wenhui et al., 2003; Tai et al., 2020). However, the molecular reasoning behind SARS-CoV-2 being considerably more widespread and transmissible is widely unknown. When comparing the two viruses, their structural and functional features are reasonably similar (Tortorici & Veesler, 2019). Furthermore, genome sequences gathered from SARS-CoV-2 spike proteins have displayed a noticeably shared sequence identity of approximately 79% with those of SARS-CoV-1 (Zhou et al., 2020; Govind et al., 2022, Hu et al., 2021). Therefore, it comes as a bit of a surprise that the RBD of the CoV-2 spike protein has been shown to have a higher affinity to the host ACE2 receptor than the CoV-1 spike protein (Tai et al., 2020). Keeping in mind that CoV-2 is more transmissible than CoV-1, it can be inferred that there potentially is some variability within the spike proteins of each coronavirus that causes them to have varying degrees

of characteristics such as this transmissibility (Walls et al., 2020; Wu et al., 2012; Petersen et al., 2020; Ferretti et al., 2020).

Many variants of SARS-CoV-2 have also been found to display different intensities of transmissibility and infectivity. Recently, many variants of the virus have shown increased contagiousness, including the Alpha, Beta, Delta, and Omicron variants, as well as other variants of interest (Campbell et al., 2021; Karim & Karim, 2021). The Delta variant particularly showed a significant increase in the effective reproduction number at +100% compared to non-variants (Campbell et al., 2021; Department of Health and Social Care, 2021). In addition, the Delta variant has been shown to be approximately 63-167% more transmissible than the Alpha variant (Earnest et al., 2022). Due to its higher viral load, longer duration of infectiousness, and ability to elude natural immunity, the Delta variant provides an excellent example of how some of the SARS-CoV-2 variants possess different characteristics that lead to varying levels of viral competence (Karim & Karim, 2021; Luo et al., 2021; Wang et al., 2021; Townsend et al., 2021).

The differing levels of infectivity and transmissibility among variants could be due to mutations in the virus's genome. Past studies have demonstrated that several regions of the SARS-CoV-2 genome have the potential to accumulate mutations, including the RBD region in spike proteins which has shown to be particularly susceptible (Kaushal et al., 2020; van Dorp et al., 2020; Isabel et al., 2020; Koyama et al., 2020; Guruprasad, 2021). One mutation that was consistently present among variants was D614G; in many studies, it was the most observed individual mutation in the genome (Guruprasad, 2021; Daniloski et al., 2021; Hou et al., 2020; Koyama et al., 2020). The D614G mutation was also one of the most frequent substitutions in presently spreading SARS-CoV-2 strains, and has been discovered in the Alpha, Beta, Gamma,

Delta, and Omicron variants specifically (Wang et al., 2021; "SARS-CoV-2 Variants of Concern," 2022; Jackson et al., 2021). D614G can be characterized as a spike protein mutation that is located in the S1 subunit and directly associated with the S2 subunit (Zhang et al., 2020; Morais et al., 2020; Laha et al., 2020; Commandatore et al., 2020). This mutation was found to play a role in the geographical segregation of subtypes of SARS-CoV-2 and ultimately became the predominant variation with time (Cortey et al., 2020; Morais et al., 2020, Furuyama et al., 2020). The dramatic increase in frequency demonstrated over a short period of time perhaps suggests that the mutation induces a transmission advantage (Zhang et al., 2020). In fact, recent studies have shown that variants that possess the D614G mutant have a selective advantage. D614G has proved to increase infectivity in vitro, suggest a higher viral load in patients, and alter receptor binding conformation to increase the likelihood of ACE2 binding (Volz et al., 2021; Korber et al., 2020; Yurkovetskiy et al., 2020). It has also shown a more efficient replication, transduction of cells, and competitive fitness than the wild-type virus (Hou et al., 2020; Daniloski et al., 2021).

Considering the numerous mutations, such as D614G, and variants of SARS-CoV-2, it is imperative to emphasize analysis that focuses more on therapeutic targets and potential hot spots of the spike proteins. It is possible that vaccines and treatments targeting only the binding interaction of the virus may prove to be ineffective due to the steady emergence of new variants and mutants. Recent studies have shown that current COVID-19 vaccines may not be as efficacious on some variants of the virus (Krause et al., 2021). Preliminary trials indicate that COVID-19 vaccines are about 30% less effective on specific variants of concern than others in some places (Fontanet et al., 2021). With this in mind, attempting to discover the long-term efficacy of vaccines and treatments by identifying regions of the spike proteins and their role in

the activation process could prove valuable in the lab. Here, the mentioned activation process refers to the conformational changes in the prefusion spike protein, which takes place prior to ACE2 binding.

Prior research on spike protein conformation does not give complete insight into the activation process that occurs before binding to the host cell receptor. Earlier studies have detailed several partial or full-length structures of the spike protein ectodomains, potentially providing a rational basis for developing therapeutics that inhibit the binding process (Lan et al., 2020; Wrapp et al., 2020; Walls et al., 2020; Yan et al., 2020). However, the dynamic behavior in the prefusion spike protein cannot be identified using methods such as X-ray crystallography and cryogenic electron microscopy (cryo-EM). These methods specifically portray static images of the proteins in their active and inactive configurations. As a result, studies such as cryo-EM could not capture intermediate states in the activation mechanism, nor can they point to why one variant of SARS-CoV-2 may have enhanced infectivity and transmissibility. But, while their scope is limited, cryo-EM studies on SARS-CoV-1 and SARS-CoV-2 have shown that there are conformational changes that the RBD undergoes before engaging with host receptors: the state that conceals the determinants of receptor binding is referred to as the “down” state, while the state that exposes this section is referred to as the “up” state (Henderson et al., 2020; Barnes et al., 2020; Wrapp et al., 2020; Yuan et al., 2017; Walls et al., 2016). Other studies have addressed the changes in RBDs that arise due to various mutations, which cause altered interactions of antibody-binding surfaces and transformed conformations in virus variants (Cerutti et al., 2022). However, these methods focus more on the RBD-receptor interaction rather than the entire spike protein activation and possibly neglect the large-scale conformational changes of the protein.

Our lab aims to focus not only on the structures of the proteins and their RBDs but also on the extensive molecular basis of spike protein activations. For this purpose, we incorporate microsecond-level unbiased and biased molecular dynamics (MD) simulations.

Currently, all-atom molecular dynamics simulation is the only tool to supply a dynamic illustration of biomolecular processes at detailed molecular levels with spatiotemporal resolutions. Molecular dynamics can be summarized as a computer simulation technique that predicts the movement of atoms based on governing interatomic interactions over a given period (Hollingsworth & Dror, 2018). It replaces a static model with a dynamic model and has the ability to view any internal motions and resulting conformational changes of macromolecules (Karplus & McCammon, 2002). The theoretical basis of the approach is relatively straightforward: a trajectory of all atoms is developed by repeatedly solving Newton's equations of motion for the interacting system (Binder et al., 2004). Given the positions of atoms in a system and appropriate potential energies between them, one can calculate forces exerted on each atom by the other atoms and ultimately yield new positions and velocities that characterize atom configurations at every point during the time interval (Hollingsworth & Dror, 2018). Applying MD simulations, we are able to examine the differential activation behavior of SARS-CoV-2 and the D614G mutation.

III. Methods

MD equilibrium simulations of D614G

We used all-atom equilibrium MD simulations to characterize the conformational dynamics of the SARS-CoV-2 D614G variant spike protein. Simulations were constructed and executed using VMD (Humphey, Dalke, & Schulten, 1999) and NAMD 3Alpha (Phillips et al., 2020),

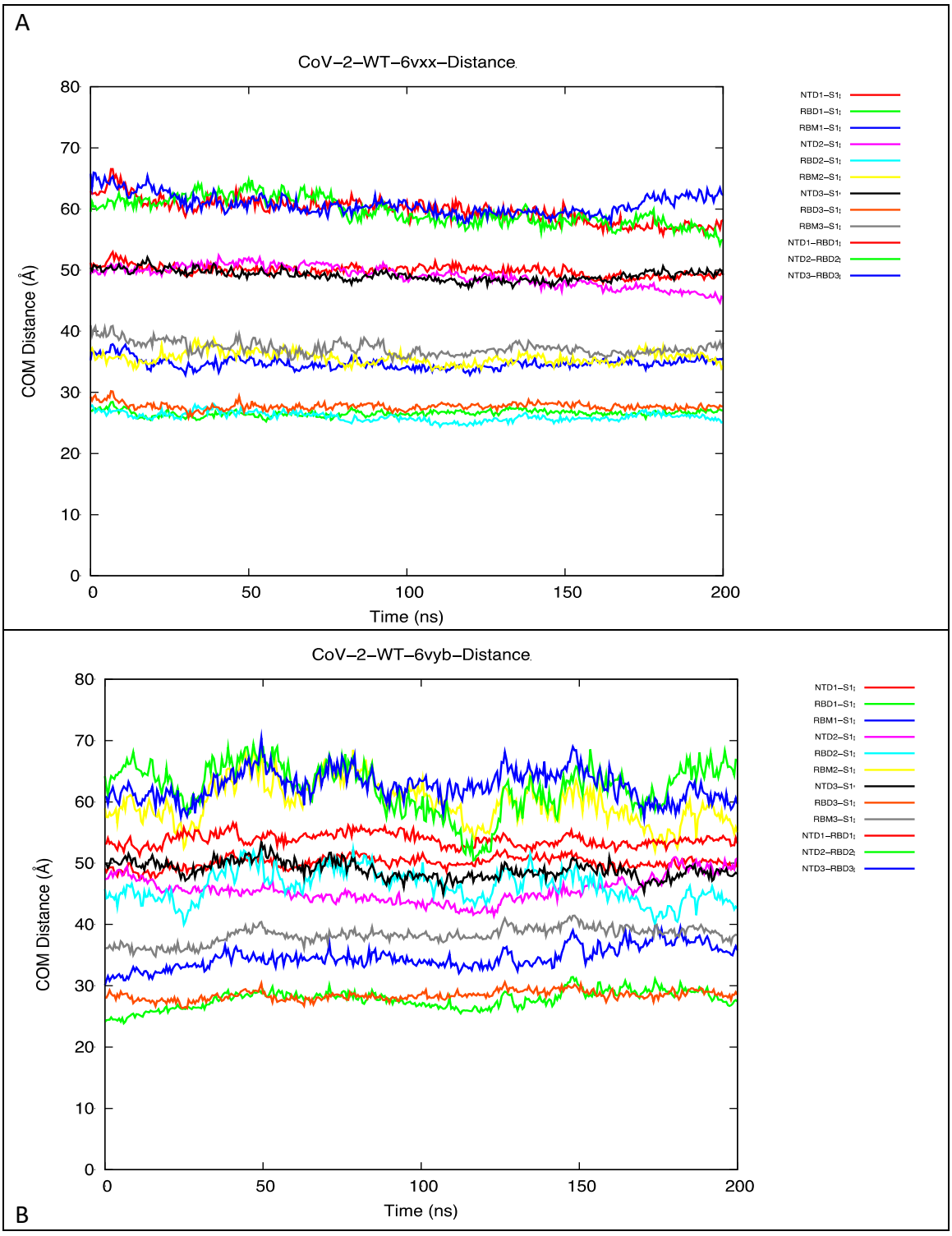
and were based on cryo-EM structures of the WT SARS-CoV-2 spike protein in the active (PDB entry:6VYB) (Walls et al., 2020) and inactive (PDB entry:6VXX) (Walls et al., 2020) states. CHARMM-GUI website was used to perform the D614G mutation on all three protomers of WT SARS-CoV-2 (Jo et al., 2008). Cumulative of 36 disulfide bonds were added to simulation models according to the information on the respective PDB files of SARS-CoV-2 (Walls et al., 2020; Jo et al., 2008; Lee et al., 2016). Systems were solvated in a cubic water box of 190 Å X 190 Å 190 Å and neutralized with 150 mM NaCl to mimic physiological conditions. An explicit water solvent (TIP3P) was used, and the systems contained approximately 800,000 atoms. A Monte Carlo algorithm within CHARMM-GUI was used to iteratively minimize the energy of the system. 10,000 Monte Carlo iterations were used to generate the initial models for the equilibrium simulations and CHARMM-GUI was then used to build the simulation systems (Jo et al., 2008; Lee et al., 2016).

Energy-minimization was performed for 10,000 steps utilizing the conjugate gradient algorithm (Reid J.K., 1971). Subsequently, systems were relaxed using restrained MD simulations in a stepwise approach utilizing the standard CHARMM-GUI relaxation step (Jo et al., 2008; Lee et al., 2016). Backbone and sidechain restraints were used for 10 ns in the following step with a force constant of 1 kcal/mol.$Å^2$ and 0.5 kcal/mol.$Å^2$ respectively. Then systems were equilibrated without bias for an extra 10 ns. The initial relaxation was performed in an NVT ensemble. The production simulations were run in an NPT ensemble. A timestep of 2 fs was used for the simulations at 310 K using a Langevin integrator with a damping coefficient of $\gamma = 0.5\ ps^{-1}$. Using the Nose-Hoover Langevin piston method the pressure in the simulation box was maintained at 1 atm (Reid J.K., 1971; Martyna G. J., Tobias D. J., & Klein M. L., 1994). For nonbonded interactions, the smoothed cutoff distance of 10 to 12 Å was used and particle mesh Ewald (PME)

method was used for computing long-range electrostatic interactions (Darden T., York D., & Pederson L., 1993). The production runs were then extended to 200 ns and conformations were collected every 100 picoseconds for trajectory analyses. In our analysis, we incorporated data from all-atom equilibrium MD simulations performed on the wild-type (WT) SARS-CoV-2 variant spike protein and used the same distance, RMSF, and salt-bridge analyses methodology delineated at (Govind et al., 2022).

IV. Results

Distance Analysis of SARS-CoV-2 WT spike protein and D614G spike protein



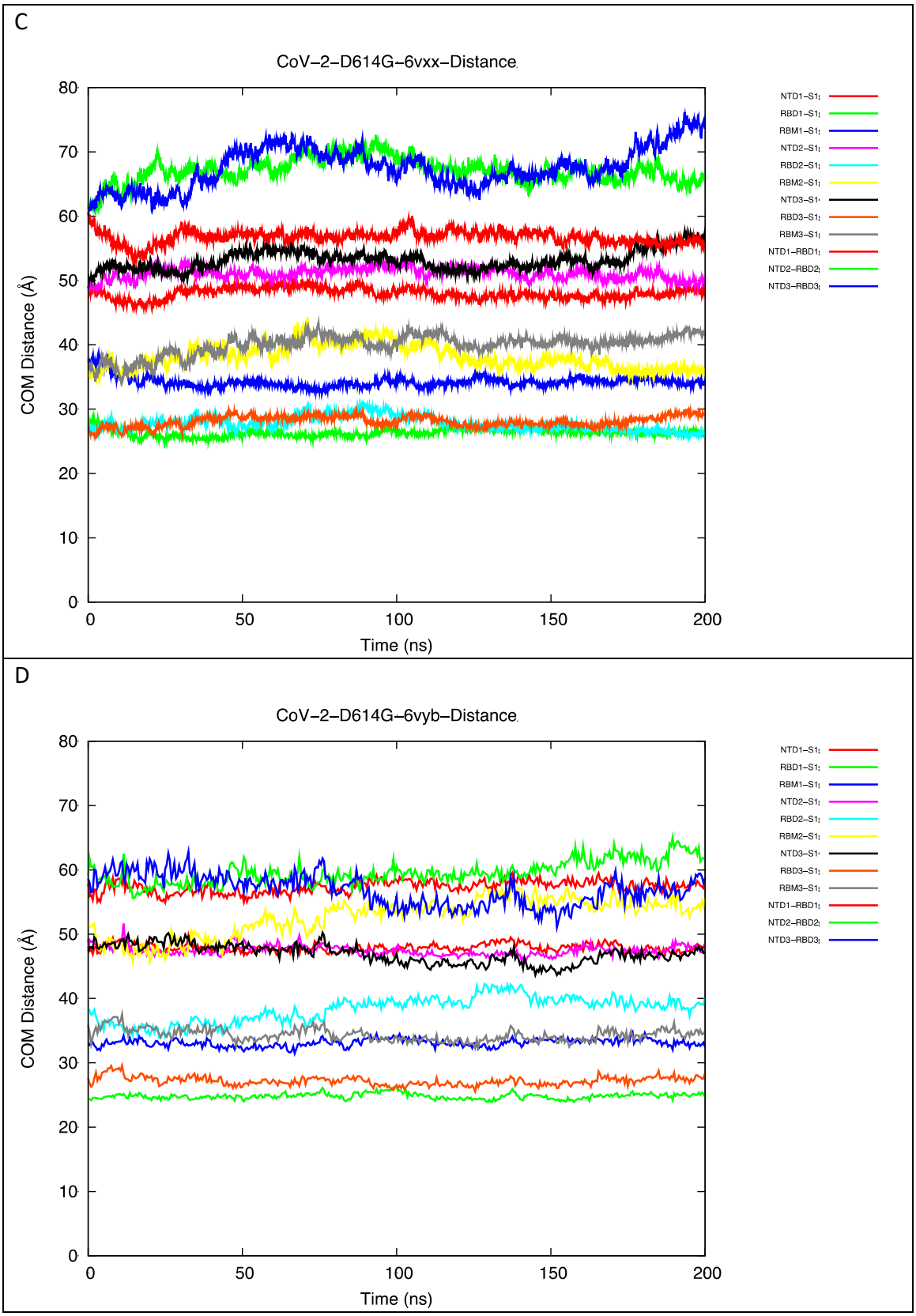

C
CoV–2–D614G–6vxx–Distance
COM Distance (Å)
Time (ns)
0
10
20
30
40
50
60
70
80
0
50
100
150
200
NTD1–S1
RBD1–S1
RBM1–S1
NTD2–S1
RBD2–S1
RBM2–S1
NTD3–S1
RBD3–S1
RBM3–S1
NTD1–RBD1
NTD2–RBD2
NTD3–RBD3
D
CoV–2–D614G–6vyb–Distance
COM Distance (Å)
Time (ns)
NTD1–S1
RBD1–S1
RBM1–S1
NTD2–S1
RBD2–S1
RBM2–S1
NTD3–S1
RBD3–S1
RBM3–S1
NTD1–RBD1
NTD2–RBD2
NTD3–RBD3

Figure 3: Distance analysis of different domains of the spike protein in the inactive and active states of both Wild Type and D614G
3A: Distance analysis of the inactive, closed Wild Type variant
3B: Distance analysis of the active, open Wild Type variant
3C: Distance analysis of the inactive, closed D614G variant
3D: Distance analysis of the active, open D614G variant

Distance analysis was conducted on all variant states to visualize relative flexibility and movement within independent internal regions of each spike protein (Fig. 3). Distance evaluation was carried out by measuring the Center of Mass (COM) distance between separate domains of the spike protein and allowed us to view how each area dynamically moved with respect to other regions. For instance, when measuring the COM distance in the RBD1-S1 complex, we were able to distinguish the relative distance between the RBD in protomer 1 and the S1 subunit and see how this distance evolved over time. This way, magnitudes of any conformational changes in that area were capable of being observed based on whether one region of the spike protein grew apart or closer to another reference region.

With this in mind, we set out to visualize how the D614G mutation may alter sets of interactions in individual areas of the spike protein. First, the D614G mutation appeared to induce an overall larger range of distances between internal domains in the inactive, closed D614G variant (3C) compared to the inactive, closed WT variant (3A). One example that expressed this sentiment was the RBM1-S1 distance ranging from ~60 Å to ~75 Å in the closed D614G state, while the RBM1-S1 distance in the closed WT state only varied from ~60 Å to ~65 Å. These results indicated that the D614G mutation allowed more freedom of movement to the inactive spike protein in multiple areas, possibly resulting in a higher potential of activation. The

D614G mutation altered interactions throughout the protein and was not limited to the vicinity that it was induced.

Further comparisons also supported this concept of modified interactions; while the RBM2-S1 and RBD2-S1 ranges increased from the inactive to active states in both variants, the distances these ranges encompassed were shown to be distinct between the active D614G (3D) and active WT (3B). To expound, the range of the RBM2-S1 complex stayed around 13 Å in both the active WT and active D614G. However, in the active WT state, the distance range varied from ~53 Å to ~66 Å, while in the active D614G state, the distance range varied from ~45 Å to ~58 Å. Similarly, the range of RBD2-S1 stayed around 9 Å in both active variants. In the active WT state, distance varied from ~40 Å to ~50 Å, while in the active D614G state, distance varied from ~34 Å to ~43 Å. In these cases, while the range was not affected due to the D614G variant, the overall distance was, implying that the mutation led to a different series of interactions that altered the structure of the spike protein to a new, unique formation.

Newly induced internal sets of interactions caused by the D614G mutation may also have influenced the preferred state of each variant. When comparing the inactive WT state (3A) to the active WT state (3B), it was reasoned that the WT variant favored the inactive state due to its increased relative stability and decreased fluctuation. However, when comparing the inactive (3C) and active (3D) states of the D614G variant, the mutated variant showed more preference for its active form; in the D614G variant, the active state of the spike protein displayed more relative stability and less fluctuation than the inactive state. This result would imply that the D614G mutation may have altered the activation mechanism of the spike protein, promoting conformational changes which further support the active state. This shift in preferred conditions

between variants of the virus could potentially be one of the reasons variants might show different sets of characteristics compared to their original counterparts.

Comparative analysis of salt bridges found in D614G variant and WT variant

Some of these internal interactions and differences in stability can possibly be attributed to salt bridges between opposing amino acid side chain charges. We analyzed the presence of salt bridges in each of the inactive and active states of the WT and D614G variants. In our investigation, we emphasized spotlighting salt bridges that were between different protomers in the spike protein and within either the NTD or RBD domains, as these areas provide major implications for the activation mechanism. It is important to note that the following salt bridges were not the only salt bridges found, and they were primarily meant to serve as examples that showed new sets of interactions between different variants. With this being said, our results revealed that there were a multitude of salt bridges that were not shared between the D614G and WT variants of the virus. These include five salt bridges that were present in the PROB - PROC complex in the inactive WT but not present in the inactive D614G: LYS 462-ASP 198, ARG 357-ASP 228, ASP 405-ARG 408, ASP 405-LYS 378, and ASP 571-ARG 44. One other salt bridge, ASP 198-LYS462, was found in the PROA-PROC complex in the inactive WT but was missing in the inactive D614G (Fig. 5).

A

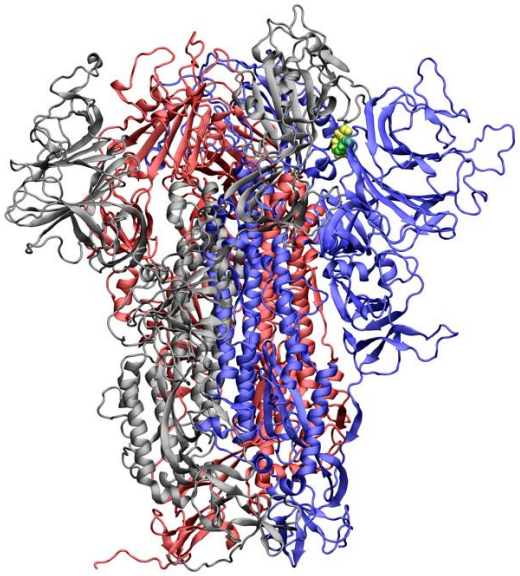


Figure 5: *Salt Bridge Analysis* **5A:** Aspartic Acid 198 (Green) and Lysine 462 (Yellow) in the inactive WT variant PRO1 – PRO3 complex

B

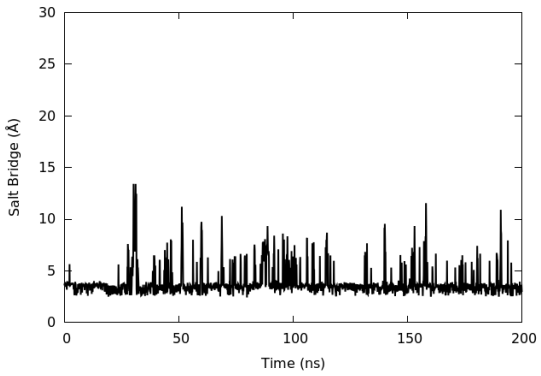


Figure 5B: This plot demonstrates that this salt bridge remains below 4 Å for about 80% of the simulation. Nitrogen – Oxygen interactions were measured for the salt bridge and normally 4 Å signifies a salt bridge.

The absence of these salt bridges in the D614G variant suggested that this mutation altered protein conformation and had the ability to create new sets of interactions throughout the protein. At the same time, the presence of these interactions in only the inactive WT variant may additionally explain why the WT variant's inactive state showed more stability than the D614G variant's inactive state. However, it is important to note that one salt bridge does not causally influence the overall stability or flexibility of a spike protein, and multiple tests would have to be performed to ensure the determinants for increased stability.

There are other factors that point towards the D614G mutation being able to induce new sets of electrostatic interactions in the spike protein. One result that backed up this claim was the appearance of a salt bridge in the PROA – PROC complex of the D614G inactive state that was

not present in the WT inactive state. This newly introduced salt bridge, ASP405-ARG408, further suggested that the D614G mutation shifted structural arrangements of the spike protein, resulting in the gain and loss of certain amino acid molecular binding (Fig. 6). As shown in Figure 6B, the salt bridge only appeared midway through the simulation, indicating that the D614G mutation altered sets of interactions and caused the salt bridge to form.

A

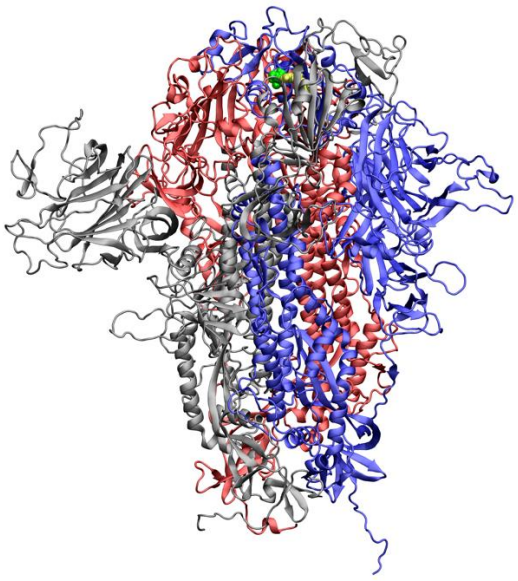


Figure 6: *Salt Bridge Analysis* **6A:** Aspartic Acid 405 (Green) and Arginine 408 (Yellow) salt bridge in the inactive D614G variant PRO1 – PRO3 complex

B

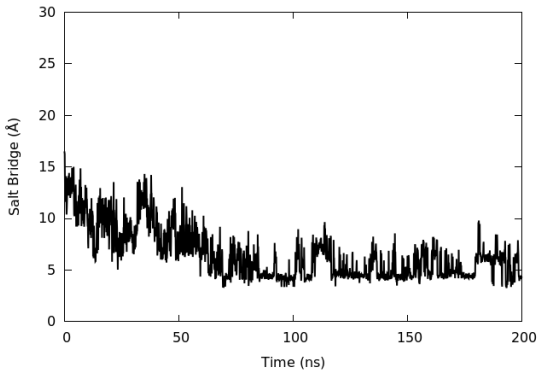


Figure 6B: This plot demonstrates that this salt bridge remains below 4 Å for much of the last half of the simulation.

Differences in interactions governing the D614G and WT variants were not limited to only the inactive states, however. Four new salt bridges were present in the D614G active state that were missing in the WT active state. These four were GLU 471-LYS 77 and GLU 484-LYS 378 in the PROA – PROB complex, ASP 571-ARG44 in the PROA – PROC complex, and ASP 571-ARG 44 in the PROB – PROC complex. Salt bridge GLU 471-LYS 77 can be seen in Figure 7 below.

A

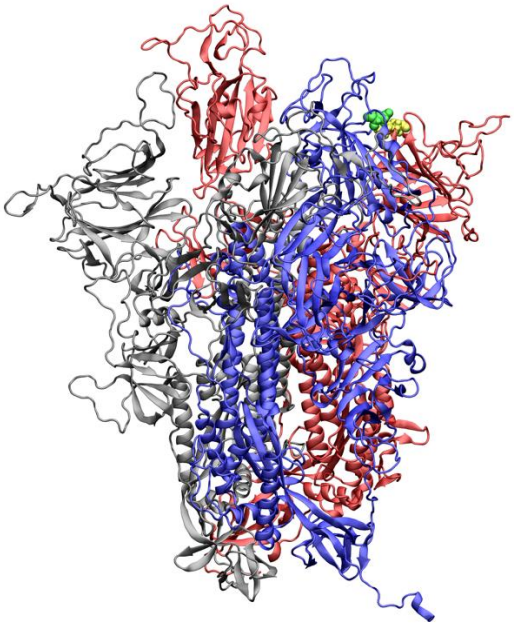


Figure 7: *Salt Bridge Analysis* **7A:** Glutamic Acid 471 (Green) and Lysine 77 (Yellow) salt bridge in the active D614G variant PRO1 – PRO2 complex

B

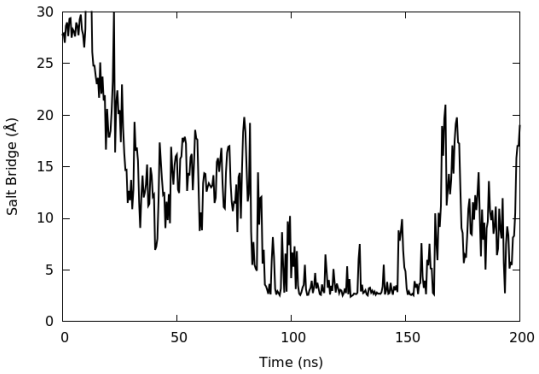


Figure 7B: This plot demonstrates the distance of the salt bridge with respect to time.

From these results, we concluded that the D614G single mutation not only altered amino acid configuration in the inactive closed state, but also in the active, open state as well. As the D614G mutation modified multiple sets of interactions between amino acids, it potentially altered structural dynamics and activation mechanisms throughout the spike protein. Also, the appearance of multiple salt bridges in the active D614G state perhaps added to the increased relative stability that this variant exhibited compared to its WT counterpart.

V. Concluding Remarks

The COVID-19 pandemic, caused by severe acute respiratory syndrome coronavirus 2, has influenced the worldwide population in many ways, including the health and safety of individuals,

the economy, and daily life in general. It has been the cause of millions of casualties and infections, and recently has been spreading due to advantageous amino acid structure mutations. Mutations such as D614G are a cause for concern, as vaccines and therapeutics targeting the initial SARS-CoV-2 have proven to be less effective on variants of the virus. Understanding the differential behavior of SARS coronaviruses and their relation to structural dynamics of spike proteins could provide insight into understanding how the spreading of SARS-CoV-2 occurs, leading to further intuition and discovery.

There were some limitations in the analysis that we performed. First, cryo-EM structures of the spike protein were relied on as our initial models for the simulations that we executed. Second, glycosylated spike proteins were not simulated due to the difficulty of modeling the correct glycan chains. Glycans have been proven to have sizable roles in both viral infection and evasion of immune response, but in our modelling's present condition, it would be difficult to determine whether conformational changes were occurring due to the internal protein dynamics or the glycosylation patterns of spike proteins. Finally, considering the information previously mentioned and the need for more research on the topic, we are unable to designate any causal inferences that the D614G mutation's induced conformational changes solely lead to the differences in virus characteristics.

Future research on the differential behavior of spike protein activation is recommended to explore this topic further. Already in the lab, free-energy calculations are being executed to analyze the energy barriers that are required in order for the spike proteins to shift from their inactive to active states. Also, as previously mentioned, further experiments involving the

glycosylated spike proteins must be performed to observe more accurate representations of the activation mechanism.

Nonetheless, the findings in this lab can prove to be significant and help lay the groundwork for future simulations. Results from studying differential behaviors showed that the D614G mutation altered spike protein sets of interactions in multiple regions, thereby changing the conformation of the structure as a whole. Results also indicated that the D614G mutation shifted the preferred state of the variants from inactive in the WT to active in the D614G. These conclusions could potentially have an impact on the activation mechanisms of spike proteins and the mediation of viral entry into a host cell.

VI. References

1) Harapan, H., Itoh, N., Yufika, A., Winardi, W., Synat K., Haypheng T., Megawati, D., Hayati, Z., Wagner, A. L., Mudatsir, M. (2020). Coronavirus disease 2019 (COVID-19): A literature review. *Journal of Infection and Public Health*, *12*(5), 667–673.
2) Ciotti, M., Angeletti, S., Minieri, M., Giovannetti, M., Benvenuto, D., Pascarella, S., ... & Ciccozzi, M. (2019). COVID-19 outbreak: an overview. Chemotherapy, 64(5-6), 215-223.
3) Hasöksüz, M., Kiliç, S., & Saraç, F. (2020). Coronaviruses and sars-cov-2. *Turkish journal of medical sciences*, *50*(SI-1), 549-556.
4) Malik, Y. A. (2020). Properties of coronavirus and SARS-CoV-2. *The Malaysian Journal of Pathology, 42*(1), 3-11.
5) World Health Organization. (2021). *WHO Coronavirus (COVID-19) Dashboard*. Retrieved from https://covid19.who.int/
6) Tortorici, M. A., & Veesler, D. (2019). Structural insights into coronavirus entry. In *Advances in virus research* (Vol. 105, pp. 93-116). Academic Press.
7) Duan, L., Zheng, Q., Zhang, H., Niu, Y., Lou, Y., & Wang, H. (2020). The SARS-CoV-2 spike glycoprotein biosynthesis, structure, function, and antigenicity: implications for the design of spike-based vaccine immunogens. *Frontiers in immunology*, 11, 2593.
8) Ke, Z., Oton, J., Qu, K., Cortese, M., Zila, V., McKeane, L., ... & Briggs, J. A. (2020). Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature, 588*(7838), 498-502.
9) Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell, 181*(2), 281-292.
10) Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., ... & Pöhlmann, S. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell, 181*(2), 271-280.
11) Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., ... & Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature*, *581*(7807), 221-224.
12) Wang, Q., Zhang, Y., Wu, L., Niu, S., Song, C., Zhang, Z., ... & Qi, J. (2020). Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell*, *181*(4), 894-904.
13) Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., ... & McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, *367*(6483), 1260-1263.
14) Kirchdoerfer, R. N., Cottrell, C. A., Wang, N., Pallesen, J., Yassine, H. M., Turner, H. L., ... & Ward, A. B. (2016). Pre-fusion structure of a human coronavirus spike protein. *Nature, 531*(7592), 118-121.
15) Millet, J. K., & Whittaker, G. R. (2015). Host cell proteases: critical determinants of coronavirus tropism and pathogenesis. *Virus research*, *202*, 120-134.
16) Tai, W., He, L., Zhang, X., Pu, J., Voronin, D., Jiang, S., ... & Du, L. (2020). Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for

development of RBD protein as a viral attachment inhibitor and vaccine. *Cellular & molecular immunology*, *17*(6), 613-620.

17) Li, F., Li, W., Farzan, M., & Harrison, S. C. (2005). Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science*, *309*(5742), 1864-1868.
18) Ge, X. Y., Li, J. L., Yang, X. L., Chmura, A. A., Zhu, G., Epstein, J. H., ... & Shi, Z. L. (2013). Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature, 503*(7477), 535-538.
19) Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., ... & Farzan, M. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*, *426*(6965), 450-454.
20) Huang, Y., Yang, C., Xu, X. F., Xu, W., & Liu, S. W. (2020). Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacologica Sinica*, *41*(9), 1141-1149.
21) Woo, H., Park, S. J., Choi, Y. K., Park, T., Tanveer, M., Cao, Y., ... & Im, W. (2020). Developing a fully glycosylated full-length SARS-CoV-2 spike protein model in a viral membrane. *The journal of physical chemistry B*, *124*(33), 7128-7137.
22) Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., ... & Shi, Z. L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *nature*, *579*(7798), 270-273.
23) Govind Kumar, V., Ogden, D. S., Isu, U. H., Polasa, A., Losey, J., & Moradi, M. (2022). Prefusion spike protein conformational changes are slower in SARS-COV-2 than in SARS-COV-1. *Journal of Biological Chemistry*, 101814. https://doi.org/10.1016/j.jbc.2022.101814
24) Hu, B., Guo, H., Zhou, P., & Shi, Z. L. (2021). Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology*, *19*(3), 141-154.
25) Wu, K., Peng, G., Wilken, M., Geraghty, R. J., & Li, F. (2012). Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. *Journal of Biological Chemistry*, *287*(12), 8904-8911.
26) Petersen, E., Koopmans, M., Go, U., Hamer, D. H., Petrosillo, N., Castelli, F., ... & Simonsen, L. (2020). Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *The Lancet infectious diseases*, *20*(9), e238-e244.
27) Ferretti, L., Wymant, C., Kendall, M., Zhao, L., Nurtay, A., Abeler-Dörner, L., ... & Fraser, C. (2020). Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*, *368*(6491), eabb6936.
28) Campbell, F., Archer, B., Laurenson-Schafer, H., Jinnai, Y., Konings, F., Batra, N., ... & De Waroux, O. L. P. (2021). Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. *Eurosurveillance*, *26*(24), 2100509.
29) Karim, S. S. A., & Karim, Q. A. (2021). Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic. *The Lancet*, *398*(10317), 2126-2128.
30) Department of Health and Social Care. New measures to boost response to the B1.617.2 variant. Press release. London: GOV.UK; 13 May 2021. Available from: https://www.gov.uk/government/news/new-measures-to-boost-response-to-the-b16172-variant

31) Earnest, R., Uddin, R., Matluk, N., Renzette, N., Turbett, S. E., Siddle, K. J., ... & Grubaugh, N. D. (2022). Comparative transmissibility of SARS-CoV-2 variants delta and alpha in New England, USA. *Cell Reports Medicine*, 100583.

32) Luo, C. H., Morris, C. P., Sachithanandham, J., Amadi, A., Gaston, D., Li, M., ... & Mostafa, H. H. (2021). Infection with the SARS-CoV-2 delta variant is associated with higher infectious virus loads compared to the alpha variant in both unvaccinated and vaccinated individuals. *medRxiv*.

33) Wang, Y., Chen, R., Hu, F., Lan, Y., Yang, Z., Zhan, C., ... & Tang, X. (2021). Transmission, viral kinetics and clinical characteristics of the emergent SARS-CoV-2 Delta VOC in Guangzhou, China. *EClinicalMedicine*, *40*, 101129.

34) Townsend, J. P., Hassler, H. B., Wang, Z., Miura, S., Singh, J., Kumar, S., ... & Dornburg, A. (2021). The durability of immunity against reinfection by SARS-CoV-2: a comparative evolutionary study. *The Lancet Microbe*, *2*(12), e666-e675.

35) Koyama, T., Platt, D., & Parida, L. (2020). Variant analysis of SARS-CoV-2 genomes. *Bulletin of the World Health Organization*, *98*(7), 495-504.

36) Isabel, S., Graña-Miraglia, L., Gutierrez, J. M., Bundalovic-Torma, C., Groves, H. E., Isabel, M. R., ... & Poutanen, S. M. (2020). Evolutionary and structural analyses of SARS-CoV-2 D614G spike protein mutation now documented worldwide. *Scientific reports*, *10*(1), 1-9.

37) Kaushal, N., Gupta, Y., Goyal, M., Khaiboullina, S. F., Baranwal, M., & Verma, S. C. (2020). Mutational frequencies of SARS-CoV-2 genome during the beginning months of the outbreak in USA. *Pathogens*, *9*(7), 565.

38) van Dorp, L., Acman, M., Richard, D., Shaw, L. P., Ford, C. E., Ormond, L., ... & Balloux, F. (2020). Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infection, Genetics and Evolution*, *83*, 104351.

39) Guruprasad, L. (2021). Human SARS CoV-2 spike protein mutations. *Proteins: Structure, Function, and Bioinformatics*, *89*(5), 569-576.

40) Zhang, L., Jackson, C. B., Mou, H., Ojha, A., Rangarajan, E. S., Izard, T., ... & Choe, H. (2020). The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. *BioRxiv*.

41) Morais, I. J., Polveiro, R. C., Souza, G. M., Bortolin, D. I., Sassaki, F. T., & Lima, A. T. M. (2020). The global population of SARS-CoV-2 is composed of six major subtypes. *Scientific reports*, *10*(1), 1-9.

42) Laha, S., Chakraborty, J., Das, S., Manna, S. K., Biswas, S., & Chatterjee, R. (2020). Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission. *Infection, genetics and evolution*, *85*, 104445.

43) Comandatore, F., Chiodi, A., Gabrieli, P., Biffignandi, G. B., Perini, M., Ricagno, S., ... & Brilli, M. (2020). Insurgence and worldwide diffusion of genomic variants in SARS-CoV-2 genomes. *bioRxiv*.

44) Cortey, M., Li, Y., Díaz, I., Clilverd, H., Darwich, L., & Mateu, E. (2020). SARS-CoV-2 amino acid substitutions widely spread in the human population are mainly located in highly conserved segments of the structural proteins. *bioRxiv*.

45) Furuyama, T. N., Antoneli, F., Carvalho, I. M., Briones, M. R., & Janini, L. M. (2020). Temporal data series of COVID-19 epidemics in the USA, Asia and Europe suggests a selective sweep of SARS-CoV-2 Spike D614G variant. *arXiv preprint arXiv:2006.11609*.

46) Yurkovetskiy, L., Wang, X., Pascal, K. E., Tomkins-Tinch, C., Nyalile, T. P., Wang, Y., ... & Luban, J. (2020). Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant. *Cell*, *183*(3), 739-751.

47) Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., ... & Montefiori, D. C. (2020). Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell*, *182*(4), 812-827.

48) Volz, E., Hill, V., McCrone, J. T., Price, A., Jorgensen, D., O'Toole, Á., ... & Allan, J. (2021). Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell*, *184*(1), 64-75.

49) Daniloski, Z., Jordan, T. X., Ilmain, J. K., Guo, X., Bhabha, G., & Sanjana, N. E. (2021). The Spike D614G mutation increases SARS-CoV-2 infection of multiple human cell types. *Elife*, *10*, e65365.

50) Hou, Y. J., Chiba, S., Halfmann, P., Ehre, C., Kuroda, M., Dinnon III, K. H., ... & Baric, R. S. (2020). SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science*, *370*(6523), 1464-1468.

51) Wang, C., Zheng, Y., Niu, Z., Jiang, X., & Sun, Q. (2021). The virological impacts of SARS-CoV-2 D614G mutation. *Journal of Molecular Cell Biology*, *13*(10), 712-720.

52) Jackson, C. B., Zhang, L., Farzan, M., & Choe, H. (2021). Functional importance of the D614G mutation in the SARS-CoV-2 spike protein. *Biochemical and Biophysical Research Communications*, *538*, 108-115.

53) SARS-CoV-2 Variants of Concern. UpToDate. https://www.uptodate.com/contents/image?imageKey=ID%2F131216. Accessed March 22, 2022.

54) Krause, P. R., Fleming, T. R., Longini, I. M., Peto, R., Briand, S., Heymann, D. L., ... & Henao-Restrepo, A. M. (2021). SARS-CoV-2 variants and vaccines. *New England Journal of Medicine*, *385*(2), 179-186.

55) Fontanet, A., Autran, B., Lina, B., Kieny, M. P., Karim, S. S. A., & Sridhar, D. (2021). SARS-CoV-2 variants and ending the COVID-19 pandemic. *The Lancet*, *397*(10278), 952-954.

56) Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., ... & Wang, X. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, *581*(7807), 215-220.

57) Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., & Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*, *367*(6485), 1444-1448.

58) Henderson, R., Edwards, R. J., Mansouri, K., Janowska, K., Stalls, V., Gobeil, S., ... & Acharya, P. (2020). Controlling the SARS-CoV-2 spike glycoprotein conformation. *Nature structural & molecular biology*, *27*(10), 925-933.

59) Barnes, C. O., Jette, C. A., Abernathy, M. E., Dam, K. M. A., Esswein, S. R., Gristick, H. B., ... & Bjorkman, P. J. (2020). SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature*, *588*(7839), 682-687.

60) Yuan, Y., Cao, D., Zhang, Y., Ma, J., Qi, J., Wang, Q., ... & Gao, G. F. (2017). Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nature communications*, *8*(1), 1-9.

61) Walls, A. C., Tortorici, M. A., Bosch, B. J., Frenz, B., Rottier, P. J., DiMaio, F., ... & Veesler, D. (2016). Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature*, *531*(7592), 114-117.

62) Cerutti, G., Guo, Y., Liu, L., Liu, L., Zhang, Z., Luo, Y., ... & Shapiro, L. (2022). Cryo-EM structure of the SARS-CoV-2 Omicron spike. *Cell reports*, *38*(9), 110428.

63) Hollingsworth, S. A., & Dror, R. O. (2018). Molecular dynamics simulation for all. *Neuron*, *99*(6), 1129-1143.

64) Binder, K., Horbach, J., Kob, W., Paul, W., & Varnik, F. (2004). Molecular dynamics simulations. *Journal of Physics: Condensed Matter*, *16*(5), S429.

65) Humphrey, W., Dalke, A. and Schulten, K., "VMD - Visual Molecular Dynamics", J. Molec. Graphics, 1996, vol. 14, pp. 33-38.

66) James C. Phillips, David J. Hardy, Julio D. C. Maia, John E. Stone, Joao V. Ribeiro, Rafael C. Bernardi, Ronak Buch, Giacomo Fiorin, Jerome Henin, Wei Jiang, Ryan McGreevy, Marcelo C. R. Melo, Brian K. Radak, Robert D. Skeel, Abhishek Singharoy, Yi Wang, Benoit Roux, Aleksei Aksimentiev, Zaida Luthey-Schulten, Laxmikant V. Kale, Klaus Schulten, Christophe Chipot, and Emad Tajkhorshid. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *Journal of Chemical Physics*, 153:044130, 2020. doi:10.1063/5.0014475

67) Jo, S., Kim, T., Iyer, V.G. & Im, W. (2008). CHARMM-GUI: A web-based graphical interface for CHARMM. J. Comp. Chem., *29*, 1859-1865.

68) Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., ... & Im, W. (2016). CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *Journal of chemical theory and computation*, *12*(1), 405-413.

69) Reid, J. K. (1971). On the method of conjugate gradients for the solution of large sparse systems of linear equations. In *Pro. the Oxford conference of institute of mathmatics and its applications* (pp. 231-254).

70) Martyna, G. J., Tobias, D. J., & Klein, M. L. (1994). Constant pressure molecular dynamics algorithms. *The Journal of chemical physics*, *101*(5), 4177-4189.

71) Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *The Journal of chemical physics*, *98*(12), 10089-10092.